Specification tests for non-Gaussian maximum likelihood estimators*

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

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

December 2016
Revised: November 2020

Abstract

We propose generalised DWH specification tests which simultaneously compare three or more likelihood-based estimators in multivariate conditionally heteroskedastic dynamic regression models. Our tests are useful for GARCH models and in many empirically relevant macro and finance applications involving VARs and multivariate regressions. We determine the rank of the differences between the estimators' asymptotic covariance matrices under correct specification, and take into account that some parameters remain consistently estimated under distributional misspecification. We provide finite sample results through Monte Carlo simulations. Finally, we analyse a structural VAR proposed to capture the relationship between macroeconomic and financial uncertainty and the business cycle.

Keywords: Durbin-Wu-Hausman Tests, Partial Adaptivity, Semiparametric Estimators, Singular Covariance Matrices, Uncertainty and the Business Cycle.

JEL: C12, C14, C22, C32, C52

*This paper draws heavily on Fiorentini and Sentana (2007). In addition to those explicitly acknowledged there, we would like to thank Dante Amengual, Christian Bontemps and Luca Fanelli for useful comments and suggestions, as well as audiences at CREST, Konstanz, Tokyo, UCLA, UCSD, the University of Liverpool 6^{th} Annual Econometrics Workshop (April 2017), the University of Southampton Finance and Econometrics Workshop (May 2017), the SanFI conference on "New Methods for the Empirical Analysis of Financial Markets" (Comillas, June 2017), the 70^{th} ESEM (Lisbon, August 2017), the Bilgi CEFIS conference on "Advances in Econometrics Methods" (Instanbul, March 2018), the Financial Engineering and Risk Management Conference (Shanghai, June 2018), the University of Kent "50 Years of Econometrics at Keynes College" Conference (Canterbury, September 2018), and the TSE Financial Econometrics Conference (Toulouse, May 2019). The co-editor (Tao Zha) and two anonymous referees have provided very useful feedback. Of course, the usual caveat applies. The second author acknowledges financial support from the Spanish Ministry of Economy, Industry & Competitiveness through grant ECO 2017-89689 and the Santander CEMFI Research Chair.

1 Introduction

Empirical studies with financial data suggest that returns distributions are leptokurtic even after controlling for volatility clustering effects. This feature has important practical consequences for standard risk management measures such as Value at Risk and recently proposed systemic risk measures such as Conditional Value at Risk or Marginal Expected Shortfall (see Adrian and Brunnermeier (2016) and Acharya et al. (2017), respectively), which could be severely mismeasured by assuming normality. Given that empirical researchers are interested in those risk measures for several probability levels, they often specify a parametric leptokurtic distribution, which then they use to estimate their models by maximum likelihood (ML).

A non-trivial by-product of these non-Gaussian ML procedures is that they deliver more efficient estimators of the mean and variance parameters, especially if the shape parameters can be fixed to their true values. The downside, though, is that they often achieve those efficiency gains under correct specification at the risk of returning inconsistent parameter estimators under distributional misspecification (see e.g. Newey and Steigerwald (1997)). This is in marked contrast with the generally inefficient Gaussian pseudo-maximum likelihood (PML) estimators advocated by Bollerslev and Wooldridge (1992) among many others, which remain root-T consistent for the mean and variance parameters under relatively weak conditions.

If researchers were only interested in those two conditional moments, the semiparametric (SP) estimators of Engle and Gonzalez-Rivera (1991) and Gonzalez-Rivera and Drost (1999) would provide an attractive solution because they are consistent and also attain full efficiency for a subset of the parameters (see Linton (1993), Drost and Klaassen (1997), Drost, Klaassen and Werker (1997) and Sun and Stengos (2006) for univariate time series examples). Unfortunately, SP estimators suffer from the curse of dimensionality when the number of series involved, N, is moderately large, which limits their use. Furthermore, Amengual, Fiorentini and Sentana (2013) show that non-parametrically estimated conditional quantiles lead to risk measures with much wider confidence intervals than their parametric counterparts even in univariate contexts. Another possibility would be the spherically symmetric semiparametric (SSP) methods considered by Hodgson and Vorkink (2003) and Hafner and Rombouts (2007), which are also partially efficient while retaining univariate rates for their nonparametric part regardless of N. However, asymmetries in the true joint distribution will contaminate these estimators too.

In any event, given that many research economist at central banks, financial institutions and economic consulting firms continue to rely on the estimators that commercial econometric software packages provide, it would be desirable that they routinely complemented their empirical results with some formal indication of the validity of the parametric assumptions they make.

The statistical and econometric literature on model specification is huge. In this paper, our focus is the adequacy of the conditional distribution under the maintained assumption that the rest of the model is correctly specified. Even so, there are various ways of assessing it. One possibility is to nest the assumed distribution within a more flexible parametric family in order to conduct a Lagrange Multiplier (LM) test of the nesting restrictions. This is the approach in Mencía and Sentana (2012), who use the generalised hyperbolic family as an instrumental nesting distribution for the multivariate Student t. In contrast, other specification tests do not consider an explicit alternative hypothesis. A case in point are consistent tests based on the difference between the theoretical and empirical cumulative distribution functions of the innovations (Bai (2003) and Bai and Zhihong (2008)) or their characteristic functions (Bierens and Wang (2012) and Amengual, Carrasco and Sentana (2019)). An alternative procedure would be the information matrix test of White (1982), which compares some or all of the elements of the expected Hessian and the variance of the score. White (1987) also proposed the application of Newey's (1985) conditional moment test to assesses the martingale difference property of the scores under correct specification. Finally, the general class of moment tests in Newey (1985) and Tauchen (1985) could also be entertained, as Bontemps and Meddahi (2012) illustrate.

But when a research economist relies on standard software for calculating some non-Gaussian estimators of $\boldsymbol{\theta}$ and their asymptotic standard errors from real data, a more natural approach to testing distributional specification would be to compare those estimators on a pairwise basis using simple Durbin-Wu-Hausman (DWH) tests.[1] As is well known, the traditional version of these tests can refute the correct specification of a model by exploiting the diverging properties under misspecification of a pair of estimators of the same parameters. Focusing on the model parameters makes sense because if they are inconsistently estimated, the conditional moments derived from them will be inconsistently estimated too.

In this paper, we take this idea one step further and propose an extension of the DWH tests which simultaneously compares three or more estimators. The rationale for our proposal is given by a novel proposition which shows that if we order the five estimators we mentioned in the preceding paragraphs as restricted and unrestricted non-Gaussian ML, SSP, SP and Gaussian PML, each estimator is "efficient" relative to all the others behind. This "Matryoshka doll" structure for their joint asymptotic covariance matrix implies that there are four asymptotically independent contiguous comparisons, and that any other pairwise comparison must be a linear combination of those four. We exploit these properties in developing the asymptotic distribution

[1] Wu (1973) compared OLS with IV in linear single equation models to assess regressor exogeneity unaware that Durbin (1954) had already suggested this. Hausman (1978) provided a procedure with far wider applicability.

of our proposed multiple comparison tests. We also explore several important issues related to the practical implementation of DWH tests, including its two score versions, their numerical invariance to reparametrisations and their application to subsets of parameters.

To design reliable tests, we first need to figure out the rank of the difference between the asymptotic covariance matrices under the null of correct specification so as to use the right number of degrees of freedom. We also need to take into account that some parameters continue to be consistently estimated under the alternative of incorrect distributional specification, thereby avoiding wasting degrees of freedom without providing any power gains.

In Fiorentini and Sentana (2019), we characterised the mean and variance parameters that distributionally misspecified ML estimators can consistently estimate, and provided simple closed-form consistent estimators for the rest. One of the most interesting results that we obtain in this paper is that the parameters that continue to be consistently estimated by the parametric estimators under distributional misspecification are those which are efficiently estimated by the semiparametric procedures. In contrast, the remaining parameters, which will be inconsistently estimated by distributionally misspecified parametric procedures, the semiparametric procedures can only estimate with the efficiency of the Gaussian PML estimator. Therefore, we will focus our tests on the comparison of the estimators of this second group of parameters, for which the usual efficiency - consistency trade off is of first-order importance.

The inclusion of means and the explicit coverage of multivariate models make our proposed tests useful not only for GARCH models but also for dynamic linear models such as VARs or multivariate regressions, which remain the workhorse in empirical macroeconomics and asset pricing contexts. This is particularly relevant in practice because researchers are increasingly acknowledging the non-normality of many macroeconomic variables (see Lanne, Meitz and Saikkonen (2017) and the references therein for recent examples of univariate and multivariate time series models with non-Gaussian innovations). Nevertheless, structural models pose some additional inference challenges, which we discuss separately. Obviously, our approach also applies in cross-sectional models with exogenous regressors, as well as in static ones.

The rest of the paper is as follows. In section 2, we provide a quick revision of DWH tests and derive several new results which we use in our subsequent analysis. Then, in section 3 we formally present the five different likelihood-based estimators that we have mentioned, and derive our proposed specification tests, paying particular attention to their degrees of freedom and power. A Monte Carlo evaluation of our tests can be found in section 4, followed by an empirical analysis of the relationship between uncertainty and the business cycle using a structural VAR. Finally, we present our conclusions in section 6. Proofs and auxiliary results are gathered in appendices.

2 Durbin-Wu-Hausman tests

2.1 Wald and score versions

Let $\hat{\boldsymbol{\theta}}_T$ and $\tilde{\boldsymbol{\theta}}_T$ denote two GMM estimators of $\boldsymbol{\theta}$ based on the average influence functions $\bar{\mathbf{m}}_T(\boldsymbol{\theta})$ and $\bar{\mathbf{n}}_T(\boldsymbol{\theta})$ and weighting matrices $\tilde{\mathcal{S}}_{mT}$ and $\tilde{\mathcal{S}}_{nT}$, respectively. When both sets of moment conditions hold, then, under standard regularity conditions (see e.g. Newey and McFadden (1994)), the estimators will be jointly root-T consistent and asymptotically Gaussian, so

$$\begin{aligned}&\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Delta}) \ \textit{and} \\ &T(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T)'\boldsymbol{\Delta}^-(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) \xrightarrow{d} \chi^2_r, \end{aligned} \tag{1}$$

where $r = \text{rank}(\boldsymbol{\Delta})$ and $^-$ denotes a generalised inverse. Consider now a sequence of local alternatives such that

$$\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) \sim N(\boldsymbol{\theta}_m - \boldsymbol{\theta}_n, \boldsymbol{\Delta}). \tag{2}$$

In this case, the asymptotic distribution of the DWH statistics (1) will become a non-central chi-square with non-centrality parameter $(\boldsymbol{\theta}_m - \boldsymbol{\theta}_n)'\boldsymbol{\Delta}^-(\boldsymbol{\theta}_m - \boldsymbol{\theta}_n)$ and the same number of degrees freedom (see e.g. Hausman (1978) or Holly (1987)). Therefore, the local power of a DWH test will be increasing in the limiting discrepancy between the two estimators, and decreasing in both the number and magnitude of the non-zero eigenvalues of $\boldsymbol{\Delta}$.

Knowing the right number of degrees of freedom is particularly important for employing the correct distribution under the null. Unfortunately, some obvious consistent estimators of $\boldsymbol{\Delta}$ might lead to inconsistent estimators of $\boldsymbol{\Delta}^-$.[2] In fact, they might not even be positive semidefinite in finite samples. We will revisit these issues in sections 3.4 and 3.6, respectively.

The calculation of the DWH test statistic (1) requires the prior computation of $\hat{\boldsymbol{\theta}}_T$ and $\tilde{\boldsymbol{\theta}}_T$. In a likelihood context, however, Theorem 5.2 of White (1982) implies that an asymptotically equivalent test can be obtained by evaluating the scores of the restricted model at the inefficient but consistent parameter estimator (see also Reiss (1983) and Ruud (1984), as well as Davidson and MacKinnon (1989)). Theorem 2.5 in Newey (1985) shows that the same equivalence holds in situations in which the estimators are defined by moment conditions. In fact, it is possible to derive not just one but two asymptotically equivalent score versions of the DWH test by evaluating the influence functions that give rise to each of the estimators at the other estimator, as explained in section 10.3 of White (1994). The following proposition, which we include for completeness, spells out those equivalences:

[2] A trivial non-random example of discontinuities is the sequence $1/T$, which converges to 0 while its generalised inverse $(1/T)^- = T$ diverges. Theorem 1 in Andrews (1987) provides conditions under which a quadratic form based on a generalised inverse of a weighting matrix converges to a chi-square distribution.

Proposition 1 *Assume that the moment conditions* $\mathbf{m}_t(\boldsymbol{\theta})$ *and* $\mathbf{n}_t(\boldsymbol{\theta})$ *are correctly specified. Then, under standard regularity conditions*

$$T(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T)'\boldsymbol{\Delta}^-(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) - T\bar{\mathbf{m}}_T'(\tilde{\boldsymbol{\theta}}_T)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\boldsymbol{\Lambda}_m^-\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T) = o_p(1) \quad (3)$$
$$\text{and } T(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T)'\boldsymbol{\Delta}^-(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) - T\bar{\mathbf{n}}_T'(\hat{\boldsymbol{\theta}}_T)\mathcal{S}_n\mathcal{J}_n(\boldsymbol{\theta}_0)\boldsymbol{\Lambda}_n^-\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n\bar{\mathbf{n}}_T(\hat{\boldsymbol{\theta}}_T) = o_p(1), \quad (4)$$

where $\boldsymbol{\Lambda}_m$ *and* $\boldsymbol{\Lambda}_n$ *are the limiting variances of* $\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T)$ *and* $\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n\sqrt{T}\bar{\mathbf{n}}_T(\hat{\boldsymbol{\theta}}_T)$*, respectively, which are such that*

$$\boldsymbol{\Delta}=\left[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right]^{-1}\boldsymbol{\Lambda}_m\left[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right]^{-1}=\left[\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n\mathcal{J}_n(\boldsymbol{\theta}_0)\right]^{-1}\boldsymbol{\Lambda}_n\left[\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n\mathcal{J}_n(\boldsymbol{\theta}_0)\right]^{-1}$$

with $\mathcal{J}_m(\boldsymbol{\theta}) = \text{plim}_{T\to\infty}\partial\bar{\mathbf{m}}_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$, $\mathcal{J}_n(\boldsymbol{\theta}) = \text{plim}_{T\to\infty}\partial\bar{\mathbf{n}}_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$, $\mathcal{S}_m = \text{plim}_{T\to\infty}\tilde{\mathcal{S}}_{mT}$, $\mathcal{S}_n = \text{plim}_{T\to\infty}\tilde{\mathcal{S}}_{nT}$ *and* $\text{rank}\left[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right] = \text{rank}\left[\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n\mathcal{J}_n(\boldsymbol{\theta}_0)\right] = p = \dim(\boldsymbol{\theta})$*, so that* $\text{rank}(\boldsymbol{\Lambda}_m) = \text{rank}(\boldsymbol{\Lambda}_n) = \text{rank}(\boldsymbol{\Delta})$.

An intuitive way of re-interpreting the asymptotic equivalence between the original DWH test in (1) and the two alternative score versions on the right hand sides of (3) and (4) is to think of the latter as original DWH tests based on two convenient reparametrisations of $\boldsymbol{\theta}$ obtained through the population version of the first order conditions that give rise to each estimator, namely $\boldsymbol{\pi}_m(\boldsymbol{\theta}) = \mathcal{J}_m'(\boldsymbol{\theta})\mathcal{S}_mE[\mathbf{m}_t(\boldsymbol{\theta})]$ and $\boldsymbol{\pi}_n(\boldsymbol{\theta}) = \mathcal{J}_n'(\boldsymbol{\theta})\mathcal{S}_nE[\mathbf{n}_t(\boldsymbol{\theta})]$. While these new parameters are equal to 0 when evaluated at the pseudo-true values of $\boldsymbol{\theta}$ implicitly defined by the exactly identified moment conditions $\mathcal{J}_m'(\boldsymbol{\theta}_m)\mathcal{S}_mE[\mathbf{m}_t(\boldsymbol{\theta}_m)]=\mathbf{0}$ and $\mathcal{J}_n'(\boldsymbol{\theta}_n)\mathcal{S}_nE[\mathbf{n}_t(\boldsymbol{\theta}_n)]=\mathbf{0}$, respectively, $\boldsymbol{\pi}_m(\boldsymbol{\theta}_n)$ and $\boldsymbol{\pi}_n(\boldsymbol{\theta}_m)$ are not necessarily so, unless the correct specification condition $\boldsymbol{\theta}_m = \boldsymbol{\theta}_n = \boldsymbol{\theta}_0$ holds.[3] The same arguments also allow us to loosely interpret the score versions of the DWH tests as distance metric tests of those moment conditions, as they compare the values of the GMM criteria at the estimator which sets those exactly identified moments to 0 with their values at the alternative estimator. We will discuss more formal links to the classical Wald, Likelihood Ratio (LR) and LM tests in a likelihood context in section 3.4.

Proposition 1 implies the choice between the three versions of the DWH test must be based on either computational ease, numerical invariance or finite sample reliability. While computational ease is model specific, we will revisit the last two issues in sections 2.2 and 4, respectively.

2.2 Numerical invariance to reparametrisations

Suppose we decide to work with an alternative parametrisation of the model for convenience or ease of interpretation. For example, we might decide to compare the logs of the estimators of a variance parameter rather than their levels. We can then state the following result:

[3] A related analogy arises in indirect estimation, in which the asymptotic equivalence between the score-based methods proposed by Gallant and Tauchen (1996) and the parameter-based methods in Gouriéroux, Monfort and Renault (1993) can be intuitively understood if we regard the expected values of the scores of the auxiliary model as a new set of auxiliary parameters that summarises all the information in the original parameters (see Calzolari, Fiorentini and Sentana (2004) for further details and a generalisation).

Proposition 2 *Consider a homeomorphic, continuously differentiable transformation* $\boldsymbol{\pi}(.)$ *from* $\boldsymbol{\theta}$ *to a new set of parameters* $\boldsymbol{\pi}$*, with* $\text{rank}\left[\partial\boldsymbol{\pi}'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\right]=p=\dim(\boldsymbol{\theta})$ *when evaluated at* $\boldsymbol{\theta}_0$*,* $\hat{\boldsymbol{\theta}}_T$ *and* $\tilde{\boldsymbol{\theta}}_T$*. Let* $\hat{\boldsymbol{\pi}}_T=\arg\min_{\boldsymbol{\pi}\in\boldsymbol{\Pi}}\bar{\mathfrak{m}}'_T(\boldsymbol{\pi})\tilde{\mathcal{S}}_{mT}\bar{\mathfrak{m}}_T(\boldsymbol{\pi})$ *and* $\tilde{\boldsymbol{\pi}}_T=\arg\min_{\boldsymbol{\pi}\in\boldsymbol{\Pi}}\bar{\mathfrak{n}}'_T(\boldsymbol{\pi})\tilde{\mathcal{S}}_{nT}\bar{\mathfrak{n}}_T(\boldsymbol{\pi})$*, where* $\mathfrak{m}_t(\boldsymbol{\pi})=\mathbf{m}_t[\boldsymbol{\theta}(\boldsymbol{\pi})]$ *and* $\mathfrak{n}_t(\boldsymbol{\pi})=\mathbf{n}_t[\boldsymbol{\theta}(\boldsymbol{\pi})]$ *are the influence functions written in terms of* $\boldsymbol{\pi}$*, with* $\boldsymbol{\theta}(\boldsymbol{\pi})$ *denoting the inverse mapping such that* $\boldsymbol{\pi}[\boldsymbol{\theta}(\boldsymbol{\pi})]=\boldsymbol{\pi}$*. Then,*

1. *The Wald versions of the DWH tests based on* $\tilde{\boldsymbol{\theta}}_T-\hat{\boldsymbol{\theta}}_T$ *and* $\tilde{\boldsymbol{\pi}}_T-\hat{\boldsymbol{\pi}}_T$ *are numerically identical if the mapping is affine, so that* $\boldsymbol{\pi}=\mathbf{A}\boldsymbol{\theta}+\mathbf{b}$*, with* $\mathbf{A}$ *and* $\mathbf{b}$ *known and* $|\mathbf{A}|\neq 0$*.*
2. *The score versions of the tests based on* $\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T)$ *and* $\bar{\mathfrak{m}}_T(\tilde{\boldsymbol{\pi}}_T)$ *are numerically identical if*

$$\boldsymbol{\Lambda}_{\mathfrak{m}T}^{\sim}=\left[\frac{\partial\boldsymbol{\theta}(\tilde{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}'}\right]^{-1}\boldsymbol{\Lambda}_{mt}^{\sim}\left[\frac{\partial\boldsymbol{\theta}'(\tilde{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\right]^{-1},$$

   where $\boldsymbol{\Lambda}_{mT}^{\sim}$ *and* $\boldsymbol{\Lambda}_{\mathfrak{m}T}^{\sim}$*, are consistent estimators of the generalised inverses of the limiting variances of* $\mathcal{J}'_m(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T)$ *and* $\mathcal{J}'_{\mathfrak{m}}(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}\bar{\mathfrak{m}}_T(\tilde{\boldsymbol{\pi}}_T)$*, respectively.*
3. *An analogous result applies to the score versions based on* $\bar{\mathbf{n}}_T(\hat{\boldsymbol{\theta}}_T)$ *and* $\bar{\mathfrak{n}}_T(\hat{\boldsymbol{\pi}}_T)$*.*

These numerical invariance results, which extend those in sections 17.4 and 22.1 of Ruud (2000), suggest that the score-based tests might be better behaved in finite samples than their "Wald" counterpart. We will provide some simulation evidence on this conjecture in section 4.

2.3 Subsets of parameters

In some examples, generalised inverses can be avoided by working with a parameter subvector. In particular, if the (scaled) difference between two estimators of the last p_2 elements of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_{2T}$ and $\tilde{\boldsymbol{\theta}}_{2T}$, converge in probability to 0, then comparing $\hat{\boldsymbol{\theta}}_{1T}$ and $\tilde{\boldsymbol{\theta}}_{1T}$ is analogous to using a generalised inverse with the entire parameter vector (see Holly and Monfort (1986) for further details).

But one may also want to focus on a subset if the means of the asymptotic distributions of $\hat{\boldsymbol{\theta}}_{2T}$ and $\tilde{\boldsymbol{\theta}}_{2T}$ coincide both under the null and the alternative, so that a DWH test involving these parameters will result in a waste of degrees of freedom, and thereby a loss of power.

The following result provides a useful interpretation of the two score versions asymptotically equivalent to a Wald-style DWH test that compares $\hat{\boldsymbol{\theta}}_{1T}$ and $\tilde{\boldsymbol{\theta}}_{1T}$:

Proposition 3 *Define*

$$\begin{aligned}\bar{\mathbf{m}}_{1T}^{\perp}(\boldsymbol{\theta},\mathcal{S}_n) &= \mathcal{J}'_{1m}(\boldsymbol{\theta})\mathcal{S}_m\bar{\mathbf{m}}_T(\boldsymbol{\theta})-\mathcal{J}'_{1m}(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{2m}(\boldsymbol{\theta})[\mathcal{J}'_{2m}(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{2m}(\boldsymbol{\theta})]^{-1}\mathcal{J}'_{2m}(\boldsymbol{\theta})\mathcal{S}_m\bar{\mathbf{m}}_T(\boldsymbol{\theta}),\\ \bar{\mathbf{n}}_{1T}^{\perp}(\boldsymbol{\theta},\mathcal{S}_n) &= \mathcal{J}'_{1n}(\boldsymbol{\theta})\mathcal{S}_n\bar{\mathbf{n}}_T(\boldsymbol{\theta})-\mathcal{J}'_{1n}(\boldsymbol{\theta})\mathcal{S}_n\mathcal{J}_{2n}(\boldsymbol{\theta})[\mathcal{J}'_{2n}(\boldsymbol{\theta})\mathcal{S}_n\mathcal{J}_{2n}(\boldsymbol{\theta})]^{-1}\mathcal{J}'_{2n}(\boldsymbol{\theta})\mathcal{S}_n\bar{\mathbf{n}}_T(\boldsymbol{\theta})\end{aligned}$$

as two sets of p_1 *transformed sample moment conditions, where*

$$\begin{aligned}\mathcal{J}_m(\boldsymbol{\theta}) &= \begin{bmatrix}\mathcal{J}_{1m}(\boldsymbol{\theta}) & \mathcal{J}_{2m}(\boldsymbol{\theta})\end{bmatrix}=\begin{bmatrix}\text{plim}_{T\to\infty}\partial\bar{\mathbf{m}}_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'_1 & \text{plim}_{T\to\infty}\partial\bar{\mathbf{m}}_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'_2\end{bmatrix},\\ \mathcal{J}_n(\boldsymbol{\theta}) &= \begin{bmatrix}\mathcal{J}_{1n}(\boldsymbol{\theta}) & \mathcal{J}_{2n}(\boldsymbol{\theta})\end{bmatrix}=\begin{bmatrix}\text{plim}_{T\to\infty}\partial\bar{\mathbf{n}}_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'_1 & \text{plim}_{T\to\infty}\partial\bar{\mathbf{n}}_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'_2\end{bmatrix}.\end{aligned}$$

If $\mathbf{m}_t(\boldsymbol{\theta})$ *and* $\mathbf{n}_t(\boldsymbol{\theta})$ *are correctly specified, then, under standard regularity conditions*

$$\begin{aligned}T(\tilde{\boldsymbol{\theta}}_T-\hat{\boldsymbol{\theta}}_T)'\boldsymbol{\Delta}_{11}^{-}(\tilde{\boldsymbol{\theta}}_T-\hat{\boldsymbol{\theta}}_T)-T\bar{\mathbf{m}}_T^{\perp\prime}(\tilde{\boldsymbol{\theta}}_T)\boldsymbol{\Lambda}_{\mathbf{m}_1^{\perp}}^{-}\bar{\mathbf{m}}_T^{\perp\prime}(\tilde{\boldsymbol{\theta}}_T) &= o_p(1)\\ \text{and } T(\tilde{\boldsymbol{\theta}}_{1T}-\hat{\boldsymbol{\theta}}_{1T})'\boldsymbol{\Delta}_{11}^{-}(\tilde{\boldsymbol{\theta}}_{1T}-\hat{\boldsymbol{\theta}}_{1T})-T\bar{\mathbf{n}}_{1T}^{\perp\prime}(\hat{\boldsymbol{\theta}}_T)\boldsymbol{\Lambda}_{\mathbf{n}_1^{\perp}}^{-}\bar{\mathbf{n}}_{1T}^{\perp}(\hat{\boldsymbol{\theta}}_T) &= o_p(1),\end{aligned}$$

where $\mathbf{\Delta}_{11}$, $\mathbf{\Lambda}_{\mathbf{m}_1^\perp}$ *and* $\mathbf{\Lambda}_{\mathbf{n}_1^\perp}$ *are the limiting variances of* $\sqrt{T}(\tilde{\boldsymbol{\theta}}_{1T} - \hat{\boldsymbol{\theta}}_{1T})$, $\sqrt{T}\bar{\mathbf{m}}_{1T}^\perp(\tilde{\boldsymbol{\theta}}_T, \mathcal{S}_m)$ *and* $\sqrt{T}\bar{\mathbf{n}}_{1T}^\perp(\hat{\boldsymbol{\theta}}_T, \mathcal{S}_n)$*, respectively, which are such that*

$$\begin{aligned}\mathbf{\Delta}_{11} &= \left[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right]^{11}\mathbf{\Lambda}_{\mathbf{m}_1^\perp}\left[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right]^{11} \\ &= \left[\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n\mathcal{J}_n(\boldsymbol{\theta}_0)\right]^{11}\mathbf{\Lambda}_{\mathbf{n}_1^\perp}\left[\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n\mathcal{J}_n(\boldsymbol{\theta}_0)\right]^{11},\end{aligned}$$

with ^{11} *denoting the diagonal block corresponding to* $\boldsymbol{\theta}_1$ *of the relevant inverse.*

Intuitively, we can understand $\bar{\mathbf{m}}_{1T}^\perp(\boldsymbol{\theta}, \mathcal{S}_n)$ and $\bar{\mathbf{n}}_{1T}^\perp(\boldsymbol{\theta}, \mathcal{S}_n)$ as moment conditions that exactly identify $\boldsymbol{\theta}_1$, but with the peculiarity that

$$\underset{T\to\infty}{\text{plim}}\frac{\partial\bar{\mathbf{m}}_{1T}^\perp(\boldsymbol{\theta}, \mathcal{S}_n)}{\partial\boldsymbol{\theta}_2'} = \underset{T\to\infty}{\text{plim}}\frac{\partial\bar{\mathbf{n}}_{1T}^\perp(\boldsymbol{\theta}, \mathcal{S}_n)}{\partial\boldsymbol{\theta}_2'} = \mathbf{0},$$

which makes them asymptotically immune to the sample variability in the estimators of $\boldsymbol{\theta}_2$.

When $\mathcal{J}_{1m}'(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{2m}(\boldsymbol{\theta}) = \mathcal{J}_{1n}'(\boldsymbol{\theta})\mathcal{S}_n\mathcal{J}_{2n}(\boldsymbol{\theta}) = \mathbf{0}$, the above moment tests will be asymptotically equivalent to tests based on $\mathcal{J}_{1m}'(\boldsymbol{\theta})\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T)$ and $\mathcal{J}_{1n}'(\boldsymbol{\theta})\mathcal{S}_n\sqrt{T}\bar{\mathbf{n}}_T(\hat{\boldsymbol{\theta}}_T)$, respectively, but in general this will not be the case.

2.4 Multiple simultaneous comparisons

All applications of DWH tests we are aware of compare two estimators of the same underlying parameters. However, as we shall see in section 3.2, there are situations in which three or more estimators are available. In those circumstances, it might not be entirely clear which pair of estimators researchers should focus on.

Ruud (1984) highlighted a special factorisation structure of the likelihood such that different pairwise comparisons give rise to asymptotically equivalent tests. He illustrated his result with three classical examples: (i) full sample vs first subsample vs second subsample in Chow tests; (ii) GLS vs within-groups vs between-groups in panel data; and (iii) Tobit vs probit vs truncated regressions. Unfortunately, Ruud's (1984) factorisation structure does not apply in our case.

In general, the best pairwise comparison, in the sense of having maximum power against a given sequence of local alternatives, would be the one with the highest non-centrality parameter among those tests with the same number of degrees of freedom.[4] But in practice, a researcher might not be able to make the required calculations without knowing the nature of the departure from the null. In those circumstances, a sensible solution would be to simultaneously compare all the alternative estimators. Such a generalisation of the DWH test is conceptually straightforward, but it requires the joint asymptotic distribution of the different estimators involved. There is one special case in which this simultaneous test takes a particularly simple form:

[4] Ranking tests with different degrees of freedom is also straightforward but more elaborate (see Holly (1987)).

Proposition 4 *Let* $\hat{\boldsymbol{\theta}}_T^j$, $j = 1, \ldots, J$ *denote an ordered sequence of asymptotically Gaussian estimators of* $\boldsymbol{\theta}$ *whose joint asymptotic covariance matrix adopts the following form:*

$$\begin{bmatrix} \boldsymbol{\Omega}_1 & \boldsymbol{\Omega}_1 & \ldots & \boldsymbol{\Omega}_1 & \boldsymbol{\Omega}_1 \\ \boldsymbol{\Omega}_1 & \boldsymbol{\Omega}_2 & \ldots & \boldsymbol{\Omega}_2 & \boldsymbol{\Omega}_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{\Omega}_1 & \boldsymbol{\Omega}_2 & \ldots & \boldsymbol{\Omega}_{J-1} & \boldsymbol{\Omega}_{J-1} \\ \boldsymbol{\Omega}_1 & \boldsymbol{\Omega}_2 & \ldots & \boldsymbol{\Omega}_{J-1} & \boldsymbol{\Omega}_J \end{bmatrix}. \tag{5}$$

Then, the DWH test comparing all J *estimators,* $T\sum_{i=2}^{J}(\hat{\boldsymbol{\theta}}_T^j - \hat{\boldsymbol{\theta}}_T^{j-1})'(\boldsymbol{\Omega}_j - \boldsymbol{\Omega}_{j-1})^+(\hat{\boldsymbol{\theta}}_T^j - \hat{\boldsymbol{\theta}}_T^{j-1})$, *is the sum of* $J-1$ *consecutive pairwise DWH tests that are asymptotically mutually independent under the null of correct specification and sequences of local alternatives.*

Hence, the asymptotic distribution of the simultaneous DWH test will be a non-central χ^2 with degrees of freedom and non-centrality parameters equal to the sum of the degrees of freedom and non-centrality parameters of the consecutive pairwise DWH tests. Moreover, the asymptotic independence of the tests implies that in large samples, the probability that at least one pairwise test will reject under the null will be $1-(1-\alpha)^{J-1}$, where α is the common significance level.

Positive semidefiniteness of the covariance structure in (5) implies that one can rank (in the usual positive semidefinite sense) the asymptotic variance of the J estimators as

$$\boldsymbol{\Omega}_J \geq \boldsymbol{\Omega}_{J-1} \geq \ldots \geq \boldsymbol{\Omega}_2 \geq \boldsymbol{\Omega}_1,$$

so that the sequence of estimators follows a decreasing efficiency order. Nevertheless, (5) goes beyond this ordering because it effectively implies that the estimators behave like Matryoshka dolls, with each one being "efficient" relative to all the others below. Therefore, Proposition 4 provides the natural multiple comparison generalisation of Lemma 2.1 in Hausman (1978).

An example of the covariance structure (5) arises in the context of sequential, general to specific tests of nested parametric restrictions (see Holly (1987) and section 22.6 of Ruud (2000)). More importantly for our purposes, the same structure also arises naturally in the comparison of parametric and semiparametric likelihood-based estimators of multivariate, conditionally heteroskedastic, dynamic regression models, to which we turn next.

3 Application to non-Gaussian likelihood estimators

3.1 Model specification

In a multivariate dynamic regression model with time-varying variances and covariances, the vector of N observed variables, $\mathbf{y}_t$, is typically assumed to be generated as:

$$\mathbf{y}_t = \boldsymbol{\mu}_t(\boldsymbol{\theta}) + \boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*,$$

where $\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}(I_{t-1}; \boldsymbol{\theta})$, $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(I_{t-1}; \boldsymbol{\theta})$, $\boldsymbol{\mu}()$ and $vech\,[\boldsymbol{\Sigma}()]$ are $N \times 1$ and $N(N+1)/2 \times 1$ vector functions describing the conditional mean vector and covariance matrix known up to the

$p \times 1$ vector of parameters $\boldsymbol{\theta}$, I_{t-1} denotes the information set available at $t-1$, which contains past values of $\mathbf{y}_t$ and possibly some contemporaneous conditioning variables, and $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ is some particular "square root" matrix such that $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{1/2\prime}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta})$. Throughout the paper, we maintain the assumption that the conditional mean and variance are correctly specified, in the sense that there is a true value of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_0$, such that $E(\mathbf{y}_t|I_{t-1}) = \boldsymbol{\mu}_t(\boldsymbol{\theta}_0)$ and $V(\mathbf{y}_t|I_{t-1}) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)$. We also maintain the high level regularity conditions in Bollerslev and Wooldridge (1992) because we want to leave unspecified the conditional mean vector and covariance matrix in order to achieve full generality. Primitive conditions for specific multivariate models can be found for example in Ling and McAleer (2003).

To complete the model, a researcher needs to specify the conditional distribution of $\boldsymbol{\varepsilon}_t^*$. In Supplemental Appendix D we study the general case. In view of the options that the dominant commercially available econometric software companies offer to their clients, though, in the main text we study the situation in which a researcher makes the assumption that, conditional on I_{t-1}, the distribution of $\boldsymbol{\varepsilon}_t^*$ is independent and identically distributed as some particular member of the spherical family with a well defined density, or $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\theta},\boldsymbol{\eta} \sim i.i.d.\ s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta})$ for short, where $\boldsymbol{\eta}$ denotes q additional shape parameters which effectively characterise the distribution of $\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$ (see Supplemental Appendix C.1 for a brief introduction to spherically symmetric distributions).[5] The most prominent example is the standard multivariate normal, which we denote by $\boldsymbol{\eta} = \mathbf{0}$ without loss of generality. Another important example favoured by empirical researchers is the standardised multivariate Student t with ν degrees of freedom, or $i.i.d.\ t(\mathbf{0},\mathbf{I}_N,\nu)$ for short. As is well known, the multivariate t approaches the multivariate normal as $\nu \to \infty$, but has generally fatter tails and allows for cross-sectional dependence beyond correlation. For tractability, we define η as $1/\nu$, which will always remain in the finite range $[0, 1/2)$ under our assumptions.[6] Obviously, in the univariate case, any symmetric distribution, including the GED (also known as the Generalised Gaussian distribution), is spherically symmetric too.[7]

3.2 Likelihood-based estimators

Let $L_T(\boldsymbol{\phi})$ denote the pseudo log-likelihood function of a sample of size T for the general model discussed in section 3.1, where $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\eta}')'$ are the $p+q$ parameters of interest, which we assume variation free. We consider up to five different estimators of $\boldsymbol{\theta}$:

1. **Restricted ML (RML)**: $\hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}})$, which is such that $\hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}}) = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L_T(\boldsymbol{\theta}, \bar{\boldsymbol{\eta}})$. Its

[5] Nevertheless, Propositions 10, 13, C2, D1, D2 and D3 already deal explicitly with the general case, while Propositions 5, 6, 7, 8 and 9 continue to be valid without sphericity.

[6] A Student t with $1 < \nu \leq 2$ implies an infinite variance, which is incompatible with the correct specification of $\boldsymbol{\Sigma}_t$, while the conditional mean will not even be properly defined if $\nu \leq 1$.

[7] See McDonald and Newey (1988) for a univariate generalised t distribution which nests both GED and Student t, and Gillier (2005) for a spherically symmetric multivariate version of the GED.

efficiency can be characterised by the $\boldsymbol{\theta},\boldsymbol{\theta}$ block of the information matrix, $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)$, provided that $\bar{\boldsymbol{\eta}} = \boldsymbol{\eta}_0$. Thus, we can interpret $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)$ as the restricted parametric efficiency bound.

2. **Joint or Unrestricted ML (UML)**: $\hat{\boldsymbol{\theta}}_T$, obtained as $(\hat{\boldsymbol{\theta}}_T, \hat{\boldsymbol{\eta}}_T) = \arg\max_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} L_T(\boldsymbol{\theta},\boldsymbol{\eta})$. In this case, the feasible parametric efficiency bound is $\mathcal{P}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)$.

3. **Spherically symmetric semiparametric (SSP)**: $\mathring{\boldsymbol{\theta}}_T$, which restricts $\boldsymbol{\varepsilon}^*_t$ to have an *i.i.d.* $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta})$ conditional distribution, but does not impose any additional structure on the distribution of $\varsigma_t = \boldsymbol{\varepsilon}^{*\prime}_t\boldsymbol{\varepsilon}^*_t$. This estimator is usually computed by means of one BHHH iteration of the spherically symmetric efficient score starting from a consistent estimator (see Supplemental Appendix C.5 for further computational details).[8] Associated to it we have the spherically symmetric semiparametric efficiency bound $\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)$.

4. **Unrestricted semiparametric (SP)**: $\ddot{\boldsymbol{\theta}}_T$, which only assumes that the conditional distribution of $\boldsymbol{\varepsilon}^*_t$ is $i.i.d.(\mathbf{0}, \mathbf{I}_N)$. It is also computed with one BHHH iteration of the efficient score starting from a consistent estimator (see Supplemental Appendix D.3 for further computational details). Associated to it we have the usual semiparametric efficiency bound $\ddot{\mathcal{S}}(\boldsymbol{\phi}_0)$.

5. **Gaussian Pseudo ML (PML)**: $\tilde{\boldsymbol{\theta}}_T = \hat{\boldsymbol{\theta}}_T(\mathbf{0})$, which imposes $\boldsymbol{\eta} = \mathbf{0}$ even though the true conditional distribution of $\boldsymbol{\varepsilon}^*_t$ might be neither normal nor spherical. As is well known, the efficiency bound for this estimator is given by $\mathcal{C}^{-1}(\boldsymbol{\phi}_0) = \mathcal{A}(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathcal{A}(\boldsymbol{\phi}_0)$, where $\mathcal{A}(\boldsymbol{\phi}_0)$ is the expected Gaussian Hessian and $\mathcal{B}(\boldsymbol{\phi}_0)$ the variance of the Gaussian score.

Propositions C1-C3 in Supplemental Appendix C and Proposition D3 in Supplemental Appendix D contain detailed expressions for all these efficiency bounds.

3.3 Covariance relationships

The next proposition provides the asymptotic covariance matrices of the different estimators presented in the previous section, and of the scores on which they are based:

Proposition 5 *If $\boldsymbol{\varepsilon}^*_t|I_{t-1};\boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta}_0)$ with bounded fourth moments, then*

$$\lim_{T\to\infty} V\left[\frac{\sqrt{T}}{T}\sum_{t=1}^{T}\begin{pmatrix}\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)\\ \mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\phi}_0)\\ \mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)\\ \ddot{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)\\ \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0})\end{pmatrix}\right] = \begin{bmatrix}\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{P}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}(\boldsymbol{\phi}_0) & \mathcal{A}(\boldsymbol{\phi}_0)\\ \mathcal{P}(\boldsymbol{\phi}_0) & \mathcal{P}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}(\boldsymbol{\phi}_0) & \mathcal{A}(\boldsymbol{\phi}_0)\\ \mathring{\mathcal{S}}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}(\boldsymbol{\phi}_0) & \mathcal{A}(\boldsymbol{\phi}_0)\\ \ddot{\mathcal{S}}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}(\boldsymbol{\phi}_0) & \mathcal{A}(\boldsymbol{\phi}_0)\\ \mathcal{A}(\boldsymbol{\phi}_0) & \mathcal{A}(\boldsymbol{\phi}_0) & \mathcal{A}(\boldsymbol{\phi}_0) & \mathcal{A}(\boldsymbol{\phi}_0) & \mathcal{B}(\boldsymbol{\phi}_0)\end{bmatrix}, \quad (6)$$

$$\text{and} \lim_{T\to\infty} V\left[\sqrt{T}\begin{pmatrix}\hat{\boldsymbol{\theta}}_T(\boldsymbol{\eta}_0)-\boldsymbol{\theta}_0\\ \hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\\ \mathring{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\\ \ddot{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\\ \tilde{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\end{pmatrix}\right] = \begin{bmatrix}\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\\ \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{P}^{-1}(\boldsymbol{\phi}_0) & \mathcal{P}^{-1}(\boldsymbol{\phi}_0) & \mathcal{P}^{-1}(\boldsymbol{\phi}_0) & \mathcal{P}^{-1}(\boldsymbol{\phi}_0)\\ \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{P}^{-1}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0)\\ \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{P}^{-1}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0)\\ \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{P}^{-1}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0) & \mathcal{C}(\boldsymbol{\phi}_0)\end{bmatrix} \quad (7)$$

[8] Hodgson, Linton and Vorkink (2002) also consider alternative estimators that iterate the semiparametric adjustment until it becomes negligible. However, since they have the same first-order asymptotic distribution, we shall not discuss them separately.

Therefore, the five estimators have the Matryoshka doll covariance structure in (5), with each estimator being "efficient" relative to all the others below. A trivial implication of this result is that one can unsurprisingly rank (in the usual positive semidefinite sense) the "information matrices" of those five estimators as follows:

$$\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) \geq \mathcal{P}(\boldsymbol{\phi}_0) \geq \mathring{\mathcal{S}}(\boldsymbol{\phi}_0) \geq \ddot{\mathcal{S}}(\boldsymbol{\phi}_0) \geq \mathcal{C}^{-1}(\boldsymbol{\phi}_0). \tag{8}$$

Proposition 5 remains valid when the distribution of $\boldsymbol{\varepsilon}_t^*$ conditional on I_{t-1} is not assumed spherical, provided that we cross out the terms corresponding to the SSP estimator $\mathring{\boldsymbol{\theta}}_T$ (see Supplemental Appendix D for further details). Therefore, the approach we develop in the next section can be straightforwardly extended to test the correct specification of any maximum likelihood estimator of multivariate conditionally heteroskedastic dynamic regression models. Such an extension would be important in practice because while the assumption of sphericity might be realistic for foreign exchange returns, it seems less plausible for stock returns.

3.4 Multiple simultaneous comparisons

Five estimators allow up to ten different possible pairwise comparisons, and it is not obvious which one researchers should focus on. If they only paid attention to the asymptotic covariance matrices of the differences between those ten combinations of estimators, expression (8) suggests that they should focus on adjacent estimators. However, the number of degrees of freedom and the diverging behaviour of the estimators also play a very important role.

Nevertheless, we also saw in section 2.4 that there is no reason why researchers should choose just one such pair, especially if they are agnostic about the alternative. In fact, the covariance structure in Proposition 5 combined with Proposition 4 implies that DWH tests of multiple simultaneous comparisons are extremely simple because non-overlapping pairwise comparisons give rise to asymptotically independent test statistics. Importantly, this result, combined with the fact that any of the ten possible pairwise comparisons can be obtained as the sum of the intermediate contiguous comparisons, implies that at the end of the day there are only four asymptotically independent pairwise comparisons. For example, the difference between the spherically symmetric estimator $\mathring{\boldsymbol{\theta}}_T$ and the Gaussian estimator $\tilde{\boldsymbol{\theta}}_T$ is numerically equal to the sum of the differences between each of those estimators and the general semiparametric estimator $\ddot{\boldsymbol{\theta}}_T$, so the limiting mean and covariance matrix of $\sqrt{T}(\mathring{\boldsymbol{\theta}}_T - \tilde{\boldsymbol{\theta}}_T)$ will be the sum of the limiting means and covariance matrices of $\sqrt{T}(\mathring{\boldsymbol{\theta}}_T - \ddot{\boldsymbol{\theta}}_T)$ and $\sqrt{T}(\ddot{\boldsymbol{\theta}}_T - \tilde{\boldsymbol{\theta}}_T)$. As a result, we can compute the non-centrality parameters of the DWH test based on $\mathring{\boldsymbol{\theta}}_T - \tilde{\boldsymbol{\theta}}_T$ from the same ingredients as the non-centrality parameters of the DWH tests that compare $\mathring{\boldsymbol{\theta}}_T - \ddot{\boldsymbol{\theta}}_T$ and $\ddot{\boldsymbol{\theta}}_T - \tilde{\boldsymbol{\theta}}_T$. This result also implies that the differences between adjacent asymptotic covariance matrices will often will be of reduced rank, a topic we will revisit in section 3.6.

Still, researchers may disregard $\ddot{\boldsymbol{\theta}}_T - \tilde{\boldsymbol{\theta}}_T$ because the semiparametric estimator and the Gaussian estimator are consistent for $\boldsymbol{\theta}_0$ regardless of the conditional distribution, at least as long as the *iid* assumption holds. For the same reason, they will also disregard $\mathring{\boldsymbol{\theta}}_T - \ddot{\boldsymbol{\theta}}_T$ if they maintain the assumption of sphericity. In practice, the main factor for deciding which estimators to compare is likely to be computational ease. For that reason many empirical researchers might prefer to compare only the three parametric estimators included in standard software packages even though increases in power might be obtained under the maintained assumption of *iid* innovations by comparing $\hat{\boldsymbol{\theta}}_T$ to $\mathring{\boldsymbol{\theta}}_T$ or $\ddot{\boldsymbol{\theta}}_T$ instead of $\tilde{\boldsymbol{\theta}}_T$. The next proposition provides detailed expressions for the necessary ingredients of the three DWH test statistics in (1), (3) and (4) when we compare the unrestricted ML estimator of $\boldsymbol{\theta}$ with its Gaussian PML counterpart.

Proposition 6 *If the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then under the null of correct specification of the conditional distribution of* $\mathbf{y}_t$

$$\begin{aligned}\lim_{T\to\infty} V[\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T)] &= \mathcal{C}(\boldsymbol{\phi}_0) - \mathcal{P}^{-1}(\boldsymbol{\phi}_0),\\ \lim_{T\to\infty} V[\sqrt{T}\bar{\mathbf{s}}'_{\boldsymbol{\theta}|\boldsymbol{\eta}T}(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta}_0)] &= \mathcal{P}(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{P}(\boldsymbol{\phi}_0) - \mathcal{P}(\boldsymbol{\phi}_0) \text{ and}\\ \lim_{T\to\infty} V[\sqrt{T}\bar{\mathbf{s}}'_{\boldsymbol{\theta}T}(\hat{\boldsymbol{\theta}}_T, \mathbf{0})] &= \mathcal{B}(\boldsymbol{\phi}_0) - \mathcal{A}(\boldsymbol{\phi}_0)\mathcal{P}^{-1}(\boldsymbol{\phi}_0)\mathcal{A}(\boldsymbol{\phi}_0),\end{aligned}$$

where $\bar{\mathbf{s}}_{\boldsymbol{\theta}|\boldsymbol{\eta}T}(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta}_0)$ *is the sample average of the unrestricted parametric efficient score for* $\boldsymbol{\theta}$ *evaluated at the Gaussian PML estimator* $\tilde{\boldsymbol{\theta}}_T$*, while* $\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\hat{\boldsymbol{\theta}}_T, \mathbf{0})$ *is the sample average of the Gaussian PML score evaluated at the unrestricted parametric ML estimator* $\hat{\boldsymbol{\theta}}_T$*.*

The next proposition provides the analogous expressions for the three DWH test statistics in (1), (3) and (4) when we compare the restricted ML estimator of $\boldsymbol{\theta}$ which fixes $\boldsymbol{\eta}$ to $\bar{\boldsymbol{\eta}}$ with its unrestricted counterpart, which simultaneously estimates these parameters.

Proposition 7 *If the regularity conditions in Crowder (1976) are satisfied, then under the null of correct specification of the conditional distribution of* $\mathbf{y}_t$

$$\lim_{T\to\infty} V\{\sqrt{T}[\hat{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}})]\} = \mathcal{P}^{-1}(\boldsymbol{\phi}_0) - \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) = \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0),$$

$$\lim_{T\to\infty} V[\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\hat{\boldsymbol{\theta}}_T, \bar{\boldsymbol{\eta}})] = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathcal{P}^{-1}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0 \text{ and}$$

$$\begin{aligned}\lim_{T\to\infty} V\{\sqrt{T}\bar{\mathbf{s}}'_{\boldsymbol{\theta}|\boldsymbol{\eta}T}[\hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}}), \bar{\boldsymbol{\eta}}]\} &= \mathcal{P}(\boldsymbol{\phi}_0) - \mathcal{P}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathcal{P}(\boldsymbol{\phi}_0)\\ &= \mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\end{aligned}$$

where $\mathcal{I}^{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0) = [\mathcal{I}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0) - \mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)]^{-1}$, $\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\hat{\boldsymbol{\theta}}_T, \bar{\boldsymbol{\eta}})$ *is the sample average of the restricted parametric score evaluated at the unrestricted parametric ML estimator* $\hat{\boldsymbol{\theta}}_T$ *and* $\bar{\mathbf{s}}_{\boldsymbol{\theta}|\boldsymbol{\eta}T}(\tilde{\boldsymbol{\theta}}_T, \bar{\boldsymbol{\eta}})$ *is the sample average of the unrestricted parametric efficient score for* $\boldsymbol{\theta}$ *evaluated at the restricted parametric ML estimator* $\hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}})$*.*

The comparison between the unrestricted and restricted parametric estimators of $\boldsymbol{\theta}$ can be regarded as a test of $H_0 : \boldsymbol{\eta} = \bar{\boldsymbol{\eta}}$. However, it is not necessarily asymptotically equivalent to the Wald, LR and LM of the same hypothesis. In fact, a straightforward application of the results in Holly (1982) implies that these four tests will be equivalent if and only if $\text{rank}[\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)] = q = \dim(\boldsymbol{\eta})$, in which case we can show that the LM test and the $\bar{\mathbf{s}}_{\boldsymbol{\theta}|\boldsymbol{\eta}T}[\hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}}), \bar{\boldsymbol{\eta}}]$ version of our DWH test numerically coincide. But Proposition C1 in Supplemental Appendix C implies that in the spherically symmetric case $\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0) = \mathbf{W}_s(\boldsymbol{\phi}_0)\text{M}_{sr}(\boldsymbol{\eta}_0)$, where $\mathbf{W}_s(\boldsymbol{\phi}_0)$ in (C28) is $p \times 1$ and $\text{M}_{sr}(\boldsymbol{\eta}_0)$ in (C18) is $1 \times q$, which in turn implies that $\text{rank}[\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)]$ is one at most. Intuitively, the reason is that the dependence between the conditional mean and variance parameters $\boldsymbol{\theta}$ and the shape parameters $\boldsymbol{\eta}$ effectively hinges on a single parameter in the spherically symmetric case, as explained in Amengual, Fiorentini and Sentana (2013). Therefore, this pairwise DWH test can only be asymptotically equivalent to the classical tests of $H_0 : \boldsymbol{\eta} = \bar{\boldsymbol{\eta}}$ when $q = 1$ and $\text{M}_{sr}(\boldsymbol{\eta}_0) \neq \mathbf{0}$, the Student t with finite degrees of freedom constituting an important example.

More generally, the asymptotic distribution of the DWH test under a sequences of local alternatives for which $\boldsymbol{\eta}_{0T} = \bar{\boldsymbol{\eta}} + \tilde{\boldsymbol{\eta}}/\sqrt{T}$ will be a non-central chi-square with $\text{rank}[\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)]$ degrees of freedom and non-centrality parameter

$$\tilde{\boldsymbol{\eta}}'\mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)[\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)]^{-}\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\tilde{\boldsymbol{\eta}}, \tag{9}$$

while the asymptotic distribution of the trinity of classical tests will be a non-central distribution with q degrees of freedom and non-centrality parameter

$$\tilde{\boldsymbol{\eta}}'[\mathcal{I}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0) - \mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)]^{-1}\tilde{\boldsymbol{\eta}}.$$

Therefore, the DWH test will have power equal to size in those directions in which $\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\tilde{\boldsymbol{\eta}} = \mathbf{0}$ but more power than the classical tests in some others (see Hausman and Taylor (1981), Holly (1982) and Davidson and MacKinnon (1989) for further discussion). For analogous reasons, it will be consistent for fixed alternatives $H_f : \boldsymbol{\eta} = \bar{\boldsymbol{\eta}} + \tilde{\boldsymbol{\eta}}$ with $\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\tilde{\boldsymbol{\eta}} \neq \mathbf{0}$.

3.5 Subsets of parameters

As in section 2.3, we may be interested in focusing on a parameter subset either to avoid generalised inverses or to increase power. In fact, we show in sections 3.6 and 3.7 that both motivations apply in our context. The next proposition provides detailed expressions for the different ingredients of the DWH test statistics in Proposition 3 when we compare the unrestricted ML estimator of a subset of the parameter vector with its Gaussian PML counterpart.

Proposition 8 *If the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satis-*

fied, then under the null of correct specification of the conditional distribution of $\mathbf{y}_t$

$$\lim_{T\to\infty} V[\sqrt{T}(\tilde{\boldsymbol{\theta}}_{1T}-\hat{\boldsymbol{\theta}}_{1T})] = \mathcal{C}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)-\mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0),$$

$$\lim_{T\to\infty} V[\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2\boldsymbol{\eta}T}(\tilde{\boldsymbol{\theta}}_T,\boldsymbol{\eta}_0)] = [\mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)]^{-1}\mathcal{C}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)[\mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)]^{-1}-[\mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)]^{-1} \text{ and}$$

$$\lim_{T\to\infty} V[\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2T}(\hat{\boldsymbol{\theta}}_T,\mathbf{0})] = [\mathcal{A}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)]^{-1}[\mathcal{C}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)-\mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)][\mathcal{A}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)]^{-1}, \text{ where}$$

$$\bar{\mathbf{s}}_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2\boldsymbol{\eta}T}(\boldsymbol{\theta},\boldsymbol{\eta})=\bar{\mathbf{s}}_{\boldsymbol{\theta}_1T}(\boldsymbol{\theta},\boldsymbol{\eta})-\left[\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\,\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\right]\begin{bmatrix}\mathcal{I}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\theta}_2\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\\ \mathcal{I}'_{\boldsymbol{\theta}_2\boldsymbol{\eta}}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\end{bmatrix}^{-1}\begin{bmatrix}\bar{\mathbf{s}}_{\boldsymbol{\theta}_2T}(\boldsymbol{\theta},\boldsymbol{\eta})\\ \bar{\mathbf{s}}_{\boldsymbol{\eta}T}(\boldsymbol{\theta},\boldsymbol{\eta})\end{bmatrix}, \tag{10}$$

$$\mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)=\left\{\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)-\left[\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\,\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\right]\begin{bmatrix}\mathcal{I}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\theta}_2\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\\ \mathcal{I}'_{\boldsymbol{\theta}_2\boldsymbol{\eta}}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\end{bmatrix}^{-1}\begin{bmatrix}\mathcal{I}'_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\\ \mathcal{I}'_{\boldsymbol{\theta}_1\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\end{bmatrix}\right\}^{-1}, \text{while}$$

$$\bar{\mathbf{s}}_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2T}(\boldsymbol{\theta},\mathbf{0}) = \bar{\mathbf{s}}_{\boldsymbol{\theta}_1T}(\boldsymbol{\theta},\mathbf{0})-\mathcal{A}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\mathcal{A}^{-1}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\bar{\mathbf{s}}_{\boldsymbol{\theta}_2T}(\boldsymbol{\theta},\mathbf{0}), \text{ and}$$

$$\mathcal{A}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0) = [\mathcal{A}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)-\mathcal{A}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\mathcal{A}^{-1}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\mathcal{A}'_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)]^{-1}.$$

The analogous result for the comparison between the unrestricted and restricted ML estimator of a subset of the parameter vector is as follows:

Proposition 9 *If the regularity conditions in Crowder (1976) are satisfied, then under the null of correct specification of the conditional distribution of* $\mathbf{y}_t$

$$\begin{aligned}
\lim_{T\to\infty} V\{\sqrt{T}[\hat{\boldsymbol{\theta}}_{1T}-\hat{\boldsymbol{\theta}}_{1T}(\bar{\boldsymbol{\eta}})]\} &= \mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)-\mathcal{I}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0),\\
\lim_{T\to\infty} V[\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2T}(\hat{\boldsymbol{\theta}}_T,\bar{\boldsymbol{\eta}})] &= [\mathcal{I}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)]^{-1}\mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)[\mathcal{I}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)]^{-1}-[\mathcal{I}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)]^{-1} \text{ and}\\
\lim_{T\to\infty} V\{\sqrt{T}\bar{\mathbf{s}}'_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2\boldsymbol{\eta}T}[\hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}}),\bar{\boldsymbol{\eta}}]\} &= [\mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)]^{-1}-[\mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)]^{-1}\mathcal{I}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)[\mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)]^{-1},
\end{aligned}$$

where $\bar{\mathbf{s}}_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2\boldsymbol{\eta}T}(\boldsymbol{\theta},\boldsymbol{\eta})$ *is defined in (10),*

$$\begin{aligned}
\bar{\mathbf{s}}_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2T}(\boldsymbol{\theta},\bar{\boldsymbol{\eta}}) &= \bar{\mathbf{s}}_{\boldsymbol{\theta}_1T}(\boldsymbol{\theta},\bar{\boldsymbol{\eta}})-\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\bar{\mathbf{s}}_{\boldsymbol{\theta}_2T}(\boldsymbol{\theta},\bar{\boldsymbol{\eta}}), \text{ and}\\
\mathcal{I}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0) &= [\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)-\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)]^{-1}.
\end{aligned}$$

In practice, we must replace $\mathcal{A}(\boldsymbol{\phi}_0)$, $\mathcal{B}(\boldsymbol{\phi}_0)$ and $\mathcal{I}(\boldsymbol{\phi}_0)$ by consistent estimators to make all the above tests operational. To guarantee the positive semidefiniteness of their weighting matrices, we will follow Ruud's (1984) suggestion and estimate all those matrices as sample averages of the corresponding conditional expressions in Propositions C1 and C2 in Supplemental Appendix C evaluated at a common estimator of $\boldsymbol{\phi}$, such as the restricted MLE $[\hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}}),\bar{\boldsymbol{\eta}}]$, its unrestricted counterpart $\hat{\boldsymbol{\phi}}_T$, or the Gaussian PML $\tilde{\boldsymbol{\theta}}_T$ coupled with the sequential ML or method of moments estimators of $\boldsymbol{\eta}$ in Amengual, Fiorentini and Sentana (2013), the latter being such that $\mathcal{B}(\boldsymbol{\theta},\boldsymbol{\eta})$ remains bounded.[9] In addition, in computing the three versions of the tests we exploit the theoretical relationships between the relevant asymptotic covariance matrices in Propositions 8 and 9 so that the required generalised inverses are internally coherent.

[9] Unfortunately, DWH tests that involve the Gaussian PMLE will not work properly with unbounded fourth moments, which violates one of the assumptions of Proposition C2 in Supplemental Appendix C.

In what follows, we will simplify the presentation by concentrating on Wald version of DWH tests in (1), but all our results can be readily applied to their two asymptotically equivalent score versions in (3) and (4) by virtue of Proposition 1, and the same applies to Proposition 3.

3.6 Choosing the correct number of degrees of freedom

Propositions 6 and 7 establish the asymptotic variances involved in the calculation of simultaneous DWH tests, but they do not determine the correct number of degrees of freedom that researchers should use. In fact, there are cases in which two or more estimators are equally efficient for all the parameters, and one instance in which this is true for all five estimators:[10]

Proposition 10 *1. If $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}_0$ is i.i.d. $N(\mathbf{0},\mathbf{I}_N)$, then*

$$\mathcal{I}_t(\boldsymbol{\theta}_0,\mathbf{0}) = V\left[\mathbf{s}_t(\boldsymbol{\theta}_0,\mathbf{0})|I_{t-1};\boldsymbol{\theta}_0,\mathbf{0}\right] = \begin{bmatrix} V\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0})|I_{t-1};\boldsymbol{\theta}_0,\mathbf{0}\right] & \mathbf{0} \\ \mathbf{0}' & \mathcal{M}_{rr}(\mathbf{0}) \end{bmatrix}, \text{ where}$$

$$V\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0})|I_{t-1};\boldsymbol{\theta}_0,\mathbf{0}\right] = -E\left[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta}_0,\mathbf{0})|I_{t-1};\boldsymbol{\theta}_0,\mathbf{0}\right] = \mathcal{A}_t(\boldsymbol{\theta}_0,\mathbf{0}) = \mathcal{B}_t(\boldsymbol{\theta}_0,\mathbf{0}).$$

2. If $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta}_0)$ with $\kappa_0 = E(\varsigma_t^2)/[N(N+2)] - 1 < \infty$, and $\mathbf{Z}_l(\boldsymbol{\phi}_0) = E[\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0] \neq \mathbf{0}$, where $\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)$ is defined in (C6), then $\ddot{\mathcal{S}}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)$ only if $\boldsymbol{\eta}_0 = \mathbf{0}$.

The first part of this proposition, which generalises Proposition 2 in Fiorentini, Sentana and Calzolari (2003), implies that $\hat{\boldsymbol{\theta}}_T$ suffers no asymptotic efficiency loss from simultaneously estimating $\boldsymbol{\eta}$ when $\boldsymbol{\eta}_0 = \mathbf{0}$. In turn, the second part, which generalises Result 2 in Gonzalez-Rivera and Drost (1999) and Proposition 6 in Hafner and Rombouts (2007), implies that normality is the only such instance within the spherical family.

For practical purposes, this result implies that a researcher who assumes multivariate normality cannot use DWH tests to assess distributional misspecification. But it also indicates that if she has specified instead a non-Gaussian distribution that nest the multivariate normal, she should not use those tests either if she suspects the true distribution may be Gaussian because the asymptotic distribution of the statistics will not be uniform. Unfortunately, one cannot always detect this problem by looking at $\hat{\boldsymbol{\eta}}_T$. For example, Fiorentini, Sentana and Calzolari (2003) prove that under normality, the ML estimator of the reciprocal of degrees of freedom of a multivariate Student t will be 0 approximately half the time only. In many empirical applications, though, normality is unlikely to be a practical concern.

There are other distributions for which some but not all of the differences will be 0:

Proposition 11 *1. If $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta}_0)$ with $-2/(N+2) < \kappa_0 < \infty$, and $\mathbf{W}_s(\boldsymbol{\phi}_0) \neq \mathbf{0}$, then $\mathring{\mathcal{S}}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)$ only if $\varsigma_t|I_{t-1};\boldsymbol{\phi}_0$ is i.i.d. Gamma with mean N and variance $N[(N+2)\kappa_0 + 2]$.*

2. If $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta}_0)$ and $\mathbf{W}_s(\boldsymbol{\phi}_0) \neq \mathbf{0}$, $\mathcal{P}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)$ only if $\mathrm{M}_{sr}(\boldsymbol{\eta}_0) = \mathbf{0}$.

[10] As we mentioned before, the restricted ML estimator $\hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}})$ is efficient provided that $\bar{\boldsymbol{\eta}} = \boldsymbol{\eta}_0$, which in this case requires that the researcher must correctly impose normality.

The first part of this proposition, which generalises the univariate results in Gonzalez-Rivera (1997), implies that the SSP estimator $\mathring{\boldsymbol{\theta}}_T$ can be fully efficient only if $\boldsymbol{\varepsilon}_t^*$ has a conditional Kotz distribution (see Kotz (1975)). This distribution nests the multivariate normal for $\kappa = 0$, but it can also be either platykurtic ($\kappa < 0$) or leptokurtic ($\kappa > 0$). Although such a nesting provides an analytically convenient generalisation of the multivariate normal that gives rise to some interesting theoretical results,[11] the density of a leptokurtic Kotz distribution has a pole at 0, which is a potential drawback from an empirical point of view.

In turn, the second part provides the necessary and sufficient condition for the information matrix to be block diagonal between the mean and variance parameters $\boldsymbol{\theta}$ on the one hand and the shape parameters $\boldsymbol{\eta}$ on the other. Although the lack of uniformity that we mentioned after Proposition 10 applies to this proposition too, its practical consequences would only become a real problem in the unlikely event that a researcher used a parametric spherical distribution for which $\text{M}_{rs} \neq 0$ in general, but which is such that $\text{M}_{rs} = 0$ in some special case. We are not aware of any non-Gaussian elliptical distribution with this property, although it might exist.[12]

There are also other more subtle but far more pervasive situations in which some, but not all elements of $\boldsymbol{\theta}$ can be estimated as efficiently as if $\boldsymbol{\eta}_0$ were known (see also Lange, Little and Taylor (1989)), a fact that would be described in the semiparametric literature as partial adaptivity. Effectively, this requires that some elements of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$ be orthogonal to the relevant tangent set after partialling out the effects of the remaining elements of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$ by regressing the former on the latter. Partial adaptivity, though, often depends on the model parametrisation. The following reparametrisation provides a general sufficient condition in multivariate dynamic models under sphericity:

Reparametrisation 1 *A homeomorphic transformation* $\mathbf{r}_s(.) = [\mathbf{r}'_{sc}(.), r'_{si}(.)]'$ *of the mean-variance parameters* $\boldsymbol{\theta}$ *into an alternative set* $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}'_c, \vartheta'_i)'$*, where* ϑ_i *is a positive scalar, and* $\mathbf{r}_s(\boldsymbol{\theta})$ *is twice continuously differentiable with* $\text{rank}\left[\partial \mathbf{r}'_s(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}\right] = p$ *in a neighbourhood of* $\boldsymbol{\theta}_0$*, such that*

$$\left.\begin{array}{r} \boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}_t(\boldsymbol{\vartheta}_c), \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \vartheta_i \boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_c) \end{array}\right\} \quad \forall t. \tag{11}$$

Expression (11) simply requires that one can construct pseudo-standardised residuals

$$\boldsymbol{\varepsilon}_t^\circ(\boldsymbol{\vartheta}_c) = \boldsymbol{\Sigma}_t^{\circ-1/2}(\boldsymbol{\vartheta}_c)[\mathbf{y}_t - \boldsymbol{\mu}_t^\circ(\boldsymbol{\vartheta}_c)]$$

which are *i.i.d.* $s(\mathbf{0}, \vartheta_i \mathbf{I}_N, \boldsymbol{\eta})$, where ϑ_i is a global scale parameter, a condition satisfied by most static and dynamic models.

[11] For example, we show in the proof of Proposition 10 that $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) = \ddot{\mathcal{S}}(\boldsymbol{\phi})$ in univariate models with Kotz innovations in which the conditional mean is correctly specified to be 0. In turn, Francq and Zakoïan (2010) show that $\mathcal{I}_{\boldsymbol{\theta\theta}}^{-1}(\boldsymbol{\phi}) = \mathcal{C}(\boldsymbol{\phi})$ in those models under exactly the same assumptions.

[12] Fiorentini and Sentana (2019) provides a very different reason for the DWH test considered in Proposition 6 to be degenerate. Specifically, Proposition 5 in that paper implies that if one uses a Student t log-likelihood function for estimating $\boldsymbol{\theta}$ but the true distribution is such that $\kappa < 0$, then $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) = o_p(1)$.

The next proposition generalises and extends earlier results by Bickel (1982), Linton (1993), Drost, Klaassen and Werker (1997) and Hodgson and Vorkink (2003):

Proposition 12 *1. If $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}$ is i.i.d. $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta})$ and (11) holds, then:*

(a) the spherically symmetric semiparametric estimator of $\boldsymbol{\vartheta}_c$ is ϑ_i-adaptive,

(b) If $\mathring{\boldsymbol{\vartheta}}_T$ denotes the iterated spherically symmetric semiparametric estimator of $\boldsymbol{\vartheta}$, then $\mathring{\vartheta}_{iT}=\vartheta_{iT}(\mathring{\boldsymbol{\vartheta}}_{cT})$, where

$$\vartheta_{iT}(\boldsymbol{\vartheta}_c) = (NT)^{-1}\sum_{t=1}^T \varsigma_t^{\circ}(\boldsymbol{\vartheta}_c), \tag{12}$$

$$\varsigma_t^{\circ}(\boldsymbol{\vartheta}_c) = [\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\vartheta}_c)]'\boldsymbol{\Sigma}_t^{\circ-1}(\boldsymbol{\vartheta}_c)[\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\vartheta}_c)], \tag{13}$$

(c) $\text{rank}[\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)-\mathcal{C}^{-1}(\boldsymbol{\phi}_0)]\le \dim(\boldsymbol{\vartheta}_c)=p-1.$

2. If in addition $E[\ln|\boldsymbol{\Sigma}_t^{\circ}(\boldsymbol{\vartheta}_c)||\boldsymbol{\phi}_0]=k\ \forall\boldsymbol{\vartheta}_c$ holds, then:

(a) $\mathcal{I}_{\boldsymbol{\vartheta}\boldsymbol{\vartheta}}(\boldsymbol{\phi}_0),\mathcal{P}(\boldsymbol{\phi}_0),\mathring{\mathcal{S}}(\boldsymbol{\phi}_0),\ddot{\mathcal{S}}(\boldsymbol{\phi}_0)$ and $\mathcal{C}(\boldsymbol{\phi}_0)$ are block-diagonal between $\boldsymbol{\vartheta}_c$ and ϑ_i.

(b) $\sqrt{T}(\mathring{\vartheta}_{iT}-\tilde{\vartheta}_{iT})=o_p(1)$, where $\tilde{\boldsymbol{\vartheta}}_T'=(\tilde{\boldsymbol{\vartheta}}_{cT}',\tilde{\vartheta}_{iT})$ is the Gaussian PMLE of $\boldsymbol{\vartheta}$, with $\tilde{\vartheta}_{iT}=\vartheta_{iT}(\tilde{\boldsymbol{\vartheta}}_{cT})$.

This proposition provides a saddle point characterisation of the asymptotic efficiency of the SSP estimator of $\boldsymbol{\vartheta}$, in the sense that in principle it can estimate $p-1$ parameters as efficiently as if we fully knew the true conditional distribution of the data, including its shape parameters, while for the remaining scalar parameter it only achieves the efficiency of the Gaussian PMLE.

The main implication of Proposition 12 for our proposed tests is that while the maximum rank of the asymptotic variance of $\sqrt{T}(\tilde{\boldsymbol{\vartheta}}_T-\mathring{\boldsymbol{\vartheta}}_T)$ will be $p-1$, the asymptotic variances of $\sqrt{T}[\hat{\boldsymbol{\vartheta}}_T-\hat{\boldsymbol{\vartheta}}_T(\bar{\boldsymbol{\eta}})]$, $\sqrt{T}(\mathring{\boldsymbol{\vartheta}}_T-\hat{\boldsymbol{\vartheta}}_T)$ and indeed $\sqrt{T}[\mathring{\boldsymbol{\vartheta}}_T-\hat{\boldsymbol{\vartheta}}_T(\bar{\boldsymbol{\eta}})]$ will have rank one at most. In fact, we can show that once we exploit the rank deficiency of the relevant matrices in the calculation of generalised inverses, the DWH tests based on $\sqrt{T}(\tilde{\boldsymbol{\vartheta}}_{cT}-\mathring{\boldsymbol{\vartheta}}_{cT})$, $\sqrt{T}[\hat{\vartheta}_{iT}-\hat{\vartheta}_{iT}(\bar{\boldsymbol{\eta}})]$, $\sqrt{T}(\mathring{\vartheta}_{iT}-\hat{\vartheta}_{iT})$ and $\sqrt{T}[\mathring{\vartheta}_{iT}-\hat{\vartheta}_{iT}(\bar{\boldsymbol{\eta}})]$ coincide with the analogous tests for the entire vector $\boldsymbol{\vartheta}$, which in turn are asymptotically equivalent to tests that look at the original parameters $\boldsymbol{\theta}$.

It is also possible to find an analogous result for the SP estimator, but at the cost of restricting further the set of parameters that can be estimated in a partially adaptive manner:

Reparametrisation 2 *A homeomorphic transformation $\mathbf{r}_g(.)=[\mathbf{r}_{gc}'(.),\mathbf{r}_{gim}'(.),\mathbf{r}_{gic}'(.)]'$ of the mean-variance parameters $\boldsymbol{\theta}$ into an alternative set $\boldsymbol{\varphi}=(\boldsymbol{\varphi}_c',\boldsymbol{\varphi}_{im}',\boldsymbol{\varphi}_{ic'}')'$, where $\boldsymbol{\varphi}_{im}$ is $N\times 1$, $\boldsymbol{\varphi}_{ic}=vech(\boldsymbol{\Phi}_{ic})$, $\boldsymbol{\Phi}_{ic}$ is an unrestricted positive definite symmetric matrix of order N and $\mathbf{r}_g(\boldsymbol{\theta})$ is twice continuously differentiable in a neighbourhood of $\boldsymbol{\theta}_0$ with* $\text{rank}\left[\partial\mathbf{r}_g'(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\right]=p$, *such that*

$$\left.\begin{array}{c}\boldsymbol{\mu}_t(\boldsymbol{\theta})=\boldsymbol{\mu}_t^{\diamond}(\boldsymbol{\varphi}_c)+\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\varphi}_c)\boldsymbol{\varphi}_{im}\\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta})=\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\varphi}_c)\boldsymbol{\Phi}_{ic}\boldsymbol{\Sigma}_t^{\diamond 1/2\prime}(\boldsymbol{\varphi}_c)\end{array}\right\}\quad \forall t. \tag{14}$$

This parametrisations simply requires the pseudo-standardised residuals

$$\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\varphi}_c)=\boldsymbol{\Sigma}_t^{\diamond -1/2}(\boldsymbol{\varphi}_c)[\mathbf{y}_t-\boldsymbol{\mu}_t^{\diamond}(\boldsymbol{\varphi}_c)] \tag{15}$$

to be *i.i.d.* with mean vector $\boldsymbol{\varphi}_{im}$ and covariance matrix $\boldsymbol{\Phi}_{ic}$.

The next proposition generalises and extends Theorems 3.1 in Drost and Klaassen (1997) and 3.2 in Sun and Stengos (2006):

Proposition 13 *1. If* $\varepsilon_t^*|I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}$ *is i.i.d.* $D(\mathbf{0},\mathbf{I}_N,\boldsymbol{\varrho})$*, and (14) holds, then*

(a) the semiparametric estimator of $\boldsymbol{\varphi}_c$, $\ddot{\boldsymbol{\varphi}}_{cT}$*, is* $\boldsymbol{\varphi}_i$*-adaptive, where* $\boldsymbol{\varphi}_i=(\boldsymbol{\varphi}'_{im},\boldsymbol{\varphi}'_{ic})'$.

(b) If $\ddot{\boldsymbol{\varphi}}_T$ *denotes the iterated semiparametric estimator of* $\boldsymbol{\varphi}$*, then* $\ddot{\boldsymbol{\varphi}}_{imT}=\boldsymbol{\varphi}_{imT}(\ddot{\boldsymbol{\varphi}}_{cT})$ *and* $\ddot{\boldsymbol{\varphi}}_{icT}=\boldsymbol{\varphi}_{icT}(\ddot{\boldsymbol{\varphi}}_{cT})$*, where*

$$\boldsymbol{\varphi}_{imT}(\boldsymbol{\varphi}_c) = T^{-1}\sum_{t=1}^T \boldsymbol{\varepsilon}_t^\diamond(\boldsymbol{\varphi}_c), \tag{16}$$

$$\boldsymbol{\varphi}_{icT}(\boldsymbol{\varphi}_c) = T^{-1}\sum_{t=1}^T vech\{[\boldsymbol{\varepsilon}_t^\diamond(\boldsymbol{\varphi}_c)-\boldsymbol{\varphi}_{imT}(\boldsymbol{\varphi}_c)][\boldsymbol{\varepsilon}_t^\diamond(\boldsymbol{\varphi}_c)-\boldsymbol{\varphi}_{imT}(\boldsymbol{\varphi}_c)]'\}. \tag{17}$$

(c) $\text{rank}[\ddot{\mathcal{S}}(\boldsymbol{\phi}_0)-\mathcal{C}^{-1}(\boldsymbol{\phi}_0)]\leq \dim(\boldsymbol{\varphi}_c)=p-N(N+3)/2.$

2. If in addition $E[\partial\boldsymbol{\mu}_t^{\diamond\prime}(\boldsymbol{\varphi}_{c0})/\partial\boldsymbol{\varphi}_c\cdot\boldsymbol{\Sigma}_t^{\diamond-1/2}(\boldsymbol{\varphi}_{c0})\big|\,\boldsymbol{\phi}_0]=\mathbf{0}$ *and* $E\{\partial vec[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\varphi}_{c0})]/\partial\boldsymbol{\varphi}_c\cdot[\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{\diamond-1/2\prime}(\boldsymbol{\varphi}_{c0})]\big|\,\boldsymbol{\phi}_0\}=\mathbf{0}$*, then*

(a) $\mathcal{I}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}(\boldsymbol{\phi}_0),\mathcal{P}(\boldsymbol{\phi}_0),\ddot{\mathcal{S}}(\boldsymbol{\phi}_0)$ *and* $\mathcal{C}(\boldsymbol{\phi}_0)$ *are block diagonal between* $\boldsymbol{\varphi}_c$ *and* $\boldsymbol{\varphi}_i$.

(b) $\sqrt{T}(\tilde{\boldsymbol{\varphi}}_{iT}-\ddot{\boldsymbol{\varphi}}_{iT})=o_p(1)$*, where* $\tilde{\boldsymbol{\varphi}}'_T=(\tilde{\boldsymbol{\varphi}}'_{cT},\tilde{\boldsymbol{\varphi}}'_{iT})$ *is the Gaussian PMLE of* $\boldsymbol{\varphi}$*, with* $\tilde{\boldsymbol{\varphi}}_{imT}=\boldsymbol{\varphi}_{imT}(\tilde{\boldsymbol{\varphi}}'_{cT})$ *and* $\tilde{\boldsymbol{\varphi}}_{icT}=\boldsymbol{\varphi}_{icT}(\tilde{\boldsymbol{\varphi}}'_{cT})$.

This proposition provides a saddle point characterisation of the asymptotic efficiency of the semiparametric estimator of $\boldsymbol{\theta}$, in the sense that in principle it can estimate $p-N(N+3)/2$ parameters as efficiently as if we fully knew the true conditional distribution of the data, while for the remaining parameters it only achieves the efficiency of the Gaussian PMLE.

The main implication of Proposition 13 for our purposes is that while the DWH test based on $\sqrt{T}(\tilde{\boldsymbol{\varphi}}_T-\ddot{\boldsymbol{\varphi}}_T)$ will have a maximum of $p-N(N+3)/2$ degrees of freedom, those based on $\sqrt{T}[\tilde{\boldsymbol{\varphi}}_T-\tilde{\boldsymbol{\varphi}}_T(\bar{\boldsymbol{\eta}})]$, $\sqrt{T}(\ddot{\boldsymbol{\varphi}}_T-\tilde{\boldsymbol{\varphi}}_T)$ and $\sqrt{T}[\ddot{\boldsymbol{\varphi}}_T-\tilde{\boldsymbol{\varphi}}_T(\bar{\boldsymbol{\eta}})]$ will have $N(N+3)/2$ at most. As before, we can show that once we exploit the rank deficiency of the relevant matrices in the calculation of generalised inverses, DWH tests based on $\sqrt{T}(\tilde{\boldsymbol{\varphi}}_{cT}-\ddot{\boldsymbol{\varphi}}_{cT})$, $\sqrt{T}[\tilde{\boldsymbol{\varphi}}_{iT}-\tilde{\boldsymbol{\varphi}}_{iT}(\bar{\boldsymbol{\eta}})]$, $\sqrt{T}(\ddot{\boldsymbol{\varphi}}_{iT}-\tilde{\boldsymbol{\varphi}}_{iT})$ and $\sqrt{T}[\ddot{\boldsymbol{\varphi}}_{iT}-\tilde{\boldsymbol{\varphi}}_{iT}(\bar{\boldsymbol{\eta}})]$ are identical to the analogous tests based on the entire vector $\tilde{\boldsymbol{\varphi}}$, which in turn are asymptotically equivalent to tests that look at the original parameters $\boldsymbol{\theta}$.

3.7 Maximising power

As we discussed in section 2.1, the local power of a pairwise DWH test depends on the difference in the pseudo-true values of the parameters under misspecification relative to the difference between the covariance matrices under the null. But Proposition 1 in Fiorentini and Sentana (2019) states that in the situation discussed in Proposition 12, $\boldsymbol{\vartheta}_c$ will be consistently estimated when the true distribution of the innovations is spherical but different from the one assumed for estimation purposes, while ϑ_i will be inconsistently estimated. Therefore, rather than losing power by disregarding all the elements of $\boldsymbol{\vartheta}_c$, we will in fact maximise power if we base our DWH tests on the overall scale parameter ϑ_i exclusively. Similarly, Proposition 3 in Fiorentini and Sentana (2019) states that in the context of Proposition 13, $\boldsymbol{\varphi}_c$ will be consistently estimated when the true distribution of the innovations is *i.i.d.* but different from the one assumed for estimation purposes, while $\boldsymbol{\varphi}_{im}$ and $\boldsymbol{\varphi}_{ic}$ will be inconsistently estimated.

Consequently, we will maximise power in that case if we base our DWH tests on the mean and covariance parameters of the pseudo standardised residuals $\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\varphi}_c)$ in (15).

3.8 Extensions to structural models

So far we have considered multivariate dynamic location scale models which directly parametrise the conditional first and second moment functions. However, non-Gaussian innovations have also become increasing popular in dynamic structural models, whose focus differs from those conditional moments. Two important examples are non-causal univariate ARMA models (see Supplemental Appendix E.2) and structural vector autoregressions (SVARs), like the one we consider in the empirical section. These models introduce some novel inference issues that we illustrate in this section by studying the following N-variate SVAR process of order p:

$$\mathbf{y}_t = \boldsymbol{\tau} + \textstyle\sum_{j=1}^{p} \mathbf{A}_j \mathbf{y}_{t-j} + \mathbf{C}\boldsymbol{\varepsilon}_t^*, \quad \boldsymbol{\varepsilon}_t^*|I_{t-1} \sim i.i.d.(\mathbf{0}, \mathbf{I}_N), \tag{18}$$

where $\mathbf{C}$ is a matrix of impact multipliers and $\boldsymbol{\varepsilon}_t^*$ are "structural" shocks. The loading matrix is sometimes reparametrised as $\mathbf{C} = \mathbf{J}\boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is a diagonal matrix whose elements contain the scale of the structural shocks, while the columns of $\mathbf{J}$, whose diagonal elements are normalised to 1, measure the relative impact effects of each of the structural shocks on all the remaining variables, so that the parameters of interest become $\mathbf{j} = veco(\mathbf{J} - \mathbf{I}_N)$ and $\boldsymbol{\psi} = vecd(\boldsymbol{\Psi})$. Similarly, the drift $\boldsymbol{\tau}$ is often written as $(\mathbf{I}_N - \boldsymbol{\Phi}_1 - \ldots - \boldsymbol{\Phi}_p)\boldsymbol{\mu}$ under the assumption of covariance stationarity, where $\boldsymbol{\mu}$ is the unconditional mean of the observed process. We will revisit these interesting alternative parametrisations below, but as we discussed in section 2.2, they all give rise to asymptotically equivalent and possibly numerically identical DWH tests.

Let $\boldsymbol{\varepsilon}_t = \mathbf{C}\boldsymbol{\varepsilon}_t^*$ denote the reduced form innovations, so that $\boldsymbol{\varepsilon}_t|I_{t-1} \sim i.i.d.(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}'$. As is well known, a Gaussian (pseudo) log-likelihood is only able to identify $\boldsymbol{\Sigma}$, which means the structural shocks $\boldsymbol{\varepsilon}_t^*$ and their loadings in $\mathbf{C}$ are only identified up to an orthogonal transformation. Specifically, we can use the so-called LQ matrix decomposition[13] to relate the matrix $\mathbf{C}$ to the Cholesky decomposition of $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_L\boldsymbol{\Sigma}_L'$ as $\mathbf{C} = \boldsymbol{\Sigma}_L\mathbf{Q}$, where $\mathbf{Q}$ is an $N \times N$ orthogonal matrix, which we can model as a function of $N(N-1)/2$ parameters $\boldsymbol{\omega}$ by assuming that $|\mathbf{Q}| = 1$.[14,15] While $\boldsymbol{\Sigma}_L$ is identified from the Gaussian log-likelihood, $\boldsymbol{\omega}$ is not. In fact,

[13]The LQ decomposition is intimately related to the QR decomposition. Specifically, $\mathbf{Q}'\boldsymbol{\Sigma}_L'$ provides the QR decomposition of the matrix $\mathbf{C}'$,which is uniquely defined if we restrict the diagonal elements of $\boldsymbol{\Sigma}_L$ to be positive (see e.g. Golub and van Loan (1993) for further details).

[14]See section 9 of Magnus, Pijls and Sentana (2020) for a detailed discussion of three ways of explicitly parametrising a rotation (or special orthogonal) matrix: (i) as the product of Givens matrices that depend on $N(N-1)/2$ Tait-Bryan angles, one for each of the strict upper diagonal elements; (ii) by using the so-called Cayley transform of a skew-symmetric matrix; and (c) by exponentiating a skew-symmetric matrix. Our procedures apply regardless of the chosen parametrisation.

[15]If $|\mathbf{Q}| = -1$ instead, we can change the sign of the i^{th} structural shock and its impact multipliers in the i^{th} column of the matrix $\mathbf{C}$ without loss of generality as long as we also modify the shape parameters of the distribution of ε_{it}^* to alter the sign of all its non-zero odd moments.

the underidentification of $\boldsymbol{\omega}$ would persist even if we assumed for estimation purposes that $\boldsymbol{\varepsilon}_t^*$ followed an elliptical distribution or a location-scale mixture of normals.

Nevertheless, Lanne, Meitz and Saikkonen (2017) show that statistical identification of both the structural shocks and $\mathbf{C}$ (up to permutations and sign changes) is possible assuming (i) cross-sectional independence of the N shocks and (ii) a non-Gaussian distribution for at least $N-1$ of them. Still, the reliability of the estimated impulse response functions (IRFs) and associated forecast error variance decomposition (FEVDs) depends on the validity of the assumed distributions. For that reason, a distributional misspecification diagnostic such our DWH test, which does not specify any particular alternative hypothesis, seems particularly appropriate.

For simplicity, in the rest of this section we assume that the N structural shocks are cross-sectionally independent with symmetric marginal distributions. One particularly important example will be $\varepsilon_{it}^*|I_{t-1} \sim i.i.d.\ t(0,1,\nu_i)$. Univariate t distributions are very popular in finance as a way of capturing fat tails while nesting the traditional Gaussian assumption. Their popularity is also on the rise in macroeconomics, as illustrated by Brunnermeier et al (2019).

Let $\boldsymbol{\theta} = [\boldsymbol{\tau}', vec'(\mathbf{A}_1), \ldots, vec'(\mathbf{A}_p), vec'(\mathbf{C})]' = (\boldsymbol{\tau}', \mathbf{a}_1', \ldots, \mathbf{a}_p', \mathbf{c}') = (\boldsymbol{\tau}', \mathbf{a}', \mathbf{c}')$ denote the structural parameters characterising the first two conditional moments of $\mathbf{y}_t$. In addition, let $\boldsymbol{\varrho} = (\boldsymbol{\varrho}_1, \ldots, \boldsymbol{\varrho}_N)'$ denote the shape parameters, so that $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\varrho}')'$. In the case of the Student t, each distribution depends on a single shape parameter $\eta_i = \nu_i^{-1}$. As in previous sections, we consider two alternative ML estimators of the structural parameters in $\boldsymbol{\theta}$: a restricted one which assumes that the shape parameters are known (RMLE), and an unrestricted one that simultaneously estimates them (UMLE).

Somewhat surprisingly, it turns out that under correct distributional specification, the UMLE is efficient for all the model parameters except the standard deviations of the structural shocks. More formally, the following proposition derives the asymptotic properties of the differences between the RMLE and UMLE under the null of correct specification:

Proposition 14 *If model (18) with cross-sectionally independent symmetric structural shocks generates a covariance stationary process, then* $\sqrt{T}[\hat{\boldsymbol{\mu}}_T - \hat{\boldsymbol{\mu}}_T(\bar{\boldsymbol{\varrho}})] = o_p(1)$, $\sqrt{T}[\hat{\mathbf{a}}_T - \hat{\mathbf{a}}_T(\bar{\boldsymbol{\varrho}})] = o_p(1)$, $\sqrt{T}[\hat{\jmath}_T - \hat{\jmath}_T(\bar{\boldsymbol{\varrho}})] = o_p(1)$, *and* $\lim_{T\to\infty} V\{\sqrt{T}[\hat{\boldsymbol{\psi}}_T - \hat{\boldsymbol{\psi}}_T(\bar{\boldsymbol{\varrho}})]\} = \mathcal{P}^{\boldsymbol{\psi\psi}}(\boldsymbol{\phi}_0) - \mathcal{I}^{\boldsymbol{\psi\psi}}(\boldsymbol{\phi}_0)$.

This result implies that we should base the DWH tests on the comparison of the restricted and unrestricted ML estimators of the elements of $\boldsymbol{\psi}$, their squares or logs, thereby avoiding the need for generalised inverses that would arise if we compared the estimators of the N^2 elements of $\mathbf{c}$ (see Proposition B1.3).[16] As usual, we can obtain two asymptotically equivalent tests by

[16]If the autoregressive polynomial $(\mathbf{I}_N - \mathbf{A}_1 L - \ldots - \mathbf{A}_p L^p)$ had some unit roots, so that (18) generated a (co-) integrated process, Proposition 14 would remain valid with $\boldsymbol{\mu}$ replaced with $\boldsymbol{\tau}$, but its proof would become more involved because of the non-standard asymptotic distribution of the estimators of the conditional mean parameters. In contrast, the distribution of the ML estimators of the conditional variance parameters would remain standard (cf. Theorem 4.2 in Phillips and Durlauf (1986)).

using the scores with respect to $\boldsymbol{\psi}$ instead of the parameter estimators (see Proposition 3). Nevertheless, one should not use any of these tests when one suspects that the innovations are Gaussian not only for the lack of uniformity mentioned after Proposition 10 in section 3.6, but also because $\boldsymbol{\psi}$ is asymptotically underidentified.

The results in Holly (1982) imply that this DWH test will be asymptotically equivalent to the LR test of $H_0 : \boldsymbol{\eta} = \bar{\boldsymbol{\eta}}$ if and only if $\text{rank}(\mathcal{I}_{\mathbf{c}\boldsymbol{\varrho}}) = N$, which we show in the proof of Proposition B1. In that case, we can prove that the version of the DWH test based on the efficient scores of the unrestricted parameter estimators evaluated at the restricted parameter estimators is numerically identical to the LM test of this null hypothesis, which is entirely analogous to the discussion that follows Proposition 7.

It might appear that one cannot compare these non-Gaussian ML estimators to the Gaussian PML ones because the Gaussian pseudo log-likelihood is flat along an $N(N-1)/2$-dimensional manifold of the structural parameters $\mathbf{c}$. However, appearances are sometimes misleading. Under correct distributional specification, the non-Gaussian estimators will efficiently estimate the reduced form covariance matrix, so it is straightforward to develop DWH specification tests based on $\boldsymbol{\mu}$ (or $\boldsymbol{\tau}$), $\mathbf{a}$ and $\boldsymbol{\sigma} = vech(\boldsymbol{\Sigma})$ or its Cholesky factor $\boldsymbol{\sigma}_L = vech(\boldsymbol{\Sigma}_L)$, and their associated scores, even though we cannot do it for for $\boldsymbol{\omega}$, let alone $\mathbf{j}$ or $\boldsymbol{\psi}$.

Proposition B2 contains the asymptotic covariance matrix of the Gaussian pseudo-ML estimators of the reduced form parameters, which are asymptotically inefficient relative to the UMLEs when the innovations are non-Gaussian. In turn, Proposition B1 provides the non-Gaussian scores and information matrix for $\boldsymbol{\tau}$ and $\mathbf{a}$. Finally, Proposition B3 provides the analogous expressions for $\boldsymbol{\sigma}_L$ and $\boldsymbol{\omega}$.[17] The only unusual feature is that in computing the asymptotic covariance of the estimators of the $N(N+1)/2$ parameters in $\boldsymbol{\sigma}_L$ in the non-Gaussian case, one must take into account the sampling variability in the estimation of the $N(N-1)/2$ structural parameters in $\boldsymbol{\omega}$, as well as the drift and autoregressive parameters.

The block diagonality of all the asymptotic covariance matrices immediately implies that we can additively decompose the DWH test that compares all the reduced form parameters into a component that compares the conditional mean parameters and another one that compares the residual covariance matrix $\boldsymbol{\Sigma}$ or its Cholesky decomposition. However, Fiorentini and Sentana (2020) show that if the true joint density of the structural shocks $\boldsymbol{\varepsilon}_t^*$ in (18) is the product of N univariate densities but they are different from the ones assumed for ML estimation purposes, then the restricted and unrestricted non-Gaussian (pseudo) ML estimators of model (18) remain consistent for $\mathbf{a}$ and $\mathbf{j}$ but not for $\boldsymbol{\tau}$ or $\boldsymbol{\psi}$. Thus, the parameters that are efficiently estimated

[17] Given that the mapping from $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma}_L$ in expression (D13) of Appendix D.1 is bijective, we can invert it to obtain the scores and information matrix for $\boldsymbol{\sigma}$ and $\boldsymbol{\omega}$ from the corresponding expression for $\boldsymbol{\sigma}_L$ and $\boldsymbol{\omega}$.

by the unrestricted ML estimator remain once again consistently estimated under distributional misspecification. Although we cannot exploit the consistency of $\mathbf{j}$ to increase the power of the DWH test that compares the ML estimators of the reduced form variance parameters with the Gaussian ones because we cannot separately identify them with a Gaussian pseudo log-likelihood, it makes sense to increase the power of the DWH test that compares the ML estimators of the mean parameters with the Gaussian ones by saving degrees of freedom and focusing on either the drifts in $\boldsymbol{\tau}$ or the unconditional means in $\boldsymbol{\mu}$ even though they do not directly affect the IRFs and FEVDs. Using the results on invariance to reparametrisation in Proposition 2, the DWH test of all the mean parameters is asymptotically equivalent whether we parametrise the model in term of $(\boldsymbol{\tau}, \mathbf{a})$ or $(\boldsymbol{\mu}, \mathbf{a})$, and in fact, some of the score versions will be numerically identical. In contrast, the DWH tests that only focus on either $\boldsymbol{\tau}$ or $\boldsymbol{\mu}$ will be different.[18]

4 Monte Carlo evidence

In this section, we assess the finite sample size and power of our proposed DWH tests in the univariate and multivariate examples that we have been considering by means of extensive Monte Carlo simulation exercises. In all cases, we evaluate the three asymptotically equivalent versions of the tests in (1), (3) and (4) using the ingredients in Propositions 8 and 9. To simplify the presentation, we denote the Wald-style test that compares parameter estimators by DWH1, the test based on the score of the more efficient estimator evaluated at the less efficient one by DWH2 and, finally, the second score-based version of the test by DWH3.

Univariate GARCH-M Let r_{Mt} denote the excess returns on a broad-based portfolio. Drost and Klaassen (1997) proposed the following model for such a series:

$$r_{Mt} = \mu_t(\boldsymbol{\theta}) + \sigma_t(\boldsymbol{\theta})\varepsilon_t^*, \quad \mu_t(\boldsymbol{\theta}) = \tau\sigma_t(\boldsymbol{\theta}), \quad \sigma_t^2(\boldsymbol{\theta}) = \omega + \alpha r_{Mt-1}^2 + \beta\sigma_{t-1}^2(\boldsymbol{\theta}). \tag{19}$$

The conditional mean and variance parameters are $\boldsymbol{\theta}' = (\tau, \omega, \alpha, \beta)$. As explained in Fiorentini and Sentana (2019), this model can also be written in terms of $\boldsymbol{\vartheta}_c = (\beta, \gamma, \delta)'$ and ϑ_i, where $\gamma = \alpha/\omega$, $\delta = \tau\omega^{1/2}$ and $\vartheta_i = \omega$ (reparametrisation 1) or $\boldsymbol{\varphi}_c = (\beta, \gamma)'$, φ_{im} and φ_{ic}, where $\gamma = \alpha/\omega$, $\varphi_{im} = \tau\omega^{1/2}$ and $\varphi_{ic} = \omega$ (reparametrisation 2).

[18] The intuition is as follows. In the case of the unconditional mean parametrisation, the block diagonality of the information matrix not only arises between the conditional mean parameters and the rest, but also between $\boldsymbol{\mu}$ and $\mathbf{a}$, with the same being true for the Gaussian PMLE covariance matrix. As a result, the DWH test of the conditional mean parameters can be additively separated between the DWH test of $\boldsymbol{\mu}$, which has all the power, and the DWH test of $\mathbf{a}$, whose asymptotic power is equal to its size. In contrast, neither the information matrix nor the Gaussian sandwich matrix are block diagonal between $\boldsymbol{\tau}$ and $\mathbf{a}$ when we rely on the parametrisation in terms of the drifts, which means that the DWH test based on the drifts is not asymptotically independent from the DWH test based the dynamic regression coefficients $\mathbf{a}$. But since both the DWH test of all the mean parameters and the DWH test for $\mathbf{a}$ are the same in both reparametrisations, the DWH test based on $\boldsymbol{\tau}$ must be different from the DWH test for $\boldsymbol{\mu}$. The ordering of the local power of these two tests is unclear.

Random draws of ε_t^* are obtained from four different distributions: two standardised Student t with $\nu = 12$ and $\nu = 8$ degrees of freedom, a standardised symmetric fourth-order Gram-Charlier expansion with an excess kurtosis of 3.2, and another standardised Gram-Charlier expansion with skewness and excess kurtosis coefficients equal to -0.9 and 3.2, respectively. For a given distribution, random draws are obtained with the NAG library G05DDF and G05FFF functions, as detailed in Amengual, Fiorentini and Sentana (2013). In all four cases, we generate 20,000 samples of length 2,000 (plus another 100 for initialisation) with $\beta = 0.85$, $\alpha = 0.1$, $\tau = 0.05$ and $\omega = 1$, which means that $\delta = \varphi_{im} = 0.05$, $\gamma = 0.1$ and $\vartheta_i = \varphi_{ic} = 1$. These parameter values ensure the strict stationarity of the observed process. Under the null, the large number of Monte Carlo replications implies that the 95% percent confidence bands for the empirical rejection percentages at the conventional 1%, 5% and 10% significance levels are $(0.86, 1.14)$, $(4.70, 5.30)$ and $(9.58, 10.42)$, respectively.

We estimate the model parameters three times: first by Gaussian PML and then by maximising the log-likelihood function of the Student t distribution with and without fixing the degrees of freedom parameter to 12. We initialise the conditional variance processes by setting σ_1^2 to $\omega(1+\gamma r_M^2)/(1-\beta)$, where $r_M^2 = \frac{1}{T}\sum_1^T r_{Mt}^2$ provides an estimate of the second moment of r_{Mt}. The Gaussian, unrestricted Student t and restricted Student t log-likelihood functions are maximised with a quasi-Newton algorithm implemented by means of the NAG library E04LBF routine with the analytical expressions for the score vector and conditional information matrix in Fiorentini, Sentana and Calzolari (2003).

Table 1 contains the empirical rejections rates of the three pairwise tests in Propositions 8 and 9, together with the corresponding three-way tests. When comparing the restricted and unrestricted ML estimators, we also compute the LR test of the null hypothesis $H_0 : \eta = \bar{\eta}$. As we mentioned in section 3.4, the asymptotically equivalent LM test of this hypothesis is numerically identical to the corresponding DWH3 test because $dim(\boldsymbol{\eta}) = 1$. Hence, we obtain exactly the same statistic whether we compare the entire parameter vector $\boldsymbol{\theta}$ or the scale parameter ϑ_i only.

When the true distribution of the standardised innovations is a Student t with 12 degrees of freedom, the empirical rejections rates of all tests should be equal to their nominal sizes. This is in fact what we found except for the DWH1 and DWH2 tests that compare the restricted and unrestricted ML estimators and scores, which are rather liberal and reject the null roughly 10% more often than expected. A closer inspection of those cases revealed that even though the small sample variance of both estimators is well approximated by the variance of their asymptotic distributions, the Monte Carlo distribution of their difference is highly leptokurtic, so the resulting critical values are larger than those expected under normality. In contrast, the

DWH3 test, which in this case is invariant to reparametrisation,[19] seems to work very well.

When the true distribution is a standardised Student t with $\nu = 8$, only the tests involving the restricted ML estimators that fix the number of degrees of freedom to 12 should show some power. And indeed, this is what the second panel of Table 1 shows, with DWH3 having the best raw (i.e. non-size adjusted) power, and the LR ranking second. In turn, the three-way tests suffer a slight loss power relative to the pairwise tests that compare the two ML estimators. Finally, the empirical rejection rates of the tests that compare the unrestricted ML and PML estimators are close to their significance levels.

For the symmetric and asymmetric standardised Gram-Charlier expansions, most tests show power close or equal to one. The only exceptions are the DWH1 and DWH2 versions of the tests comparing the unrestricted ML and PML estimators. Overall, the DWH3 version our proposed tests seems to outperform the two other versions.

In addition, we find almost no correlation between the DWH tests that compare the restricted and unrestricted ML estimators and the one that compare the Gaussian PMLE with the unrestricted MLE, as expected from Propositions 4 and 5. This confirms that the distribution of the simultaneous test can be well approximated by the distribution of the sum of the two pairwise DWH tests.

Multivariate market model Let $\mathbf{r}_t$ denote the excess returns on a vector of N assets traded on the same market as r_{MT}. A very popular model is the so-called market model

$$\mathbf{r}_t = \mathbf{a} + \mathbf{b}r_{Mt} + \mathbf{\Omega}^{1/2}\boldsymbol{\varepsilon}_t^*. \tag{20}$$

The conditional mean and variance parameters are $\boldsymbol{\theta}' = (\mathbf{a}', \mathbf{b}', \boldsymbol{\omega}')$, where $\boldsymbol{\omega} = vech(\mathbf{\Omega})$ and $\mathbf{\Omega} = \mathbf{\Omega}^{1/2}\mathbf{\Omega}'^{1/2}$. In this case, Fiorentini and Sentana (2019) show that can write it in terms of $\boldsymbol{\vartheta}_c' = (\mathbf{a}', \mathbf{b}', \boldsymbol{\varpi}')$ and ϑ_i, with $\vartheta_i = |\mathbf{\Omega}|^{1/N}$ and $\mathbf{\Omega}^\circ(\boldsymbol{\varpi}) = \mathbf{\Omega}/|\mathbf{\Omega}|^{1/N}$ (reparametrisation 1) or $\boldsymbol{\varphi}_c = \mathbf{b}$, $\boldsymbol{\varphi}_{im} = \mathbf{a}$ and $\boldsymbol{\varphi}_{ic} = vech(\mathbf{\Phi}_{ic}) = vech(\mathbf{\Omega})$ (reparametrisation 2).

We consider four standardised multivariate distributions for $\boldsymbol{\varepsilon}_t^*$, including two multivariate Student t with $\nu = 12$ and $\nu = 8$ degrees of freedom, a discrete scale mixture of two normals (DSMN) with mixing probability 0.2 and variance ratio 10, and an asymmetric, location-scale mixture (DLSMN) with the same parameters but a difference in the mean vectors of the two components $\boldsymbol{\delta} = .5\ell_N$, where ℓ_N is a vector of N ones (see Amengual and Sentana (2010) and Appendix E.1, respectively, for further details). For each distribution, we generate 20,000 samples of dimension $N = 3$ and length $T = 500$ with $\mathbf{a} = .112\ell_3$, $\mathbf{b} = \ell_3$ and $\mathbf{\Omega} = \mathbf{D}^{1/2}\mathbf{R}\mathbf{D}^{1/2}$, with $\mathbf{D} = 3.136\ \mathbf{I}_3$ and the off diagonal terms of the correlation matrix $\mathbf{R}$ equal to 0.3. Finally,

[19]Proposition 2 implies that the score tests will be numerically invariant to reparametrisations if the Jacobian used to recompute the conditional expected values of the Hessian matrices $\mathcal{A}_t$ and $\mathcal{I}_t$ and the conditional covariance matrix of the scores $\mathcal{B}_t$ are evaluated at the same parameter estimators as the Jacobian involved in recomputing the scores with respect to the transformed parameters by means of the chain rule.

in each replication we generate the strongly exogenous regressor r_{Mt} as an *i.i.d.* normal with an annual mean return of 7% and standard deviation of 16%.

Table 2 show the results of the size and power assessment of our proposed DWH tests. As in the previous example, the DWH3 version of the test appears to be the best one here too, although not uniformly so. When we compare restricted and unrestricted MLE, all versions of the DWH test perform very well both in terms of size and power despite the fact that the number of parameters involved is much higher now (three intercepts, three variances and three covariances). On the other hand, the tests that compare PMLE and unrestricted MLE show some small sample size distortions, which nevertheless disappear in simulations with larger sample lengths not reported here.

When the distribution is asymmetric, the DWH2 versions of the test that focus on the scale parameter are powerful but not extremely so, the rationale being that they are designed to detect departures from the Student t distribution within the spherical family. In contrast, when we simultaneously compare $\mathbf{a}$ and $vech(\mathbf{\Omega})$, power becomes virtually 1 at all significance levels.

Once again, we find little correlation between the statistics that compare the restricted and unrestricted ML estimators and the ones that compare the Gaussian PMLE with the unrestricted MLE, as expected from Propositions 4 and 5. This confirms that we can safely approximate the distribution of the simultaneous test by the distribution of the sum of the two pairwise tests.

Structural VAR Finally, we focus on the model in section 3.8 by simulating samples from the following bivariate SVAR(1) process:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} 1.2 \\ 0.5 \end{pmatrix} + \begin{pmatrix} 0.7 & 0.5 \\ -0.2 & 0.8 \end{pmatrix} \mathbf{y}_{t-1} + \begin{pmatrix} 1 & 0.313 \\ 0.583 & 1 \end{pmatrix} \begin{pmatrix} 1.2 & 0 \\ 0 & 1.6 \end{pmatrix} \begin{pmatrix} \varepsilon^*_{1t} \\ \varepsilon^*_{2t} \end{pmatrix}.$$

In the size experiment, ε^*_{1t} and ε^*_{2t} are two independent standardised Student ts with $\eta_1 = 0.15$ and $\eta_2 = 0.10$ respectively, but in the power experiment ε^*_{1t} is drawn from a symmetric DSMN with mixing probability 0.52 and variance ratio 0.06 while ε^*_{2t} follows an asymmetric DLSMN with mixing probability 0.3, variance ratio 0.2 and $\delta = 0.5$. The sample length is $T = 2,000$.

We consider three estimators, the Gaussian PMLE, the UMLE that assumes two independent Student t for the structural shocks, and the RMLE that fixes the shape parameters at their true values in the size experiment and at $\nu_1 = 8$ and $\nu_2 = 24$ in the power experiment.

Since the main purpose of SVARs is policy analysis, it is of interest to compare the Monte Carlos means of the estimated IRFs and FEVDs to their true values. Under correct specification, all curves are virtually indistinguishable, confirming that the identification and estimation strategy in Lanne et al (2017) works remarkably well. As Figure 1 shows, though, under incorrect specification, the IRFs and FEVDs of the first variable are markedly biased even though the pattern of the IRFs is correct because $(\mathbf{I}-\mathbf{A}L)^{-1}\mathbf{J}$ is consistently estimated, as we explained at the end of section 3.8. Remarkably, the RMLE curves show very little bias, but this is a fluke that disappears by fixing the values of η_1 and η_2 to the pseudo-true values of the UMLEs.

Table 3 displays the finite sample size and power of our tests. Given the larger sample size, we observe lower finite sample size distortions than in the multivariate market model.[20] The three versions of the test show a similar behaviour, with no version uniformly superior to the others. When the distribution is not Student, power is remarkable and reaches 1 for all tests except the one that compares the PML and UML estimators of the drifts $\boldsymbol{\tau}$. Even then, the percentage of rejections of the DWH2 statistic is above 92% at the 1% nominal level. The fact that in this design only one of the shocks is asymmetric, while the tests based on $\boldsymbol{\tau}$ only have power under asymmetric shocks, might explain why we do not observe a 100% rejection rate.

5 Empirical illustrations

In Fiorentini and Sentana (2019), we illustrated the empirical relevance of our proposed consistent estimators by fitting the univariate Garch-M model (19) to the daily returns of 200 large cap stocks from the main eurozone markets between 2014 and 2018. When we compared Gaussian and unrestricted Student t MLEs by means of the score versions of our tests, we rejected the null at the 5% significance level for 36.5% of the series if we focused on symmetric alternatives (ϑ_i) and for 41% when we allowed for asymmetric ones ($\varphi_{im}, \varphi_{is}$). In addition, the DWH test that checks the adequacy of the Student t distribution with 4 degrees of freedom rejected the null at the 5% significance level for 39.5% of series, while the joint test obtained by adding the previous statistics up rejected the null for more than half of the series under analysis.

In this section, we apply our procedures to the trivariate Svar in Angelini el al (2019), who revisited the empirical analysis in Ludvigson, Ma and Ng (2015) and Carriero, Clark and Marcellino (2018). Figure 2 displays the data, which we downloaded from the JAE data archive at *http://qed.econ.queensu.ca/jae/2019-v34.3/angelini-et-al/*. It consists of monthly observations from August 1960 to April 2015 on a macro uncertainty index taken from Jurado, Ludvigson, and Ng (2015), the rate of growth of the industrial production index, and a financial uncertainty index constructed by Ludvigson, Ma and Ng (2018). As all these authors convincingly argue, a joint model of financial and macroeconomic uncertainty is crucial to understand the relationship between uncertainty and the business cycle. We adopt the original Var(4) specification in Angelini el al (2019), which implies that $T = 653$ after initialization of the log-likelihood with

[20] As expected from Proposition 10, though, size distortions become a serious problem in a separate Monte Carlo exercise in which ε^*_{1t} and ε^*_{2t} are two independent standardised Student t with with $66.\dot{6}$ and 100 degrees of freedom, respectively, which are rather difficult to distinguish from Gaussian random variables in finite samples.

4 pre-sample observations. Our main point of departure is that we assume that the structural innovations follow three independent standardised Student t distributions with ν_i degrees of freedom, which allows us to identify the entire matrix of impact multipliers $\mathbf{C} = \mathbf{J\Psi}$. Thus, the unrestricted ML procedure estimates $2N + (p+1)N^2 = 51$ parameters, while the restricted MLE fixes $\nu_1 = \nu_2 = \nu_3 = 8$ (We tried different values of ranging from 6 to 10 but results were very similar). Finally, the Gaussian PMLE estimates $N(N-1)/2 = 3$ parameters less because it can only identify $\mathbf{CC}' = \mathbf{J\Psi}^2\mathbf{J}' = \mathbf{\Sigma}$.

Our PML estimators of the autoregressive matrices coincide with those in Angelini et al (2019). Further, the restricted and unrestricted MLEs of those parameters are also very similar because the three estimators are consistent under weak conditions, as we explained in section 3.8. The estimates of the drift, the (scaled) impact multiplier matrix $\mathbf{J}$, the standard deviations of the structural shocks in $\mathbf{\Psi}$ and the unconditional variance of the one period ahead forecast errors $\mathbf{\Sigma}$ are reported in Table 4. As can be seen, the three estimators of the drift parameters are quite similar for the first two series, while for the last one the sign of the UML and RML estimators is reversed with respect to the PML one. A look at the estimators of $\mathbf{\Sigma}$ reveals both an unbalanced scaling of the data, and a low predictability in the rate of growth of the industrial production index. The restricted and unrestricted MLEs of $\mathbf{J}$ are rather similar. In fact, the consistency of the non-Gaussian ML estimators of the matrix $\mathbf{J}$ is indirectly confirmed by the extremely high (=.995) time series correlation between the (non-standardised) estimates of each structural shock obtained as $\mathbf{J}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ evaluated at the RMLE and UMLE. In contrast, there is a striking difference in the standard deviation of the third structural shock, which strongly points to distributional misspecification. However, this conjecture needs to be confirmed by our formal DWH test statistics, which account for the sampling variability of the estimators.

The three versions of our DWH tests produce qualitatively similar results. For that reason, in Table 5 we only report the results of the versions that evaluate the score of the more efficient estimators at the less efficient ones (e.g. the unrestricted Student t scores at the Gaussian PMLE). According the Monte Carlo results in the previous section, these are the most conservative ones. As expected, we conclude that the null of correct specification of the structural innovation distributions is clearly rejected. The test statistics that compares the unrestricted ML estimator of the variance of the Wold innovations $\hat{\mathbf{J}}\hat{\mathbf{\Psi}}^2\hat{\mathbf{J}}'$ with its PML counterpart $\bar{\mathbf{\Sigma}}$ has a tiny p-value. Similarly, if we compare the same estimators of the drift parameters, the p-value of our DWH statistic is .001. Given the additivity of these two test statistics mentioned at the end of section in section 3.8, the p-value of the joint test is virtually zero. As for the comparison between the restricted and unrestricted MLEs of the diagonal elements of $\mathbf{\Psi}$, which contain the standard deviations of the structural shocks, the DWH tests massively reject once again. This rejection is confirmed by the asymptotically equivalent LR test of $H_0 : \nu_1 = \nu_2 = \nu_3 = 8$.

To gauge the extent to which are results might be driven by events in the first part of our sample, we also consider a subsample that uses the second half of the available observations. Specifically, it begins in 1988:05, thereby avoiding the October 87 market crash. As can be seen from Table 6, the model is still rejected but not overwhelmingly so.

In summary, the assumption of independent, non-Gaussian structural shocks is very attractive because it allows the identification of all the model parameters without any additional restrictions, but it entails distributional misspecification risks. Our empirical results confirm that those risks cannot be ignored.

6 Conclusions and directions for further research

We propose an extension of the Durbin-Wu-Hausman specification tests which simultaneously compares three or more likelihood-based estimators of the parameters of general multivariate dynamic models with non-zero conditional means and possibly time-varying variances and covariances. Although we focus most of our discussion on the comparison of the three estimators offered by the dominant commercial econometric packages, namely, the Gaussian PML estimator, as well as ML estimators based on a non-Gaussian distribution, which either jointly estimate the additional shape parameters or fix them to some plausible values, we also consider two semiparametric estimators, one of which imposes the assumption that the standardised innovations follow a spherical distribution.

We also explore several important issues related to the practical implementation of our proposed tests, including the different versions, their numerical invariance to reparametrisations and their application to subsets of parameters. By explicitly considering a multivariate framework with non-zero conditional means we are able to cover many empirically relevant applications. Our results also apply to dynamic structural models, whose focus differs from the conditional mean and variance, and raise some interesting inference issues that we also study in detail. Extensions to stochastic volatility models in which the log-likelihood cannot be obtained in closed-form are conceptually possible as long as the ML estimators and their asymptotic variances are available, but we leave the interesting computational considerations that they raise for further research.

To select the right number of degrees of freedom, we need to figure out the rank of the difference between the estimators' asymptotic covariance matrices. In this respect, we discuss several situations in which some of the estimators are equally efficient for some of the parameters and prove that the semiparametric estimators share a saddle point efficiency property: they are as inefficient as the Gaussian PMLE for the parameters that they cannot estimate adaptively.

A comparison of our results with those in Fiorentini and Sentana (2019) imply that the parameters that are efficiently estimated by the semiparametric procedures continue to be consistently estimated by the parametric estimators under distributional misspecification. In contrast, the remaining parameters, which the semiparametric procedures can only estimate with the efficiency of the Gaussian PML estimator, will be inconsistently estimated by distributionally misspecified parametric procedures. For that reason, we focus our tests on the comparison of the estimators of this second group of parameters, for which the usual efficiency - consistency trade off is of first-order importance.

Our Monte Carlo experiments indicate that many of our proposed tests work quite well, but some versions show noticeable size distortions in small samples. Since we have a fully specified model under the null, parametric bootstrap versions might be worth exploring. An interesting extension of our Monte Carlo analysis would look at the power of our tests in models with time-varying shape parameters or misspecified first and second moment dynamics.

Given the increased popularity of Independent Component Analysis in econometric applications, as illustrated by the SVARs in section 3.8, specification tests that directly target the maintained assumptions of non-normality and independence of the structural shocks provide a particularly appropriate complement to our proposed tests (see Amengual, Fiorentini and Sentana (2020)). We could also extend our theoretical results to a broad class of models for which a pseudo log-likelihood function belonging to the linear exponential family leads to consistent estimators of the conditional mean parameters (see Gouriéroux, Monfort and Trognon (1984a)). For example, we could use a DWH test to assess the correct distributional specification of Lanne's (2006) multiplicative error model for realised volatility by comparing his ML estimator based on a two-component Gamma mixture with the Gamma-based consistent pseudo ML estimators in Engle and Gallo (2006). Similarly, we could also use the same approach to test the correct specification of the count model for patents in Hausman, Hall and Griliches (1984) by comparing their ML estimator, which assumes a Poisson model with unobserved gamma heterogeneity, with the consistent pseudo ML estimators in Gouriéroux. Monfort and Trognon (1984b)). All these extensions constitute interesting avenues for further research.

References

Acharya V., Pedersen, L.H., Philippon, T. and Richardson, M. (2017), "Measuring systemic risk", *Review of Financial Studies* 30, 2-47.

Adrian, T. and Brunnermeier, M.K. (2016): "CoVaR", *American Economic Review* 106, 1705-1741.

Amengual, D., Carrasco, M. and Sentana, E. (2019): "Testing distributional assumptions using a continuum of moments", forthcoming in the *Journal of Econometrics.*

Amengual, D., Fiorentini, G. and Sentana, E. (2013): "Sequential estimators of shape parameters in multivariate dynamic models", *Journal of Econometrics* 177, 233-249.

Amengual, D., Fiorentini, G. and Sentana, E. (2020): "Specification tests for independent component analysis", mimeo, CEMFI.

Amengual, D. and Sentana, E. (2010): "A comparison of mean-variance efficiency tests", *Journal of Econometrics* 154, 16-34.

Angelini, G., Bacchiocchi, E., Caggiano, G. and Fanelli, L. (2019): "Uncertainty across volatility regimes", *Journal of Applied Econometrics* 34, 437–455.

Andrews, D.W.K. (1987): "Asymptotic results for generalized Wald tests", *Econometric Theory* 3, 348-358.

Bai, J. (2003): "Testing parametric conditional distributions of dynamic models", *Review of Economics and Statistics* 85, 531–549.

Bai, J. and Zhihong, C. (2008): "Testing multivariate distributions in GARCH models", *Journal of Econometrics* 143, 19–36.

Bickel, P.J. (1982): "On adaptive estimation", *Annals of Statistics* 10, 647-671.

Bierens, H.J. and L. Wang (2012): "Integrated conditional moment tests for parametric conditional distributions", *Econometric Theory* 28, 328–362.

Bollerslev, T., and Wooldridge, J. M. (1992): "Quasi maximum likelihood estimation and inference in dynamic models with time-varying covariances", *Econometric Reviews* 11, 143-172.

Bontemps, C. and Meddahi, N. (2012): "Testing distributional assumptions: a GMM approach", *Journal of Applied Econometrics* 27, 978–1012.

Brunnermeier, M., Palia, D., Sastry, K.A. and Sims, C.A. (2019): "Feedbacks: financial markets and economic activity", mimeo, Princeton University.

Calzolari, G., Fiorentini, G. and Sentana, E. (2004): "Constrained indirect estimation", *Review of Economic Studies* 71, 945-973.

Carriero, A., Clark, T.E. and Marcellino, M. (2018): "Measuring uncertainty and its impact on the economy", *Review of Economics and Statistics* 100, 799-815.

Crowder, M.J. (1976): “Maximum likelihood estimation for dependent observations”, *Journal of the Royal Statistical Society* B, 38, 45-53.

Davidson, R. and MacKinnon, J.G. (1989): “Testing for consistency using artificial regressions”, *Econometric Theory* 5, 363-384.

Drost, F.C. and Klaassen, C.A.J. (1997): “Efficient estimation in semiparametric GARCH models”, *Journal of Econometrics* 80, 193-221.

Drost, F.C., Klaassen, C.A.J. and Werker, B.J.M. (1997): “Adaptive estimation in time series models”, *Annals of Statistics* 25, 786-817.

Durbin, J. (1954): “Errors in variables”, *Review International Statistical Institute* 22, 23–32.

Engle, R.F. and Gallo, G.M. (2006): “A multiple indicators model for volatility using intra-daily data”, *Journal of Econometrics* 131, 3–27.

Engle, R.F. and Gonzalez-Rivera, G. (1991): “Semiparametric ARCH models”, *Journal of Business and Economic Statistics* 9, 345-360.

Fiorentini, G. and Sentana, E. (2007): “On the efficiency and consistency of likelihood estimation in multivariate conditionally heteroskedastic dynamic regression models”, CEMFI Working Paper 0713.

Fiorentini, G. and Sentana, E. (2019): “Consistent non-Gaussian pseudo maximum likelihood estimators”, *Journal of Econometrics* 213, 321-358.

Fiorentini, G. and Sentana, E. (2020): “Discrete mixtures of normals pseudo maximum likelihood estimators of structural vector autoregressions”, CEMFI Working Paper 2023.

Fiorentini, G. Sentana, E. and Calzolari, G. (2003): “Maximum likelihood estimation and inference in multivariate conditionally heteroskedastic dynamic regression models with Student t innovations”, *Journal of Business and Economic Statistics* 21, 532-546.

Francq, C. and Zakoïan, J.-M. (2010): GARCH *models: structure, statistical inference and financial applications*, Wiley.

Gallant, A.R. and Tauchen, G. (1996): “Which moments to match?” *Econometric Theory* 12, 657-681.

Gillier, G.L. (2005): “A generalized error distribution”, http://dx.doi.org/10.2139/ssrn.2265027

Golub, G.H. and van Loan, C.F. (2013): *Matrix computations* (4th ed.), Johns Hopkins.

Gonzalez-Rivera, G. (1997): “A note on adaptation in GARCH models”, *Econometric Reviews* 16, 55-68.

Gonzalez-Rivera, G. and Drost, F.C. (1999): “Efficiency comparisons of maximum-likelihood-based estimators in GARCH models”, *Journal of Econometrics* 93, 93-111.

Gouriéroux, C., Monfort A. and Renault E. (1993): “Indirect inference”, *Journal of Applied*

Econometrics 8, S85-S118.

Gouriéroux, C., Monfort A. and Trognon, A. (1984a): "Pseudo maximum likelihood methods: theory", *Econometrica* 52, 681-700.

Gouriéroux, C., Monfort A. and Trognon, A. (1984b): "Pseudo maximum likelihood methods: applications to Poisson models", *Econometrica* 52, 701-720.

Hafner, C.M. and Rombouts, J.V.K. (2007): "Semiparametric multivariate volatility models", *Econometric Theory* 23, 251-280.

Hausman, J. (1978): "Specification tests in econometrics", *Econometrica* 46, 1273-1291.

Hausman, J., Hall, B. and Griliches, Z. (1984): "Econometric models for count data with an application to the patents R&D relationship", *Econometrica* 52, 909-938.

Hausman, J. and Taylor, W. (1981): "A generalised specification test", *Economics Letters* 8, 239-245.

Hodgson, D.J., Linton, O. and Vorkink, K.P. (2002): "Testing the Capital Asset Pricing Model efficiently under elliptical symmetry: a semiparametric approach", *Journal of Applied Econometrics* 17, 617-639.

Hodgson, D.J. and Vorkink, K.P. (2003): "Efficient estimation of conditional asset pricing models", *Journal of Business and Economic Statistics* 21, 269-283.

Holly, A. (1982): "A remark on Hausman's specification test", *Econometrica* 50, 749-759.

Holly A. (1987): "Specification tests: an overview", chap. 2 in Bewley, T.F. (ed.) *Advances in Econometrics - Fifth Word Congress*, Cambridge University Press.

Holly, A. and Monfort, A. (1986): "Some useful equivalence properties of Hausman's test", *Economics Letters* 20, 39-43.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015): "Measuring uncertainty", *American Economic Review* 105, 1177–1216.

Kotz, S. (1975): "Multivariate distributions at a cross-road", in G. P. Patil, S. Kotz and J.K. Ord (eds.) *Statistical distributions in scientific work*, vol. I, 247-270, Reidel.

Lange, K.L., Little, R.J.A., and Taylor, J.M.G. (1989): "Robust statistical modeling using the t distribution", *Journal of the American Statistical Association* 84, 881-896.

Lanne, M. (2006): "A mixture multiplicative error model for realized volatility", *Journal of Financial Econometrics* 4, 594–616.

Lanne, M., Meitz, M. and Saikkonen, P. (2017): "Identification and estimation of non-Gaussian structural vector autoregressions", *Journal of Econometrics* 196, 288-304.

Ling, S. and McAleer, M. (2003): "Asymptotic theory for a vector ARMA-GARCH model", *Econometric Theory* 19, 280-310.

Linton, O. (1993): "Adaptive estimation in Arch models", *Econometric Theory* 9, 539-569.

Ludvigson, S. C., Ma, S., and Ng, S. (2018): "Uncertainty and business cycles: exogenous impulse or endogenous response?", mimeo, NYU.

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

Magnus, J.R., Pijls, H.G.J. and Sentana, E. (2020): "The Jacobian of the exponential function", CEMFI Working Paper 2005.

Magnus, J.R. and Sentana, E. (2020): "Zero-diagonality as a linear structure", *Economics Letters* 196, 109513.

McDonald, J.B. and Newey, W.K. (1988): "Partially adaptive estimation of regression models via the generalized t distribution", *Econometric Theory* 4, 428-457.

Mencía, J. and Sentana, E. (2012): "Distributional tests in multivariate dynamic models with Normal and Student t innovations", *Review of Economics and Statistics* 94, 133-152.

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

Newey, W.K. and McFadden, D.L. (1994): "Large sample estimation and hypothesis testing", in R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics* vol. IV, 2111-2245, Elsevier.

Newey, W.K. and Steigerwald, D.G. (1997): "Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models", *Econometrica* 65, 587-99.

Phillips, P.C.B. and Durlauf, S.N. (1986): "Multiple time series regression with integrated processes", *Review of Economic Studies* 53, 473-495.

Reiss, P. (1983): "A note on the selection of parameters and estimators in the Hausman specification test", Stanford University Graduate School of Business Research Paper 708.

Ruud, P.A. (1984): "Tests of specification in econometrics", *Econometric Reviews* 3, 211-242.

Ruud, P.A. (2000): *An introduction to classical econometric theory*, Oxford University Press.

Sun, Y. and Stengos, T. (2006): "Semiparametric efficient adaptive estimation of asymmetric Garch models", *Journal of Econometrics* 127, 373-386.

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

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

White, H. (1987): "Specification testing in dynamic models", chap. 1 in Bewley, T.F. (ed.) *Advances in Econometrics - Fifth Word Congress*, Cambridge University Press.

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

Wu, D-M (1973): "Alternative tests of independence between stochastic regressors and disturbances", *Econometrica.* 41, 733–750.

Appendix

A Proofs

A.1 Proposition 1

Assuming that $\boldsymbol{\theta}_0$ belongs to the interior of its admissible parameter space, the estimators of $\boldsymbol{\theta}$ will be characterised with probability tending to 1 by the first order conditions

$$\frac{\partial \bar{\mathbf{m}}'_T(\hat{\boldsymbol{\theta}}_T)}{\partial \boldsymbol{\theta}} \tilde{\mathcal{S}}_{mT} \bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T) = \mathbf{0}, \tag{A1}$$

$$\frac{\partial \bar{\mathbf{n}}'_T(\tilde{\boldsymbol{\theta}}_T)}{\partial \boldsymbol{\theta}} \tilde{\mathcal{S}}_{nT} \bar{\mathbf{n}}_T(\tilde{\boldsymbol{\theta}}_T) = \mathbf{0}. \tag{A2}$$

By analogy, $\boldsymbol{\theta}_m$ and $\boldsymbol{\theta}_n$ will be the pseudo-true values of $\boldsymbol{\theta}$ implicitly defined by the exactly identified moment conditions

$$\begin{aligned} \mathcal{J}'_m(\boldsymbol{\theta}_m)\mathcal{S}_m E[\mathbf{m}_t(\boldsymbol{\theta}_m)] &= \mathbf{0}, \\ \mathcal{J}'_n(\boldsymbol{\theta}_n)\mathcal{S}_n E[\mathbf{n}_t(\boldsymbol{\theta}_n)] &= \mathbf{0}. \end{aligned}$$

Under the null hypothesis that both sets of moments are correctly specified, $\boldsymbol{\theta}_m = \boldsymbol{\theta}_n = \boldsymbol{\theta}_0$.

The Wald version of the DWH test in (1) is based on the difference between $\tilde{\boldsymbol{\theta}}_T$ and $\hat{\boldsymbol{\theta}}_T$. Under standard regularity conditions (see e.g. Newey and McFadden (1994)), first-order Taylor expansions of (A1) and (A2) around $\boldsymbol{\theta}_0$ imply that

$$\begin{aligned} \sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) &= -\left[\mathcal{J}'_m(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right]^{-1} \mathcal{J}'_m(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_0) + o_p(1), \\ \sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) &= -\left[\mathcal{J}'_n(\boldsymbol{\theta}_0)\mathcal{S}_n(\boldsymbol{\theta}_0)\mathcal{J}_n(\boldsymbol{\theta}_0)\right]^{-1} \mathcal{J}'_n(\boldsymbol{\theta}_0)\mathcal{S}_n\sqrt{T}\bar{\mathbf{n}}_T(\boldsymbol{\theta}_0) + o_p(1). \end{aligned} \tag{A3}$$

Therefore,

$$\begin{aligned} \sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) = &\left\{ \left[\mathcal{J}'_m(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right]^{-1} \mathcal{J}'_m(\boldsymbol{\theta}_0)\mathcal{S}_m \quad -\left[\mathcal{J}'_n(\boldsymbol{\theta}_0)\mathcal{S}_n\mathcal{J}_n(\boldsymbol{\theta}_0)\right]^{-1} \mathcal{J}'_n(\boldsymbol{\theta}_0)\mathcal{S}_n \right\} \\ &\times \begin{bmatrix} \sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_0) \\ \sqrt{T}\bar{\mathbf{n}}_T(\boldsymbol{\theta}_0) \end{bmatrix} + o_p(1). \end{aligned} \tag{A4}$$

On the other hand, the first score version of the DWH test is as a test of the moment restrictions

$$\mathcal{J}'_m(\boldsymbol{\theta}_n)\mathcal{S}_m E[\mathbf{m}_t(\boldsymbol{\theta}_n)] = \mathbf{0}. \tag{A5}$$

If we knew $\boldsymbol{\theta}_n$, it would be straightforward to test whether (A5) holds. But since we do not know it, we replace it by its consistent estimator $\tilde{\boldsymbol{\theta}}_T$, which satisfies (A2). To account for the sampling variability that this introduces under the null, we can use again a first-order Taylor expansion of the sample version of (A5) evaluated at $\tilde{\boldsymbol{\theta}}_T$ around $\boldsymbol{\theta}_0$. Given the assumed root-T

consistency of $\tilde{\boldsymbol{\theta}}_T$ for $\boldsymbol{\theta}_0$, we can use (A3) to write this expansion as

$$\mathcal{J}_m'(\tilde{\boldsymbol{\theta}}_T)\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T) = \mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_0) + \mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) + o_p(1)$$

$$= \mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_0)$$

$$-[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)][\mathcal{J}_n(\boldsymbol{\theta}_0)\mathcal{S}_n(\boldsymbol{\theta}_0)\mathcal{J}_n'(\boldsymbol{\theta}_0)]^{-1}\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n\sqrt{T}\bar{\mathbf{n}}_T(\boldsymbol{\theta}_0) + o_p(1). \quad \text{(A6)}$$

But a comparison between (A6) and (A4) makes clear that

$$\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) = \left[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right]^{-1}[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T)] + o_p(1), \quad \text{(A7)}$$

which confirms that the Wald and score versions of the test are asymptotically equivalent because $\text{rank}[\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n\mathcal{J}_n(\boldsymbol{\theta}_0)] = \dim(\boldsymbol{\theta})$ in first-order identified models. Given that $\bar{\mathbf{m}}_T(\boldsymbol{\theta})$ and $\bar{\mathbf{n}}_T(\boldsymbol{\theta})$ are exchangeable, the second equivalence condition trivially holds too. □

A.2 Proposition 2

The Wald-type version of the Hausman test for the original parameters in (1) is infeasible when $\boldsymbol{\Delta}$ is unknown, in which case it must be computed as

$$T(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T)'\boldsymbol{\Delta}_T^{\sim}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T), \quad \text{(A8)}$$

where $\boldsymbol{\Delta}_T^{\sim}$ denotes a consistent estimator of a generalised inverse of $\boldsymbol{\Delta}$, i.e. the asymptotic covariance matrix of $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T)$, which does not necessarily coincide with a generalised inverse of a consistent estimator of $\boldsymbol{\Delta}$ because of the potential discontinuities of generalised inverses. Given the assumed regularity of the reparametrisation, we can apply the delta method to show that the asymptotic covariance matrix of $\sqrt{T}(\tilde{\boldsymbol{\pi}}_T - \hat{\boldsymbol{\pi}}_T)$ will be

$$\frac{\partial\boldsymbol{\theta}'(\boldsymbol{\pi}_0)}{\partial\boldsymbol{\pi}}\boldsymbol{\Delta}\frac{\partial\boldsymbol{\theta}(\boldsymbol{\pi}_0)}{\partial\boldsymbol{\pi}'},$$

which in turn implies that we can use

$$\left[\frac{\partial\boldsymbol{\theta}(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}'}\right]^{-1}\boldsymbol{\Delta}_T^{\sim}\left[\frac{\partial\boldsymbol{\theta}'(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\right]^{-1}$$

as a consistent estimator of its generalised inverse provided that $\dot{\boldsymbol{\pi}}_T$ is a consistent estimator of $\boldsymbol{\pi}_0$. Therefore, the Wald-type version of the Hausman test for the original parameters will be

$$T(\tilde{\boldsymbol{\pi}}_T - \hat{\boldsymbol{\pi}}_T)'\left[\frac{\partial\boldsymbol{\theta}(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}'}\right]^{-1}\boldsymbol{\Delta}_T^{\sim}\left[\frac{\partial\boldsymbol{\theta}'(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\right]^{-1}(\tilde{\boldsymbol{\pi}}_T - \hat{\boldsymbol{\pi}}_T). \quad \text{(A9)}$$

Lemma 1 in Supplemental Appendix B states the numerical invariance of GMM estimators and criterion functions to reparametrisations when the weighting matrix remains the same, so that

$$\tilde{\boldsymbol{\pi}}_T - \hat{\boldsymbol{\pi}}_T = \mathbf{r}(\tilde{\boldsymbol{\theta}}_T) - \mathbf{r}(\hat{\boldsymbol{\theta}}_T).$$

In general, though, one would expect (A8) and (A9) to differ. However, when the mapping from $\boldsymbol{\theta}$ to $\boldsymbol{\pi}$ is affine, the Jacobian of the inverse transformation is the constant matrix $\mathbf{A}^{-1}$, yielding

$$T(\tilde{\boldsymbol{\pi}}_T - \hat{\boldsymbol{\pi}}_T)'\mathbf{A}'^{-1}\boldsymbol{\Delta}_T^{\sim}\mathbf{A}^{-1}(\tilde{\boldsymbol{\pi}}_T - \hat{\boldsymbol{\pi}}_T) = T(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T)'\boldsymbol{\Delta}_T^{\sim}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T),$$

as required.

Let us now look at one of the score versions of the DWH test in terms of the original parameters, the other one being entirely analogous. We saw in the proof of the previous proposition that the first-order condition for $\hat{\boldsymbol{\theta}}_T$ is (A1). Therefore, we can compute the alternative DWH test in practice as

$$T\bar{\mathbf{m}}_T'(\tilde{\boldsymbol{\theta}}_T)\tilde{\mathcal{S}}_{mT}\frac{\partial\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T)}{\partial\boldsymbol{\theta}'}\boldsymbol{\Lambda}_{mT}^{\sim}\frac{\partial\bar{\mathbf{m}}_T'(\tilde{\boldsymbol{\theta}}_T)}{\partial\boldsymbol{\theta}}\tilde{\mathcal{S}}_{mT}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T). \tag{A10}$$

Lemma 1 also implies that $\bar{\mathbb{m}}_T(\boldsymbol{\pi}) = \bar{\mathbf{m}}_T[\boldsymbol{\theta}(\boldsymbol{\pi})]$ and $\tilde{\boldsymbol{\theta}}_T = \boldsymbol{\theta}(\tilde{\boldsymbol{\pi}}_T)$ when the weighting matrix used to compute $\tilde{\boldsymbol{\theta}}_T$ and $\tilde{\boldsymbol{\pi}}_T$ is common. Given the assumed regularity of the reparametrisation, we can easily show that the asymptotic covariance matrix of $\mathcal{J}_{\mathbb{m}}'(\boldsymbol{\pi}_0)\mathcal{S}_m\sqrt{T}\bar{\mathbb{m}}_T(\tilde{\boldsymbol{\pi}}_T)$ will be

$$\boldsymbol{\Lambda}_{\mathbb{m}} = \frac{\partial\boldsymbol{\theta}'(\boldsymbol{\pi}_0)}{\partial\boldsymbol{\pi}}\boldsymbol{\Lambda}_m\frac{\partial\boldsymbol{\theta}(\boldsymbol{\pi}_0)}{\partial\boldsymbol{\pi}'}.$$

As a consequence, it seems natural to use

$$\left[\frac{\partial\boldsymbol{\theta}(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}'}\right]^{-1}\boldsymbol{\Lambda}_{mT}^{\sim}\left[\frac{\partial\boldsymbol{\theta}'(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\right]^{-1} \tag{A11}$$

as a consistent estimator of a generalised inverse of $\boldsymbol{\Lambda}_{\mathbb{m}}$, provided that $\dot{\boldsymbol{\pi}}_T$ is a consistent estimator of $\boldsymbol{\pi}_0$. Therefore, we can compute the analogous test in terms of $\boldsymbol{\pi}$ as

$$T\bar{\mathbb{m}}_T'(\tilde{\boldsymbol{\pi}}_T)\tilde{\mathcal{S}}_{mT}\frac{\partial\bar{\mathbb{m}}_T(\tilde{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}'}\left[\frac{\partial\boldsymbol{\theta}(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}'}\right]^{-1}\boldsymbol{\Lambda}_{mT}^{\sim}\left[\frac{\partial\boldsymbol{\theta}'(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\right]^{-1}\frac{\partial\bar{\mathbb{m}}_T'(\tilde{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\tilde{\mathcal{S}}_{mT}\bar{\mathbb{m}}_T(\tilde{\boldsymbol{\pi}}_T). \tag{A12}$$

Combining the chain rule for derivatives with the results in Lemma 1, we can prove that

$$\frac{\partial\bar{\mathbb{m}}_T'(\tilde{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\tilde{\mathcal{S}}_{mT}\bar{\mathbb{m}}_T(\tilde{\boldsymbol{\pi}}_T) = \frac{\partial\boldsymbol{\theta}'(\tilde{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\frac{\partial\bar{\mathbf{m}}_T'(\tilde{\boldsymbol{\theta}}_T)}{\partial\boldsymbol{\theta}}\tilde{\mathcal{S}}_{mT}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T),$$

which in turn implies that

$$\begin{aligned}
&\bar{\mathbb{m}}_T'(\tilde{\boldsymbol{\pi}}_T)\tilde{\mathcal{S}}_{mT}\frac{\partial\bar{\mathbb{m}}_T(\tilde{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}'}\left[\frac{\partial\boldsymbol{\theta}(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}'}\right]^{-1}\boldsymbol{\Lambda}_{mT}^{\sim}\left[\frac{\partial\boldsymbol{\theta}'(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\right]^{-1}\frac{\partial\bar{\mathbb{m}}_T'(\tilde{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\tilde{\mathcal{S}}_{mT}\bar{\mathbb{m}}_T(\tilde{\boldsymbol{\pi}}_T)\\
= \;&\bar{\mathbf{m}}_T'(\tilde{\boldsymbol{\theta}}_T)\tilde{\mathcal{S}}_{mT}\frac{\partial\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T)}{\partial\boldsymbol{\theta}'}\frac{\partial\boldsymbol{\theta}(\tilde{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}'}\left[\frac{\partial\boldsymbol{\theta}(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}'}\right]^{-1}\boldsymbol{\Lambda}_{mT}^{\sim}\left[\frac{\partial\boldsymbol{\theta}'(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\right]^{-1}\frac{\partial\boldsymbol{\theta}'(\tilde{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\frac{\partial\bar{\mathbf{m}}_T'(\tilde{\boldsymbol{\theta}}_T)}{\partial\boldsymbol{\theta}}\tilde{\mathcal{S}}_{mT}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T).
\end{aligned}$$

Therefore, (A10) and (A12) will be numerically identical if

$$\frac{\partial\boldsymbol{\theta}(\tilde{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}'}\left[\frac{\partial\boldsymbol{\theta}(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}'}\right]^{-1} = \mathbf{I}_p.$$

Sufficient conditions for this to happen are that the mapping is affine, or that we use $\dot{\boldsymbol{\pi}}_T = \tilde{\boldsymbol{\pi}}_T$ in computing (A11). □

A.3 Proposition 3

Again, we focus on the first result, as the second one is entirely analogous. Let us start from the asymptotic equivalence relationship (A7). Given that

$$\begin{aligned}
\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0) &= \begin{bmatrix} \mathcal{J}_{1m}'(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{1m}(\boldsymbol{\theta}) & \mathcal{J}_{1m}'(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{2m}(\boldsymbol{\theta}) \\ \mathcal{J}_{2m}'(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{1m}(\boldsymbol{\theta}) & \mathcal{J}_{2m}'(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{2m}(\boldsymbol{\theta}) \end{bmatrix} \textit{ and} \\
\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T) &= \begin{bmatrix} \mathcal{J}_{1m}'(\boldsymbol{\theta})\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T) \\ \mathcal{J}_{2m}'(\boldsymbol{\theta})\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T) \end{bmatrix},
\end{aligned}$$

the application of the partitioned inverse formula yields

$$\begin{aligned}
\sqrt{T}(\tilde{\boldsymbol{\theta}}_{1T} - \hat{\boldsymbol{\theta}}_{1T}) &= \left[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right]^{11}\bar{\mathbf{m}}_{1T}^{\perp}(\tilde{\boldsymbol{\theta}}_T,\mathcal{S}_m), \textit{ where} \\
\left[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right]^{11} &= \begin{bmatrix} \mathcal{J}_{1m}'(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{1m}(\boldsymbol{\theta}) \\ -\mathcal{J}_{1m}'(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{2m}(\boldsymbol{\theta})[\mathcal{J}_{2m}'(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{2m}(\boldsymbol{\theta})]^{-1}\mathcal{J}_{2m}'(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{1m}(\boldsymbol{\theta}) \end{bmatrix}^{-1}.
\end{aligned}$$

Given that $\left[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right]^{11}$ will have rank p_1 because $[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)]$ has rank p, the Wald version of the DWH test that focuses on $\boldsymbol{\theta}_1$ only is equivalent to a score version that looks at $\bar{\mathbf{m}}_{1T}^{\perp}(\tilde{\boldsymbol{\theta}}_T,\mathcal{S}_n)$. □

A.4 Proposition 4

Given that

$$\begin{pmatrix} \hat{\boldsymbol{\theta}}_T^2 - \hat{\boldsymbol{\theta}}_T^1 \\ \hat{\boldsymbol{\theta}}_T^3 - \hat{\boldsymbol{\theta}}_T^2 \\ \vdots \\ \hat{\boldsymbol{\theta}}_T^{J-1} - \hat{\boldsymbol{\theta}}_T^{J-2} \\ \hat{\boldsymbol{\theta}}_T^J - \hat{\boldsymbol{\theta}}_T^{J-1} \end{pmatrix} = \begin{bmatrix} -\mathbf{I} & \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} & \mathbf{I} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & -\mathbf{I} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & -\mathbf{I} & \mathbf{I} \end{bmatrix} \begin{pmatrix} \hat{\boldsymbol{\theta}}_T^1 \\ \hat{\boldsymbol{\theta}}_T^2 \\ \hat{\boldsymbol{\theta}}_T^3 \\ \vdots \\ \hat{\boldsymbol{\theta}}_T^{J-2} \\ \hat{\boldsymbol{\theta}}_T^{J-1} \\ \hat{\boldsymbol{\theta}}_T^J \end{pmatrix}, \quad \text{(A13)}$$

it follows immediately from (5) that

$$\lim_{T\to\infty} V\left[\begin{pmatrix} \hat{\boldsymbol{\theta}}_T^2 - \hat{\boldsymbol{\theta}}_T^1 \\ \hat{\boldsymbol{\theta}}_T^3 - \hat{\boldsymbol{\theta}}_T^2 \\ \vdots \\ \hat{\boldsymbol{\theta}}_T^{J-1} - \hat{\boldsymbol{\theta}}_T^{J-2} \\ \hat{\boldsymbol{\theta}}_T^J - \hat{\boldsymbol{\theta}}_T^{J-1} \end{pmatrix}\right] = \begin{bmatrix} \boldsymbol{\Omega}_2-\boldsymbol{\Omega}_1 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_3-\boldsymbol{\Omega}_2 & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Omega}_{J-1}-\boldsymbol{\Omega}_{J-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \boldsymbol{\Omega}_J-\boldsymbol{\Omega}_{J-1} \end{bmatrix}, \quad \text{(A14)}$$

which in turn implies the asymptotic independence of non-overlapping DWH test statistics of the form (1). But since (A13) holds for any T, all $J(J-1)/2$ possible differences between any two of the J estimators will be linear combinations of the $J-1$ adjacent differences in (A14).□

A.5 Proposition 5

Given that Propositions C1-C3 in Supplemental Appendix C and Proposition D3 in Supplemental Appendix D derive all the information bounds, we simply need to compute the off-diagonal elements. Let us start with the first row. Straightforward manipulations imply that

$$
\begin{aligned}
E[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathbf{s}'_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}] &= E\{\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})[\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi})-\mathbf{s}'_{\boldsymbol{\eta}t}(\boldsymbol{\phi})\mathcal{I}^{-1}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi})\mathcal{I}'_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi})]|\boldsymbol{\phi}\} \\
&= \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi})-\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi})\mathcal{I}^{-1}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi})\mathcal{I}'_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi})=\mathcal{P}(\boldsymbol{\phi}).
\end{aligned}
$$

Intuitively, $\mathcal{P}(\boldsymbol{\phi}_0)$ is the covariance matrix of the residuals in the multivariate theoretical regression of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$ on $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}_0)$, which trivially coincides with the covariance matrix between those residuals and $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$. Next,

$$
\begin{aligned}
&E[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathring{\mathbf{s}}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}] = E[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\{\mathbf{e}'_{dt}(\boldsymbol{\phi})\mathbf{Z}'_{dt}(\boldsymbol{\theta})-[\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi})-\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathcal{K}}(0)]\mathbf{Z}'_{d}(\boldsymbol{\phi})\}|\boldsymbol{\phi}] \\
&= E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\phi})\mathbf{Z}_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}\right]-E\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})[\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi})-\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathcal{K}}(0)]\mathbf{Z}'_{d}(\boldsymbol{\phi})|\boldsymbol{\phi}\} \\
&= \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)-\mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}'_s(\boldsymbol{\phi}_0)\cdot\left\{\left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta}_0)-1\right]-\frac{4}{N\left[(N+2)\kappa_0+2\right]}\right\}=\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)
\end{aligned}
$$

by virtue of the law of iterated expectations, together with expressions (C33), (C34) and (C35) in Supplemental Appendix C. Intuitively, $\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)$ is the variance of the error in the least squares projection of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$ onto the Hilbert space spanned by all the time-invariant functions of $\varsigma_t(\boldsymbol{\theta}_0)$ with bounded second moments that have zero conditional means and are conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$, which trivially coincides with the covariance matrix between those residuals and $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$. Given that this Hilbert space includes the linear span of $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}_0)$, it follows immediately that $\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)$ is smaller than $\mathcal{P}(\boldsymbol{\phi}_0)$ in the positive semidefinite sense.

We also know from the proof of proposition D3 in Supplemental Appendix D that

$$
\begin{aligned}
E[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\ddot{\mathbf{s}}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}] &= E[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\left\{\mathbf{e}'_{dt}(\boldsymbol{\phi})\mathbf{Z}'_{dt}(\boldsymbol{\theta})-\left[\mathbf{e}'_{dt}(\boldsymbol{\phi})-\mathbf{e}'_{dt}(\boldsymbol{\theta},0)\mathcal{K}^{+}(\boldsymbol{\varrho})\,\mathcal{K}(0)\right]\mathbf{Z}'_{d}(\boldsymbol{\phi})\right\}|\boldsymbol{\phi}] \\
&= E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})\mathbf{e}'_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})\mathbf{Z}_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}\right] \\
&\quad -E\left\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\left[\mathbf{e}'_{dt}(\boldsymbol{\phi})-\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathcal{K}^{+}(\boldsymbol{\varrho})\,\mathcal{K}(0)\right]\mathbf{Z}'_{d}(\boldsymbol{\phi})|\boldsymbol{\phi}\right\} \\
&= \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi})-\mathbf{Z}_{d}(\boldsymbol{\phi})\left[\mathcal{M}_{dd}(\boldsymbol{\varrho}_0)-\mathcal{K}(0)\,\mathcal{K}^{+}(\boldsymbol{\varrho}_0)\mathcal{K}(0)\right]\mathbf{Z}'_{d}(\boldsymbol{\phi})=\ddot{\mathcal{S}}(\boldsymbol{\phi}_0)
\end{aligned}
$$

by virtue of the law of iterated expectations, together with expressions (B3) and (C22) in appendices B and C, respectively. Intuitively, $\ddot{\mathcal{S}}(\boldsymbol{\phi}_0)$ is the covariance matrix of the errors in the projection of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$ onto the Hilbert space spanned by all the time-invariant functions of $\boldsymbol{\varepsilon}^*_t$ with zero conditional means and bounded second moments that are conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$, which trivially coincides with the covariance matrix between those residuals and $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$. The fact that the residual variance of a multivariate regression cannot increase as we increase the number of regressors explains why $\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)$ is at least as large (in the positive semidefinite matrix sense) as $\ddot{\mathcal{S}}(\boldsymbol{\phi}_0)$, reflecting the fact that the relevant tangent sets become increasing larger. Finally,

$$
E[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}]=-\partial E[\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}]/\partial\boldsymbol{\theta}=\mathcal{A}(\boldsymbol{\phi})
$$

thanks to the generalised information equality.

Let us now move on to the second row, and in particular to

$$E[\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\phi})\mathring{\mathbf{s}}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}] = E[\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})$$
$$-\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi})\mathcal{I}^{-1}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi})\mathbf{e}_{rt}(\boldsymbol{\phi})\}\left\{\mathbf{e}'_{dt}(\boldsymbol{\phi})\mathbf{Z}'_{dt}(\boldsymbol{\theta}) - \left[\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi}) - \mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathcal{K}}(0)\right]\mathbf{Z}'_{d}(\boldsymbol{\phi})\right\}|\boldsymbol{\phi}]$$
$$= E[\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\mathbf{Z}'_{dt}(\boldsymbol{\phi}_0)|\boldsymbol{\phi}] - E[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi})\mathbf{Z}'_{dt}(\boldsymbol{\phi}_0)|\boldsymbol{\phi}]$$
$$+E[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathcal{K}^{+}(\boldsymbol{\varrho}_0)\mathcal{K}(0)\,\mathbf{Z}'_{d}(\boldsymbol{\phi})|\boldsymbol{\phi}] - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi})\mathcal{I}^{-1}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi})E[\mathbf{e}_{rt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\phi})\mathbf{Z}'_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}]$$
$$+\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi})\mathcal{I}^{-1}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi})E[\mathbf{e}_{rt}(\boldsymbol{\phi})\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi})\mathbf{Z}'_{d}(\boldsymbol{\theta})|\boldsymbol{\phi}] - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi})\mathcal{I}^{-1}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi})E[\mathbf{e}_{rt}(\boldsymbol{\phi})\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathcal{K}}(0)\,\mathbf{Z}'_{d}(\boldsymbol{\phi})|\boldsymbol{\phi}]$$
$$= \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}) - \mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}'_s(\boldsymbol{\phi}_0)\cdot\left\{\left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1\right] - \frac{4}{N\left[(N+2)\kappa_0+2\right]}\right\} = \mathring{\mathcal{S}}(\boldsymbol{\phi}_0)$$

where we have used the fact that

$$\begin{aligned} E[\mathbf{e}_{rt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\phi})|\boldsymbol{\phi}] &= E\{E[\mathbf{e}_{rt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\phi})|\varsigma_t,\boldsymbol{\phi}]|\boldsymbol{\phi}\} = E[\mathbf{e}_{rt}(\boldsymbol{\phi})\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi})|\boldsymbol{\phi}] \\ &= E\left\{\mathbf{e}_{rt}(\boldsymbol{\phi})\left[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N) - 1\right]|\boldsymbol{\phi}\right\}\left[\begin{array}{cc}\mathbf{0} & vec'(\mathbf{I}_N)\end{array}\right] \textit{ and} \\ E[\mathbf{e}_{rt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}] &= E\{E[\mathbf{e}_{rt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})|\varsigma_t,\boldsymbol{\phi}]|\boldsymbol{\phi}\} = E[\mathbf{e}_{rt}(\boldsymbol{\phi})\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}] \\ &= E\left\{\mathbf{e}_{rt}(\boldsymbol{\phi})\left[(\varsigma_t/N) - 1\right]|\boldsymbol{\phi}\right\}\left[\begin{array}{cc}\mathbf{0} & vec'(\mathbf{I}_N)\end{array}\right] = \mathbf{0} \end{aligned}$$

by virtue of Lemma 3 in Supplemental Appendix B. Similarly,

$$E[\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\phi})\ddot{\mathbf{s}}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}] = E[\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})$$
$$-\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi})\mathcal{I}^{-1}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi})\mathbf{e}_{rt}(\boldsymbol{\phi})\}\{\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)[\mathbf{Z}'_{dt}(\boldsymbol{\phi}_0) - \mathbf{Z}'_{d}(\boldsymbol{\phi})] - \mathbf{e}'_{dt}(\boldsymbol{\theta}_0,\mathbf{0})\mathcal{K}^{+}(\boldsymbol{\varrho}_0)\mathcal{K}(0)\,\mathbf{Z}'_{d}(\boldsymbol{\phi})\}|\boldsymbol{\phi}]$$
$$= E[\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\mathbf{Z}'_{dt}(\boldsymbol{\phi}_0)|\boldsymbol{\phi}] - E[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\mathbf{Z}'_{d}(\boldsymbol{\phi})]|\boldsymbol{\phi}]$$
$$-E[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathcal{K}^{+}(\boldsymbol{\varrho}_0)\mathcal{K}(0)\,\mathbf{Z}'_{d}(\boldsymbol{\theta})|\boldsymbol{\phi}]$$
$$= \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\phi})\left[\mathcal{M}_{dd}(\boldsymbol{\varrho}_0) - \mathcal{K}(0)\,\mathcal{K}^{+}(\boldsymbol{\varrho}_0)\mathcal{K}(0)\right]\mathbf{Z}'_{d}(\boldsymbol{\phi}) = \ddot{\mathcal{S}}(\boldsymbol{\phi}_0)$$

because $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})$ is orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})$ by virtue of Lemma 3 and

$$E[\mathbf{e}_{rt}(\boldsymbol{\phi})\}\{\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)[\mathbf{Z}'_{dt}(\boldsymbol{\phi}_0) - \mathbf{Z}'_{d}(\boldsymbol{\phi})]\}|\boldsymbol{\phi}] = \mathbf{0}$$

by the law of iterated expectations. Finally,

$$E[\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\phi})\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}] = E[\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi})\mathcal{I}^{-1}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi})\mathbf{e}_{rt}(\boldsymbol{\phi})\}\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathbf{Z}'_{dt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = \mathcal{A}(\boldsymbol{\phi})$$

because of the generalised information equality and the orthogonality of $\mathbf{e}_{rt}(\boldsymbol{\phi})$ and $\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})$.

Let us start the third row with

$$E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\ddot{\mathbf{s}}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}] = E[\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\phi})[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}) - \mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})]\}$$
$$\times\{\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)[\mathbf{Z}'_{dt}(\boldsymbol{\phi}_0) - \mathbf{Z}'_{d}(\boldsymbol{\phi})] - \mathbf{e}'_{dt}(\boldsymbol{\theta}_0,\mathbf{0})\mathcal{K}^{+}(\boldsymbol{\varrho}_0)\mathcal{K}(0)\,\mathbf{Z}'_{d}(\boldsymbol{\phi})\}|\boldsymbol{\phi}]$$
$$= \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\phi})\left[\mathcal{M}_{dd}(\boldsymbol{\varrho}_0) - \mathcal{K}(0)\,\mathcal{K}^{+}(\boldsymbol{\varrho}_0)\mathcal{K}(0)\right]\mathbf{Z}'_{d}(\boldsymbol{\phi}) = \ddot{\mathcal{S}}(\boldsymbol{\phi}_0) \textit{ because}$$
$$E\{[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}) - \mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})]\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)[\mathbf{Z}'_{dt}(\boldsymbol{\phi}_0) - \mathbf{Z}'_{d}(\boldsymbol{\phi})]|\boldsymbol{\phi}\} = \mathbf{0}$$

by the law of iterated expectations. In addition, we have that

$$E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}] = \mathcal{A}(\boldsymbol{\phi}), \tag{A15}$$

which follows immediately from (A21) and the generalised information matrix equality.

Turning to the last off-diagonal element, we can show that

$$E[\ddot{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}] = E[\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathbf{Z}_{d}(\boldsymbol{\theta},\boldsymbol{\varrho})\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathcal{K}(0)\mathcal{K}^{+}(\boldsymbol{\varrho})\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\}$$
$$\times\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathbf{Z}'_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}] = \mathcal{A}(\boldsymbol{\theta})$$

because $\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})$ is conditionally orthogonal to $[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathcal{K}(0)\mathcal{K}^{+}(\boldsymbol{\varrho})\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})]$ by construction. This result also proves the positive semidefiniteness of $\ddot{\mathcal{S}}(\boldsymbol{\phi}_0) - \mathcal{A}(\boldsymbol{\theta})\mathcal{B}^{-1}(\boldsymbol{\phi})\mathcal{A}(\boldsymbol{\theta})$ because this expression coincides with the residual covariance matrix in the theoretical regression of the semiparametric efficient score on the Gaussian pseudo-score.

To prove the second part of the proposition, it is convenient to regard each estimator as an exactly identified GMM estimator based on the corresponding score, whose asymptotic variance depends on the asymptotic variance of this score and the corresponding expected Jacobian. In this regard, note that the information matrix equality applied to the restricted and unrestricted versions of the efficient score implies that

$$-\partial E[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}]/\partial\boldsymbol{\theta}' = E[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}] = \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}) \text{ and}$$
$$-\partial E[\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}]/\partial\boldsymbol{\theta}' = E[\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\phi})\mathbf{s}'_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}] = \mathcal{P}(\boldsymbol{\phi}).$$

Similarly, we can use the generalised information matrix equality together with some of the arguments in the proof of Proposition C3 in Supplemental Appendix C to show that

$$-\partial E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}]/\partial\boldsymbol{\theta} = E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)|\boldsymbol{\phi}] = E[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0)\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\mathbf{Z}'_{dt}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0]$$
$$-E\left\{\mathbf{W}_s(\boldsymbol{\phi}_0)\left[\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}-1\right] - \frac{2}{(N+2)\kappa_0+2}\left(\frac{\varsigma_t}{N}-1\right)\right]\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\mathbf{Z}'_{dt}(\boldsymbol{\theta}_0)\right|\boldsymbol{\phi}_0\right\}$$
$$= \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)E\left\{\left[\left\{\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}-1\right\} - \frac{2}{(N+2)\kappa_0+2}\left(\frac{\varsigma_t}{N}-1\right)\right]\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\right|\boldsymbol{\phi}_0\right\}\mathbf{Z}_d(\boldsymbol{\theta}_0)$$
$$= \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)\text{-}\mathbf{W}_s(\boldsymbol{\phi}_0)E\left[\left\{\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}\text{-}1\right]\text{-}\frac{2}{(N+2)\kappa_0+2}\left(\frac{\varsigma_t}{N}\text{-}1\right)\right\}\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}\text{-}1\right]\right|\boldsymbol{\phi}_0\right]\mathbf{W}'_s(\boldsymbol{\phi}_0)$$
$$= \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}'_s(\boldsymbol{\phi}_0)\cdot\left\{\left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta}_0)-1\right] - \frac{4}{N[(N+2)\kappa_0+2]}\right\}$$
$$= \mathring{\mathcal{S}}(\boldsymbol{\phi}_0) = E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathring{\mathbf{s}}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}]. \tag{A16}$$

The generalised information matrix equality also implies that

$$-\frac{\partial E[\ddot{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)|\boldsymbol{\phi}_0]}{\partial\boldsymbol{\theta}} = E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)|\boldsymbol{\phi}] = E[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0)\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\mathbf{Z}'_{dt}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0].$$

On this basis, we can use standard first-order expansions of $\sqrt{T}[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\eta}_0)-\boldsymbol{\theta}_0]$ and $\sqrt{T}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)$ to show that

$$\lim_{T\to\infty} E\{T[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\eta}_0)-\boldsymbol{\theta}_0](\hat{\boldsymbol{\theta}}_T'-\boldsymbol{\theta}_0')\} = \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\phi}) \lim_{T\to\infty} E[T\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\boldsymbol{\phi})\bar{\mathbf{s}}_{\boldsymbol{\theta}|\boldsymbol{\eta}T}'(\boldsymbol{\phi})\}\mathcal{P}^{-1}(\boldsymbol{\phi}) = \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\phi}).$$

All the remaining asymptotic covariances are obtained analogously. □

A.6 Proposition 6

Given the efficiency of $\hat{\boldsymbol{\theta}}_T$ relative to $\tilde{\boldsymbol{\theta}}_T$, it follows from Lemma 2 in Hausman (1978) that

$$\sqrt{T}(\tilde{\boldsymbol{\theta}}_T-\hat{\boldsymbol{\theta}}_T) \to N\left[\mathbf{0}, \mathcal{C}(\boldsymbol{\phi}_0)-\mathcal{P}^{-1}(\boldsymbol{\phi}_0)\right].$$

The other two results follow directly from Proposition 1 after taking into account that

$$\begin{aligned} -\partial E[\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}]/\partial\boldsymbol{\theta}' &= \mathcal{P}(\boldsymbol{\phi}) \\ -\partial E[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}]/\partial\boldsymbol{\theta}' &= \mathcal{A}(\boldsymbol{\phi}) \end{aligned} \quad \text{(A17)}$$

by the generalised information matrix equality. □

A.7 Proposition 7

The efficiency of $\hat{\boldsymbol{\theta}}_T(\boldsymbol{\eta})$ relative to $\hat{\boldsymbol{\theta}}_T$ and Lemma 2 in Hausman (1978) imply that

$$\sqrt{T}[\hat{\boldsymbol{\theta}}_T-\hat{\boldsymbol{\theta}}_T(\boldsymbol{\eta})] \to N\left[\mathbf{0}, \mathcal{I}^{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)-\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\phi}_0)\right]$$

under then null of correct specification. The other two results follow directly from Proposition 1 and the partitioned inverse formula after taking into account that (A17) and

$$-\partial E[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\bar{\boldsymbol{\eta}})|\boldsymbol{\phi}]/\partial\boldsymbol{\theta}' = \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi})$$

by the information matrix equality. □

A.8 Proposition 8

The proof of Proposition 6 immediately implies that

$$\sqrt{T}(\tilde{\boldsymbol{\theta}}_T-\hat{\boldsymbol{\theta}}_T) \to N\left[\mathbf{0}, \mathcal{C}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)-\mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)\right]$$

under the null. If we combine this result with Proposition 3, we obtain the expressions for the asymptotic variances of the two asymptotically equivalent score versions. □

A.9 Proposition 9

The proof of Proposition 7 immediately implies that

$$\sqrt{T}[\hat{\boldsymbol{\theta}}_{1T}-\hat{\boldsymbol{\theta}}_{1T}(\boldsymbol{\eta})] \to N\{\mathbf{0}, [\mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)-\mathcal{I}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)]\}$$

under the null. If we combine this result with Proposition 3, we obtain the expressions for the asymptotic variances of the two asymptotically equivalent score versions. □

A.10 Proposition 10

The proof of the first part is trivial, except perhaps for the fact that $\mathcal{M}_{sr}(\mathbf{0}) = \mathbf{0}$, which follows from Lemma 3 in Supplemental Appendix B because $\mathbf{e}_{st}(\boldsymbol{\theta}_0, \mathbf{0})$ coincides with $\mathbf{e}_{st}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0)$ under normality.

To prove the second part, we use the fact that after some tedious algebraic manipulations we can write $\mathcal{M}_{dd}(\boldsymbol{\eta}) - \mathcal{K}(0)\mathcal{K}^{+}(\kappa)\mathcal{K}(0)$ in the spherical case as

$$\left\{ \begin{array}{cc} [\mathrm{M}_{ll}(\boldsymbol{\eta})\text{-}1]\mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \left[\mathrm{M}_{ss}(\boldsymbol{\eta})\text{-}\frac{1}{\kappa+1}\right](\mathbf{I}_{N^2}+\mathbf{K}_{NN}) + \left[\mathrm{M}_{ss}(\boldsymbol{\eta}_0)\text{-}1+\frac{2\kappa}{(\kappa+1)[(N+2)\kappa+2]}\right] vec(\mathbf{I}_N)vec'(\mathbf{I}_N) \end{array} \right\}.$$

Therefore, given that $\mathbf{Z}_l(\boldsymbol{\phi}_0) \neq \mathbf{0}$, $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) - \ddot{\mathcal{S}}(\boldsymbol{\phi})$ will be zero only if $\mathrm{M}_{ll}(\boldsymbol{\eta}) = 1$, which in turn requires that the residual variance in the multivariate regression of $\delta(\varsigma_t, \boldsymbol{\eta}_0)\boldsymbol{\varepsilon}_t^*$ on $\boldsymbol{\varepsilon}_t^*$ is zero for all t, or equivalently, that $\delta(\varsigma_t, \boldsymbol{\eta}_0) = 1$. But since the solution to this differential equation is $g(\varsigma_t, \boldsymbol{\eta}) = -.5\varsigma_t + C$, then the result follows from (C19) in Supplemental Appendix C.

If the true conditional mean were 0, and this was taken into account in estimation, then the first diagonal block would disappear, and $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) - \ddot{\mathcal{S}}(\boldsymbol{\phi})$ could also be 0 if

$$\mathbf{Z}_d(\boldsymbol{\theta}, \boldsymbol{\varrho})\left[\mathcal{M}_{dd}(\boldsymbol{\varrho}) - \mathcal{K}(0)\mathcal{K}^{+}(\boldsymbol{\varrho})\mathcal{K}(0)\right]\mathbf{Z}_d'(\boldsymbol{\theta}, \boldsymbol{\varrho}) = \mathbf{0}.$$

Although this condition is unlikely to hold otherwise, it does not strictly speaking require normality. For example, Amengual, Fiorentini and Sentana (2013), correcting an earlier typo in Amengual and Sentana (2010), show that

$$\mathrm{M}_{ss}(\boldsymbol{\eta}_0) = \frac{N\kappa + 2}{(N+2)\kappa + 2}$$

for the Kotz distribution, which immediately implies that

$$\begin{aligned} \mathrm{M}_{ss}(\boldsymbol{\eta}) - \frac{1}{\kappa+1} &= \frac{N\kappa^2}{(\kappa+1)(2\kappa+N\kappa+2)} \; \mathit{and} \\ \mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1 + \frac{2\kappa}{(\kappa+1)[(N+2)\kappa+2]} &= -\frac{2\kappa^2}{(\kappa+1)(2\kappa+N\kappa+2)}. \end{aligned}$$

When $N = 1$, $(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) = 2$ and $vec(\mathbf{I}_N)vec'(\mathbf{I}_N) = 1$, which trivially implies that $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) - \ddot{\mathcal{S}}(\boldsymbol{\phi}) = 0$. However, this result fails to hold for $N \geq 2$. Specifically, using the explicit expressions for the commutation matrix in Magnus (1988), it is straightforward to show that

$$\begin{aligned} &\frac{\kappa^2}{(\kappa+1)(4\kappa+2)}\begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} - \frac{\kappa^2}{(\kappa+1)(2\kappa+1)}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{\kappa^2}{(\kappa+1)(2\kappa+1)} & 0 & 0 & -\frac{\kappa^2}{(\kappa+1)(2\kappa+1)} \\ 0 & \frac{\kappa^2}{(\kappa+1)(2\kappa+1)} & \frac{\kappa^2}{(\kappa+1)(2\kappa+1)} & 0 \\ 0 & \frac{\kappa^2}{(\kappa+1)(2\kappa+1)} & \frac{\kappa^2}{(\kappa+1)(2\kappa+1)} & 0 \\ -\frac{\kappa^2}{(\kappa+1)(2\kappa+1)} & 0 & 0 & \frac{\kappa^2}{(\kappa+1)(2\kappa+1)} \end{pmatrix}, \end{aligned}$$

which can only be 0 under normality. □

A.11 Proposition 11

Note that $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) - \mathring{\mathcal{S}}(\boldsymbol{\phi})$ is $\mathbf{W}_s(\boldsymbol{\phi})\mathbf{W}_s'(\boldsymbol{\phi})$ times the residual variance in the theoretical regression of $\delta(\varsigma_t, \boldsymbol{\eta}_0)\varsigma_t/N - 1$ on $(\varsigma_t/N) - 1$. Therefore, given that $\mathbf{W}_s(\boldsymbol{\phi}) \neq \mathbf{0}$, $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) - \mathring{\mathcal{S}}(\boldsymbol{\phi})$ can only be 0 if that regression residual is identically 0 for all t. The solution to the resulting differential equation is

$$g(\varsigma_t, \boldsymbol{\eta}) = -\frac{N(N+2)\kappa}{2[(N+2)\kappa+2]}\ln\varsigma_t - \frac{1}{[(N+2)\kappa+2]}\varsigma_t + C,$$

which in view of (C19) in Supplemental Appendix C implies that

$$h(\varsigma_t;\boldsymbol{\eta})\propto\varsigma_t^{\frac{N}{(N+2)\kappa+2}-1}\exp\left\{-\frac{1}{[(N+2)\kappa+2]}\varsigma_t\right\},$$

i.e. the density of Gamma random variable with mean N and variance $N[(N+2)\kappa_0+2]$. In this sense, it is worth recalling that $\kappa \geq -2/(N+2)$ for all spherical distributions, with the lower limit corresponding to the uniform.

As for the second part, expression (C27) in Supplemental Appendix C implies that in the spherically symmetric case the difference between $\mathcal{P}(\boldsymbol{\phi}_0)$ and $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)$ is given by

$$\mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}_s'(\boldsymbol{\phi}_0)\cdot\left[\mathrm{M}_{sr}(\boldsymbol{\eta}_0)\mathcal{M}_{rr}^{-1}(\boldsymbol{\eta}_0)\mathrm{M}_{sr}'(\boldsymbol{\eta}_0)\right],$$

which is the product of a rank one matrix times a non-negative scalar. Therefore, given that $\mathbf{W}_s(\boldsymbol{\phi}) \neq \mathbf{0}$ and $\mathcal{M}_{rr}(\boldsymbol{\eta}_0)$ has full rank, $\mathcal{P}(\boldsymbol{\phi}_0)$ can only coincide with $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)$ if the $1\times q$ vector $\mathrm{M}_{sr}(\boldsymbol{\eta}_0)$ is identically 0. □

A.12 Proposition 12

Given our assumptions on the mapping $\mathbf{r}_s(.)$, we can directly work in terms of the $\boldsymbol{\vartheta}$ parameters. In this sense, since the conditional covariance matrix of $\mathbf{y}_t$ is of the form $\vartheta_i\boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_c)$, it is straightforward to show that

$$\mathbf{Z}_{dt}(\boldsymbol{\vartheta}) = \left\{\begin{array}{c} \vartheta_i^{-1/2}\left[\partial\boldsymbol{\mu}_t'(\boldsymbol{\vartheta}_c)/\partial\boldsymbol{\vartheta}_c\right]\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_c) \\ 0 \end{array}\right.$$

$$\left.\begin{array}{c} \frac{1}{2}\{\partial vec'[\Sigma_t^\circ(\boldsymbol{\vartheta}_c)]/\partial\boldsymbol{\vartheta}_c\}[\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_c)\otimes\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_c)] \\ \frac{1}{2}\vartheta_i^{-1}vec'(\mathbf{I}_N) \end{array}\right\} = \left[\begin{array}{cc} \mathbf{Z}_{\vartheta_c lt}(\boldsymbol{\vartheta}) & \mathbf{Z}_{\vartheta_c st}(\boldsymbol{\vartheta}) \\ 0 & \mathbf{Z}_{\vartheta_i st}(\boldsymbol{\vartheta}) \end{array}\right]. \quad \text{(A18)}$$

Thus, the score vector for $\boldsymbol{\vartheta}$ will be

$$\left[\begin{array}{c} \mathbf{s}_{\vartheta_c t}(\boldsymbol{\vartheta},\boldsymbol{\eta}) \\ s_{\vartheta_i t}(\boldsymbol{\vartheta},\boldsymbol{\eta}) \end{array}\right] = \left[\begin{array}{c} \mathbf{Z}_{\vartheta_c lt}(\boldsymbol{\vartheta})\mathbf{e}_{lt}(\boldsymbol{\vartheta},\boldsymbol{\eta}) + \mathbf{Z}_{\vartheta_c st}(\boldsymbol{\vartheta})\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta}) \\ \mathbf{Z}_{\vartheta_i st}(\boldsymbol{\vartheta})\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta}) \end{array}\right], \quad \text{(A19)}$$

where $\mathbf{e}_{lt}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ and $\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ are given in expressions (C8) and (C9) in Supplemental Appendix C, respectively.

It is then easy to see that the unconditional covariance between $\mathbf{s}_{\vartheta_c t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ and $s_{\vartheta_i t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ is

$$
\begin{aligned}
& E\left\{\left[\begin{array}{cc}\mathbf{Z}_{\boldsymbol{\vartheta}_c lt}(\boldsymbol{\vartheta}) & \mathbf{Z}_{\boldsymbol{\vartheta}_c st}(\boldsymbol{\vartheta})\end{array}\right]\left[\begin{array}{cc}\mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathbf{0} \\ \mathbf{0} & \mathcal{M}_{ss}(\boldsymbol{\eta})\end{array}\right]\left[\begin{array}{c}0 \\ \mathbf{Z}'_{\vartheta_i st}(\boldsymbol{\vartheta})\end{array}\right]\right|\boldsymbol{\vartheta},\boldsymbol{\eta}\right\} \\
= & \frac{\{2\mathrm{M}_{ss}(\boldsymbol{\eta})+N[\mathrm{M}_{ss}(\boldsymbol{\eta})-1]\}}{2\vartheta_i} E\left\{\frac{1}{2}\frac{\partial vec'[\boldsymbol{\Sigma}_t^{\circ}(\boldsymbol{\vartheta}_c)]}{\partial \boldsymbol{\vartheta}_c}[\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_c)\otimes \boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_c)]\right|\boldsymbol{\vartheta},\boldsymbol{\eta}\right\} vec(\mathbf{I}_N) \\
= & \frac{\{2\mathrm{M}_{ss}(\boldsymbol{\eta})+N[\mathrm{M}_{ss}(\boldsymbol{\eta})-1]\}}{2\vartheta_i}\mathbf{Z}_{\boldsymbol{\vartheta}_c s}(\boldsymbol{\vartheta},\boldsymbol{\eta}) vec(\mathbf{I}_N),
\end{aligned}
$$

with $\mathbf{Z}_{\boldsymbol{\vartheta}_c s}(\boldsymbol{\vartheta},\boldsymbol{\eta}) = E[\mathbf{Z}_{\boldsymbol{\vartheta}_c st}(\boldsymbol{\vartheta})|\boldsymbol{\vartheta},\boldsymbol{\eta}]$, where we have exploited the serial independence of $\boldsymbol{\varepsilon}_t^*$, as well as the law of iterated expectations, together with the results in Proposition C1 in Supplemental Appendix C.

We can use the same arguments to show that the unconditional variance of $s_{\vartheta_i t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ will be given by

$$
\begin{aligned}
& E\left\{\left[\begin{array}{cc}0 & \mathbf{Z}_{\vartheta_i st}(\boldsymbol{\vartheta})\end{array}\right]\left[\begin{array}{cc}\mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathbf{0} \\ \mathbf{0} & \mathcal{M}_{ss}(\boldsymbol{\eta})\end{array}\right]\left[\begin{array}{c}0 \\ \mathbf{Z}'_{\vartheta_i st}(\boldsymbol{\vartheta})\end{array}\right]\right|\boldsymbol{\vartheta},\boldsymbol{\eta}\right\} \\
= & \frac{1}{4\vartheta_i^2} vec'(\mathbf{I}_N)[\mathrm{M}_{ss}(\boldsymbol{\eta})\,(\mathbf{I}_{N^2}+\mathbf{K}_{NN})+[\mathrm{M}_{ss}(\boldsymbol{\eta})-1])vec(\mathbf{I}_N)vec'(\mathbf{I}_N)]vec(\mathbf{I}_N) \\
= & \frac{\{2\mathrm{M}_{ss}(\boldsymbol{\eta})+N[\mathrm{M}_{ss}(\boldsymbol{\eta})-1]\}N}{4\vartheta_i^2}.
\end{aligned}
$$

Hence, the residuals from the unconditional regression of $\mathbf{s}_{\boldsymbol{\vartheta}_c t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ on $s_{\vartheta_i t}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ will be:

$$
\begin{aligned}
\mathbf{s}_{\vartheta_1|\vartheta_i t}(\boldsymbol{\vartheta},\boldsymbol{\eta}) &= \mathbf{Z}_{\boldsymbol{\vartheta}_c lt}(\boldsymbol{\vartheta})\mathbf{e}_{lt}(\boldsymbol{\vartheta},\boldsymbol{\eta}) + \mathbf{Z}_{\boldsymbol{\vartheta}_c st}(\boldsymbol{\vartheta})\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta}) \\
-\frac{4\vartheta_i^2}{\{2\mathrm{M}_{ss}(\boldsymbol{\eta})+N[\mathrm{M}_{ss}(\boldsymbol{\eta})\text{-}1]\}N} & \frac{\{2\mathrm{M}_{ss}(\boldsymbol{\eta})+N[\mathrm{M}_{ss}(\boldsymbol{\eta})\text{-}1]\}}{2\vartheta_i}\mathbf{Z}_{\boldsymbol{\vartheta}_c s}(\boldsymbol{\vartheta})vec(\mathbf{I}_N)\frac{1}{2\vartheta_i}vec'(\mathbf{I}_N)\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta}) \\
&= \mathbf{Z}_{\boldsymbol{\vartheta}_c lt}(\boldsymbol{\vartheta})\mathbf{e}_{lt}(\boldsymbol{\vartheta},\boldsymbol{\eta}) + [\mathbf{Z}_{\boldsymbol{\vartheta}_c st}(\boldsymbol{\vartheta}) - \mathbf{Z}_{\boldsymbol{\vartheta}_c s}(\boldsymbol{\vartheta},\boldsymbol{\eta})]\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta}).
\end{aligned}
$$

The first term of $\mathbf{s}_{\vartheta_c|\vartheta_i t}(\boldsymbol{\vartheta}_0,\boldsymbol{\eta}_0)$ is clearly conditionally orthogonal to any function of $\varsigma_t(\boldsymbol{\vartheta}_0)$. In contrast, the second term is not conditionally orthogonal to functions of $\varsigma_t(\boldsymbol{\vartheta}_0)$, but since the conditional covariance between any such function and $\mathbf{e}_{st}(\boldsymbol{\vartheta}_0,\boldsymbol{\eta}_0)$ will be time-invariant, it will be unconditionally orthogonal by the law of iterated expectations. As a result, $\mathbf{s}_{\boldsymbol{\vartheta}_c|\vartheta_i t}(\boldsymbol{\vartheta}_0,\boldsymbol{\eta}_0)$ will be unconditionally orthogonal to the spherically symmetric tangent set, which in turn implies that the spherically symmetric semiparametric estimator of $\boldsymbol{\vartheta}_c$ will be ϑ_i-adaptive.

To prove Part 1b, note that Proposition C3 in Supplemental Appendix C and (A18) imply that the spherically symmetric semiparametric efficient score corresponding to ϑ_i will be

$$
\begin{aligned}
\mathring{s}_{\vartheta_i t}(\boldsymbol{\vartheta}) &= -\frac{1}{2\vartheta_i}vec'(\mathbf{I}_N)vec\left\{\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]\boldsymbol{\varepsilon}_t^*(\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\vartheta})-\mathbf{I}_N\right\} \\
&\quad -\frac{N}{2\vartheta_i}\left\{\left[\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]\frac{\varsigma_t(\boldsymbol{\vartheta})}{N}-1\right]-\frac{2}{(N+2)\kappa+2}\left[\frac{\varsigma_t(\boldsymbol{\vartheta})}{N}-1\right]\right\} \\
&= \frac{1}{2\vartheta_i}\{\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]\varsigma_t(\boldsymbol{\vartheta})-N\} - \frac{N}{2\vartheta_i}\left\{\left[\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]\frac{\varsigma_t(\boldsymbol{\vartheta})}{N}-1\right]-\frac{2}{(N+2)\kappa+2}\left[\frac{\varsigma_t(\boldsymbol{\vartheta})}{N}-1\right]\right\} \\
&= \frac{N}{\vartheta_i[(N+2)\kappa+2]}\left[\frac{\varsigma_t(\boldsymbol{\vartheta})}{N}-1\right].
\end{aligned}
$$

But since the iterated spherically symmetric semiparametric estimator of $\boldsymbol{\vartheta}$ must set to 0 the sample average of this modified score, it must be the case that $\sum_{t=1}^{T}\varsigma_t(\mathring{\boldsymbol{\vartheta}}_T)=\sum_{t=1}^{T}\varsigma_t^{\circ}(\mathring{\boldsymbol{\vartheta}}_{cT})/\mathring{\vartheta}_{iT}=NT$, which is equivalent to (12).

To prove Part 1c note that

$$\mathbf{s}_{\vartheta_i t}(\boldsymbol{\vartheta},\mathbf{0})=\frac{1}{2\vartheta_i}\left[\varsigma_t(\boldsymbol{\vartheta})-N\right] \tag{A20}$$

is proportional to the spherically symmetric semiparametric efficient score $\mathring{s}_{\vartheta_i t}(\boldsymbol{\vartheta})$, which means that the residual covariance matrix in the theoretical regression of this efficient score on the Gaussian score will have rank $p-1$ at most. But this residual covariance matrix coincides with $\mathring{\mathcal{S}}(\boldsymbol{\phi})-\mathcal{A}(\boldsymbol{\phi})\mathcal{B}^{-1}(\boldsymbol{\phi})\mathcal{A}(\boldsymbol{\phi})$ since

$$E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}]=E[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathbf{Z}'_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}]=\mathcal{A}(\boldsymbol{\theta}) \tag{A21}$$

because the regression residual

$$\left[\delta(\varsigma_t,\boldsymbol{\eta})\frac{\varsigma_t}{N}-1\right]-\frac{2}{(N+2)\kappa_0+2}\left(\frac{\varsigma_t}{N}-1\right)$$

is conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$ by the law of iterated expectations, as shown in the proof of proposition C3 in Supplemental Appendix C.

Tedious algebraic manipulations that exploit the block-triangularity of (A18) and the constancy of $\mathbf{Z}_{\vartheta_i st}(\boldsymbol{\vartheta})$ show that the different information matrices will be block diagonal when $\mathbf{W}_{\boldsymbol{\vartheta}_c s}(\boldsymbol{\phi}_0)$ is 0. Then, part 2a follows from the fact that $\mathbf{W}_{\boldsymbol{\vartheta}_c s}(\boldsymbol{\phi}_0)=-E\left\{\partial d_t(\boldsymbol{\vartheta}_0)/\partial\boldsymbol{\vartheta}_c|\,\boldsymbol{\phi}_0\right\}$ will trivially be 0 if $E[\ln|\boldsymbol{\Sigma}_t^{\circ}(\boldsymbol{\vartheta}_c)||\boldsymbol{\phi}_0]=k\ \forall\boldsymbol{\vartheta}_c$.

Finally, to prove Part 2b note that (A20) implies that the Gaussian PMLE will also satisfy (12). But since the asymptotic covariance matrices in both cases will be block-diagonal between $\boldsymbol{\vartheta}_c$ and ϑ_i when $E[\ln|\boldsymbol{\Sigma}_t^{\circ}(\boldsymbol{\vartheta}_c)||\boldsymbol{\phi}_0]=k\ \forall\boldsymbol{\vartheta}_c$, the effect of estimating $\boldsymbol{\vartheta}_c$ becomes irrelevant. □

A.13 Proposition 13

We can directly work in terms of the $\boldsymbol{\varphi}$ parameters thanks to our assumptions on the mapping $\mathbf{r}_g(.)$. Given the specification for the conditional mean and variance in (14), and the fact that $\boldsymbol{\varepsilon}_t^*$ is assumed to be *i.i.d.* conditional on $\mathbf{z}_t$ and I_{t-1}, it is tedious but otherwise straightforward to show that the score vector will be

$$\begin{bmatrix}\mathbf{s}_{\boldsymbol{\varphi}_1 t}(\boldsymbol{\varphi},\boldsymbol{\varrho})\\ \mathbf{s}_{\boldsymbol{\varphi}_{ic} t}(\boldsymbol{\varphi},\boldsymbol{\varrho})\\ \mathbf{s}_{\boldsymbol{\varphi}_{im} t}(\boldsymbol{\varphi},\boldsymbol{\varrho})\end{bmatrix}=\begin{bmatrix}\mathbf{Z}_{\boldsymbol{\varphi}_1 lt}(\boldsymbol{\varphi})\mathbf{e}_{lt}(\boldsymbol{\varphi},\boldsymbol{\varrho})+\mathbf{Z}_{\boldsymbol{\varphi}_1 st}(\boldsymbol{\varphi})\mathbf{e}_{st}(\boldsymbol{\varphi},\boldsymbol{\varrho})\\ \mathbf{Z}_{\boldsymbol{\varphi}_{ic} st}(\boldsymbol{\varphi})\mathbf{e}_{st}(\boldsymbol{\varphi},\boldsymbol{\varrho})\\ \mathbf{Z}_{\boldsymbol{\varphi}_{im} lt}(\boldsymbol{\varphi})\mathbf{e}_{lt}(\boldsymbol{\varphi},\boldsymbol{\varrho})\end{bmatrix}, \tag{A22}$$

where

$$\left.\begin{array}{l}\mathbf{Z}_{\boldsymbol{\varphi}_1 lt}(\boldsymbol{\varphi})=\left\{\partial\boldsymbol{\mu}_t^{\diamond\prime}(\boldsymbol{\varphi}_1)/\partial\boldsymbol{\varphi}_1+\partial vec'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\varphi}_1)]/\partial\boldsymbol{\varphi}_1\cdot(\boldsymbol{\varphi}_{im}\otimes\mathbf{I}_N)\right\}\boldsymbol{\Sigma}_t^{\diamond-1/2\prime}(\boldsymbol{\varphi}_1)\boldsymbol{\Phi}_2^{-1/2\prime},\\ \mathbf{Z}_{\boldsymbol{\varphi}_1 st}(\boldsymbol{\varphi})=\partial vec'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\varphi}_1)]/\partial\boldsymbol{\varphi}_1\cdot[\boldsymbol{\Phi}_2^{1/2}\otimes\boldsymbol{\Sigma}_t^{\diamond-1/2\prime}(\boldsymbol{\varphi}_1)\boldsymbol{\Phi}_2^{-1/2\prime}],\\ \mathbf{Z}_{\boldsymbol{\varphi}_{im} lt}(\boldsymbol{\varphi})=\boldsymbol{\Phi}_2^{-1/2\prime}=\mathbf{Z}_{\boldsymbol{\varphi}_{im} l}(\boldsymbol{\varphi}),\\ \mathbf{Z}_{\boldsymbol{\varphi}_{ic} st}(\boldsymbol{\varphi})=\partial vec'(\boldsymbol{\Phi}^{1/2})/\partial\boldsymbol{\varphi}_{ic}\cdot(\mathbf{I}_N\otimes\boldsymbol{\Phi}_2^{-1/2\prime})=\mathbf{Z}_{\boldsymbol{\varphi}_{ic} s}(\boldsymbol{\varphi}),\end{array}\right\} \tag{A23}$$

$\mathbf{e}_{lt}(\boldsymbol{\varphi},\boldsymbol{\varrho})$ and $\mathbf{e}_{st}(\boldsymbol{\varphi},\boldsymbol{\varrho})$ are given in (D4) in Supplemental Appendix D, with

$$\boldsymbol{\varepsilon}_t^*(\boldsymbol{\varphi}) = \boldsymbol{\Phi}_{ic}^{-1/2}\boldsymbol{\Sigma}_t^{\diamond-1/2}(\boldsymbol{\varphi}_c)[\mathbf{y}_t - \boldsymbol{\mu}_t^{\diamond}(\boldsymbol{\varphi}_c) - \boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\varphi}_c)\boldsymbol{\varphi}_{im}]. \tag{A24}$$

It is then easy to see that the unconditional covariance between $\mathbf{s}_{\boldsymbol{\varphi}_c t}(\boldsymbol{\varphi},\boldsymbol{\varrho})$ and the remaining elements of the score will be given by

$$\left[\begin{array}{cc} \mathbf{Z}_{\boldsymbol{\varphi}_c l}(\boldsymbol{\varphi},\boldsymbol{\varrho}) & \mathbf{Z}_{\boldsymbol{\varphi}_c s}(\boldsymbol{\varphi},\boldsymbol{\varrho}) \end{array}\right]\left[\begin{array}{cc} \mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho}) \\ \mathcal{M}_{ls}'(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho}) \end{array}\right]\left[\begin{array}{cc} \mathbf{0} & \mathbf{Z}_{\boldsymbol{\varphi}_{im} l}'(\boldsymbol{\varphi}) \\ \mathbf{Z}_{\boldsymbol{\varphi}_{ic} s}'(\boldsymbol{\varphi}) & \mathbf{0} \end{array}\right]$$

with $\mathbf{Z}_{\boldsymbol{\varphi}_c l}(\boldsymbol{\varphi},\boldsymbol{\varrho}) = E[\mathbf{Z}_{\boldsymbol{\varphi}_c lt}(\boldsymbol{\varphi})|\boldsymbol{\varphi},\boldsymbol{\varrho}]$ and $\mathbf{Z}_{\boldsymbol{\varphi}_c s}(\boldsymbol{\varphi},\boldsymbol{\varrho}) = E[\mathbf{Z}_{\boldsymbol{\varphi}_c st}(\boldsymbol{\varphi})|\boldsymbol{\varphi},\boldsymbol{\varrho}]$, where we have exploited the serial independence of $\boldsymbol{\varepsilon}_t^*$ and the constancy of $\mathbf{Z}_{\boldsymbol{\varphi}_{ic} st}(\boldsymbol{\varphi})$ and $\mathbf{Z}_{\boldsymbol{\varphi}_{im} lt}(\boldsymbol{\varphi})$, together with the law of iterated expectations and the definition

$$\left[\begin{array}{cc} \mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho}) \\ \mathcal{M}_{ls}'(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho}) \end{array}\right] = V\left[\left.\begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\varphi},\boldsymbol{\varrho}) \\ \mathbf{e}_{st}(\boldsymbol{\varphi},\boldsymbol{\varrho}) \end{array}\right|\boldsymbol{\varphi},\boldsymbol{\varrho}\right].$$

Similarly, the unconditional covariance matrix of $\mathbf{s}_{\boldsymbol{\varphi}_{ic} t}(\boldsymbol{\varphi},\boldsymbol{\varrho})$ and $\mathbf{s}_{\boldsymbol{\varphi}_{im} t}(\boldsymbol{\varphi},\boldsymbol{\varrho})$ will be

$$\left[\begin{array}{cc} \mathbf{0} & \mathbf{Z}_{\boldsymbol{\varphi}_{ic} s}(\boldsymbol{\varphi}) \\ \mathbf{Z}_{\boldsymbol{\varphi}_{im} l}(\boldsymbol{\varphi}) & \mathbf{0} \end{array}\right]\left[\begin{array}{cc} \mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho}) \\ \mathcal{M}_{ls}'(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho}) \end{array}\right]\left[\begin{array}{cc} \mathbf{0} & \mathbf{Z}_{\boldsymbol{\varphi}_{im} l}'(\boldsymbol{\varphi}) \\ \mathbf{Z}_{\boldsymbol{\varphi}_{ic} s}'(\boldsymbol{\varphi}) & \mathbf{0} \end{array}\right].$$

Thus, the residuals from the unconditional least squares projection of $\mathbf{s}_{\boldsymbol{\varphi}_c t}(\boldsymbol{\varphi},\boldsymbol{\varrho})$ on $\mathbf{s}_{\boldsymbol{\varphi}_{ic} t}(\boldsymbol{\varphi},\boldsymbol{\varrho})$ and $\mathbf{s}_{\boldsymbol{\varphi}_{im} t}(\boldsymbol{\varphi},\boldsymbol{\varrho})$ will be:

$$\begin{aligned}
\mathbf{s}_{\boldsymbol{\varphi}_c|\boldsymbol{\varphi}_{ic},\boldsymbol{\varphi}_{im} t}(\boldsymbol{\varphi},\boldsymbol{\varrho}) &= \mathbf{Z}_{\boldsymbol{\varphi}_c lt}(\boldsymbol{\varphi})\mathbf{e}_{lt}(\boldsymbol{\varphi},\boldsymbol{\varrho}) + \mathbf{Z}_{\boldsymbol{\varphi}_c st}(\boldsymbol{\varphi})\mathbf{e}_{st}(\boldsymbol{\varphi},\boldsymbol{\varrho}) \\
&\quad - \left[\begin{array}{cc} \mathbf{Z}_{\boldsymbol{\varphi}_c l}(\boldsymbol{\varphi},\boldsymbol{\varrho}) & \mathbf{Z}_{\boldsymbol{\varphi}_c s}(\boldsymbol{\varphi},\boldsymbol{\varrho}) \end{array}\right]\left[\begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\varphi},\boldsymbol{\varrho}) \\ \mathbf{e}_{st}(\boldsymbol{\varphi},\boldsymbol{\varrho}) \end{array}\right] \\
&= [\mathbf{Z}_{\boldsymbol{\varphi}_c lt}(\boldsymbol{\varphi}) - \mathbf{Z}_{\boldsymbol{\varphi}_c l}(\boldsymbol{\varphi},\boldsymbol{\varrho})]\mathbf{e}_{lt}(\boldsymbol{\varphi},\boldsymbol{\varrho}) + [\mathbf{Z}_{\boldsymbol{\varphi}_c st}(\boldsymbol{\varphi}) - \mathbf{Z}_{\boldsymbol{\varphi} cs}(\boldsymbol{\varphi},\boldsymbol{\varrho})]\mathbf{e}_{st}(\boldsymbol{\varphi},\boldsymbol{\varrho}),
\end{aligned}$$

because both $\mathbf{Z}_{\boldsymbol{\varphi}_{ic} s}(\boldsymbol{\varphi})$ and $\mathbf{Z}_{\boldsymbol{\varphi}_{im} l}(\boldsymbol{\varphi})$ have full row rank when $\boldsymbol{\Phi}_{ic}$ has full rank in view of the discussion that follows expression (D13) in Supplemental Appendix D.

Although neither $\mathbf{e}_{lt}(\boldsymbol{\varphi},\boldsymbol{\varrho})$ nor $\mathbf{e}_{st}(\boldsymbol{\varphi},\boldsymbol{\varrho})$ will be conditionally orthogonal to arbitrary functions of $\boldsymbol{\varepsilon}_t^*$, their conditional covariance with any such function will be time-invariant. Hence, $\mathbf{s}_{\boldsymbol{\varphi}_c|\boldsymbol{\varphi}_{ic},\boldsymbol{\varphi}_{im} t}(\boldsymbol{\varphi},\boldsymbol{\varrho})$ will be unconditionally orthogonal to $\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\varphi});\boldsymbol{\varrho}]/\partial\boldsymbol{\varrho}$ by virtue of the law of iterated expectations, which in turn implies that the unrestricted semiparametric estimator of $\boldsymbol{\varphi}_c$ will be $\boldsymbol{\varphi}_i$-adaptive.

To prove Part 1b note that the semiparametric efficient scores corresponding to $\boldsymbol{\varphi}_{ic}$ and $\boldsymbol{\varphi}_{im}$ will be given by

$$\left[\begin{array}{cc} \mathbf{0} & \mathbf{Z}_{\boldsymbol{\varphi}_{ic} s}(\boldsymbol{\varphi}) \\ \mathbf{Z}_{\boldsymbol{\varphi}_{im} l}(\boldsymbol{\varphi}) & \mathbf{0} \end{array}\right]\mathcal{K}(0)\,\mathcal{K}^+(\boldsymbol{\varrho}_0)\left\{\begin{array}{c} \boldsymbol{\varepsilon}_t^*(\boldsymbol{\varphi}) \\ vec[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\varphi})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\varphi}) - \mathbf{I}_N] \end{array}\right\}$$

because $\mathbf{Z}_{\boldsymbol{\varphi}_{ic} st}(\boldsymbol{\vartheta}) = \mathbf{Z}_{\boldsymbol{\varphi}_{ic} s}(\boldsymbol{\vartheta})$ and $\mathbf{Z}_{\boldsymbol{\varphi}_{im} lt}(\boldsymbol{\vartheta}) = \mathbf{Z}_{\boldsymbol{\varphi}_{im} l}(\boldsymbol{\vartheta})$ $\forall t$. But if (17) and (16) hold, then the sample averages of $\mathbf{e}_{lt}[\boldsymbol{\varphi}_c, \boldsymbol{\varphi}_{ic}(\boldsymbol{\varphi}_c), \boldsymbol{\varphi}_{im}(\boldsymbol{\varphi}_c); \mathbf{0}]$ and $\mathbf{e}_{st}[\boldsymbol{\varphi}_c, \boldsymbol{\varphi}_{ic}(\boldsymbol{\varphi}_c), \boldsymbol{\varphi}_{im}(\boldsymbol{\varphi}_c); \mathbf{0}]$ will be 0, and the same is true of the semiparametric efficient score.

To prove Part 1c note that

$$\left[\begin{array}{c} \mathbf{s}_{\boldsymbol{\varphi}_{ic}t}(\boldsymbol{\varphi},\mathbf{0}) \\ \mathbf{s}_{\boldsymbol{\varphi}_{im}t}(\boldsymbol{\varphi},\mathbf{0}) \end{array}\right] = \left[\begin{array}{cc} \mathbf{0} & \mathbf{Z}_{\boldsymbol{\varphi}_{ic}s}(\boldsymbol{\varphi}) \\ \mathbf{Z}_{\boldsymbol{\varphi}_{im}l}(\boldsymbol{\varphi}) & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \boldsymbol{\varepsilon}_t^*(\boldsymbol{\varphi}) \\ vec[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\varphi})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\varphi}) - \mathbf{I}_N] \end{array}\right], \quad \text{(A25)}$$

which implies that the residual covariance matrix in the theoretical regression of the semiparametric efficient score on the Gaussian score will have rank $p-N(N+3)/2$ at most because both $\mathbf{Z}_{\boldsymbol{\varphi}_{ic}s}(\boldsymbol{\varphi})$ and $\mathbf{Z}_{\boldsymbol{\varphi}_{im}l}(\boldsymbol{\varphi})$ have full row rank when $\boldsymbol{\Phi}_{ic}$ has full rank. But as we saw in the proof of Proposition 5, that residual covariance matrix coincides with $\ddot{\mathcal{S}}(\boldsymbol{\phi}_0) - \mathcal{A}(\boldsymbol{\theta})\mathcal{B}^{-1}(\boldsymbol{\phi})\mathcal{A}(\boldsymbol{\theta})$.

Tedious algebraic manipulations that exploit the block structure of (A23) and the constancy of $\mathbf{Z}_{\boldsymbol{\varphi}_{ic}st}(\boldsymbol{\varphi})$ and $\mathbf{Z}_{\boldsymbol{\varphi}_{im}lt}(\boldsymbol{\varphi})$ show that the different information matrices will be block diagonal when $\mathbf{Z}_{\boldsymbol{\varphi}_c l}(\boldsymbol{\varphi},\boldsymbol{\varrho})$ and $\mathbf{Z}_{\boldsymbol{\varphi}_c s}(\boldsymbol{\varphi},\boldsymbol{\varrho})$ are both 0. But those are precisely the necessary and sufficient conditions for $\mathbf{s}_{\boldsymbol{\varphi}_c t}(\boldsymbol{\varphi},\boldsymbol{\varrho})$ to be equal to $\mathbf{s}_{\boldsymbol{\varphi}_c|\boldsymbol{\varphi}_{ic},\boldsymbol{\varphi}_{im}t}(\boldsymbol{\varphi},\boldsymbol{\varrho})$, which is also guaranteed by two conditions in the statement of part 2. In this sense, please note that the reparametrisation of $\boldsymbol{\varphi}_{ic}$ and $\boldsymbol{\varphi}_{im}$ that satisfies those conditions will be such that the Jacobian matrix of $vech[\mathbf{K}^{-1/2}(\boldsymbol{\varphi}_c)\boldsymbol{\Phi}_{ic}\mathbf{K}^{-1/2\prime}(\boldsymbol{\varphi}_c)]$ and $\mathbf{K}^{-1/2}(\boldsymbol{\varphi}_c)\boldsymbol{\varphi}_{im} - \mathbf{l}(\boldsymbol{\varphi}_c)$ with respect to $\boldsymbol{\varphi}$ evaluated at the true values is equal to

$$\left\{ -V^{-1}\left[\begin{array}{c} \mathbf{s}_{\boldsymbol{\varphi}_{ic}t}(\boldsymbol{\varphi}_0) \\ \mathbf{s}_{\boldsymbol{\varphi}_{im}t}(\boldsymbol{\varphi}_0) \end{array}\middle|\boldsymbol{\phi}_0\right] E\left[\begin{array}{c} \mathbf{s}_{\boldsymbol{\varphi}_{ic}t}(\boldsymbol{\varphi}_0)\mathbf{s}'_{\boldsymbol{\varphi}_{ct}}(\boldsymbol{\varphi}_0) \\ \mathbf{s}_{\boldsymbol{\varphi}_{im}t}(\boldsymbol{\varphi}_0)\mathbf{s}'_{\boldsymbol{\varphi}_{ct}}(\boldsymbol{\varphi}_0) \end{array}\middle|\boldsymbol{\phi}_0\right] \middle| \begin{array}{c} \mathbf{I}_{N(N+1)/2} \\ \mathbf{0} \end{array} \middle| \begin{array}{c} \mathbf{0} \\ \mathbf{I}_N \end{array} \right\}.$$

Finally, to prove Part 2b simply note that (A25) implies the Gaussian PMLE will also satisfy (17) and (16). But since the asymptotic covariance matrices in both cases will be block-diagonal between $\boldsymbol{\varphi}_c$ and $\boldsymbol{\varphi}_i$ when the two conditions in the statement of part 2 hold, the effect of estimating $\boldsymbol{\varphi}_c$ becomes irrelevant. □

A.14 Proposition 14

The proof builds up on Proposition B1 in Supplemental Appendix B. Assuming mean stationarity, the relationship vector of drift parameters $\boldsymbol{\tau}$ and the unconditional mean $\boldsymbol{\mu}$ is given by $(\mathbf{I}_N - \mathbf{A}_1 - \ldots - \mathbf{A}_p)\boldsymbol{\mu}$. Hence, the Jacobian from one vector of parameters to the other is

$$\frac{\partial\left(\begin{array}{c}\boldsymbol{\tau}\\ \mathbf{a}\end{array}\right)}{\partial(\boldsymbol{\mu}',\mathbf{a}')} = \left(\begin{array}{cccc} \mathbf{I}_N - \mathbf{A}_1 - \ldots - \mathbf{A}_p & -\boldsymbol{\mu}'\otimes\mathbf{I}_N & \ldots & -\boldsymbol{\mu}'\otimes\mathbf{I}_N \\ \mathbf{0} & \mathbf{I}_{N^2} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{I}_{N^2} \end{array}\right).$$

Consequently, $\mathbf{Z}_{lt}(\boldsymbol{\theta})$ for $(\boldsymbol{\mu}',\mathbf{a}',\mathbf{c}')$ becomes

$$\left(\begin{array}{c} (\mathbf{I}_N - \mathbf{A}_1 - \ldots - \mathbf{A}_p)\mathbf{C}^{-1\prime} \\ (\mathbf{y}_{t-1}-\boldsymbol{\mu})\otimes\mathbf{C}^{-1\prime} \\ \vdots \\ (\mathbf{y}_{t-p}-\boldsymbol{\mu})\otimes\mathbf{C}^{-1\prime} \\ \mathbf{0}_{N^2\times N} \end{array}\right),$$

so that

$$\mathcal{I}_{\boldsymbol{\mu}\boldsymbol{\mu}} = (\mathbf{I}_N - \mathbf{A}_1 - \ldots - \mathbf{A}_p)\mathbf{C}^{-1\prime}\mathcal{M}_{ll}\mathbf{C}^{-1}(\mathbf{I}_N - \mathbf{A}_1 - \ldots - \mathbf{A}_p)',$$

$$\mathcal{I}_{\mathbf{aa}} = \begin{bmatrix} \boldsymbol{\Gamma}(0) & \ldots & \boldsymbol{\Gamma}(p-1) \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}'(p-1) & \ldots & \boldsymbol{\Gamma}(0) \end{bmatrix} \otimes \mathbf{C}^{-1\prime}\mathcal{M}_{ll}\mathbf{C}^{-1},$$

and $\mathcal{I}_{\boldsymbol{\mu}\mathbf{a}} = \mathbf{0}$. Consequently, the asymptotic variances of the restricted and unrestricted ML estimators of $\boldsymbol{\mu}$ and $\mathbf{a}$ will be given by

$$\mathcal{I}^{-1}_{\boldsymbol{\mu}\boldsymbol{\mu}} = (\mathbf{I}_N - \mathbf{A}_1 - \ldots - \mathbf{A}_p)^{-1\prime}\mathbf{C}\mathcal{M}^{-1}_{ll}\mathbf{C}'(\mathbf{I}_N - \mathbf{A}_1 - \ldots - \mathbf{A}_p)^{-1},$$

$$\mathcal{I}^{-1}_{\mathbf{aa}} = \begin{bmatrix} \boldsymbol{\Gamma}(0) & \ldots & \boldsymbol{\Gamma}(p-1) \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}'(p-1) & \ldots & \boldsymbol{\Gamma}(0) \end{bmatrix}^{-1} \otimes \mathbf{C}\mathcal{M}^{-1}_{ll}\mathbf{C}',$$

where $\boldsymbol{\Gamma}(j)$ is the j^{th} autocovariance matrix of $\mathbf{y}_t$.

Let us now look at the conditional variance parameters. The product rule for differentials $d\mathbf{C} = (d\mathbf{J})\boldsymbol{\Psi} + \mathbf{J}(d\boldsymbol{\Psi})$ immediately implies that

$$dvec(\mathbf{C}) = (\boldsymbol{\Psi} \otimes \mathbf{I}_N)\boldsymbol{\Delta}_N dveco(\mathbf{J}) + (\mathbf{I}_N \otimes \mathbf{J})\mathbf{E}_N dveco(\boldsymbol{\Psi}),$$

where $\mathbf{E}_N$ is the $N^2 \times N$ matrix such that $vec(\boldsymbol{\Psi}) = \mathbf{E}_N vecd(\boldsymbol{\Psi})$ for any diagonal matrix $\boldsymbol{\Psi}$, where $vecd(\boldsymbol{\Psi})$ places the elements in the main diagonal of $\boldsymbol{\Psi}$ in a column vector, and $\boldsymbol{\Delta}_N$ is an $N^2 \times N(N-1)$ matrix such that $vec(\mathbf{J} - \mathbf{I}_N) = \boldsymbol{\Delta}_N veco(\mathbf{J} - \mathbf{I}_N)$, with $veco(\mathbf{J} - \mathbf{I}_N)$ stacking by columns all the elements of the zero-diagonal matrix $\mathbf{J} - \mathbf{I}_N$ except those that appear in its diagonal. Therefore, the Jacobian will be

$$\frac{\partial vec(\mathbf{C})}{\partial(\mathbf{j}', \boldsymbol{\psi}')} = \begin{bmatrix} (\boldsymbol{\Psi} \otimes \mathbf{I}_N)\boldsymbol{\Delta}_N & (\mathbf{I}_N \otimes \mathbf{J})\mathbf{E}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Delta}_N(\boldsymbol{\Psi} \otimes \mathbf{I}_{N-1}) & (\mathbf{I}_N \otimes \mathbf{J})\mathbf{E}_N \end{bmatrix}, \quad \text{(A26)}$$

where we have used that $\boldsymbol{\Upsilon}\boldsymbol{\Delta}_N = \boldsymbol{\Delta}_N(\boldsymbol{\Delta}'_N\boldsymbol{\Upsilon}\boldsymbol{\Delta}_N)$ for any diagonal matrix $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Delta}'_N(\boldsymbol{\Psi} \otimes \mathbf{I}_N)\boldsymbol{\Delta}_N = (\boldsymbol{\Psi} \otimes \mathbf{I}_{N-1})$ (see Proposition 6 in Magnus and Sentana (2020)).

As a result, the scores with respect to $\mathbf{j}$ and $\boldsymbol{\psi}$ will be

$$\begin{bmatrix} (\boldsymbol{\Psi} \otimes \mathbf{I}_{N-1})\boldsymbol{\Delta}'_N \\ \mathbf{E}'_N(\mathbf{I}_N \otimes \mathbf{J}') \end{bmatrix} (\mathbf{I}_N \otimes \mathbf{J}^{-1\prime})(\mathbf{I}_N \otimes \boldsymbol{\Psi}^{-1})\mathbf{e}_{st}(\boldsymbol{\phi})$$

$$= \begin{bmatrix} (\boldsymbol{\Psi} \otimes \mathbf{I}_{N-1})\boldsymbol{\Delta}'_N(\mathbf{I}_N \otimes \mathbf{J}^{-1\prime})(\mathbf{I}_N \otimes \boldsymbol{\Psi}^{-1}) \\ \boldsymbol{\Psi}^{-1}\mathbf{E}'_N \end{bmatrix} \mathbf{e}_{st}(\boldsymbol{\phi}) = \begin{bmatrix} \boldsymbol{\Delta}'_N(\mathbf{I}_N \otimes \mathbf{J}^{-1\prime})(\boldsymbol{\Psi} \otimes \boldsymbol{\Psi}^{-1}) \\ \boldsymbol{\Psi}^{-1}\mathbf{E}'_N \end{bmatrix} \mathbf{e}_{st}(\boldsymbol{\phi}).$$

Similarly, the information matrix of the unrestricted ML estimators of $(\mathbf{j}, \boldsymbol{\psi}, \boldsymbol{\varrho})$ will be

$$\left\{\begin{array}{c}\begin{bmatrix}\boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\\ \boldsymbol{\Psi}^{-1}\mathbf{E}_N'\end{bmatrix}\mathcal{M}_{ss}\begin{bmatrix}(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})(\mathbf{I}_N\otimes\mathbf{J}^{-1})\boldsymbol{\Delta}_N & \mathbf{E}_N\boldsymbol{\Psi}^{-1}\end{bmatrix}\\ \mathrm{M}_{sr}'\mathbf{E}_N'\begin{bmatrix}(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})(\mathbf{I}_N\otimes\mathbf{J}^{-1})\boldsymbol{\Delta}_N & \mathbf{E}_N\boldsymbol{\Psi}^{-1}\end{bmatrix}\end{array}\right.$$
$$\left.\begin{array}{c}\begin{bmatrix}\boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\\ \boldsymbol{\Psi}^{-1}\mathbf{E}_N'\end{bmatrix}\mathbf{E}_N\mathrm{M}_{sr}\\ \mathcal{M}_{rr}\end{array}\right\}$$
$$=\left[\begin{array}{c}\boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathcal{M}_{ss}(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})(\mathbf{I}_N\otimes\mathbf{J}^{-1})\boldsymbol{\Delta}_N\\ \boldsymbol{\Psi}^{-1}\mathbf{E}_N'\mathcal{M}_{ss}(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})(\mathbf{I}_N\otimes\mathbf{J}^{-1})\boldsymbol{\Delta}_N\\ \mathrm{M}_{sr}'\mathbf{E}_N'(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})(\mathbf{I}_N\otimes\mathbf{J}^{-1})\boldsymbol{\Delta}_N\end{array}\right.$$
$$\left.\begin{array}{cc}\boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathcal{M}_{ss}\mathbf{E}_N\boldsymbol{\Psi}^{-1} & \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathbf{E}_N\mathrm{M}_{sr}\\ \boldsymbol{\Psi}^{-1}\mathbf{E}_N'\mathcal{M}_{ss}\mathbf{E}_N\boldsymbol{\Psi}^{-1} & \boldsymbol{\Psi}^{-1}\mathbf{E}_N'\mathbf{E}_N\mathrm{M}_{sr}\\ \mathrm{M}_{sr}'\mathbf{E}_N'\mathbf{E}_N\boldsymbol{\Psi}^{-1} & \mathcal{M}_{rr}\end{array}\right]$$
$$=\left[\begin{array}{c}\boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})\mathcal{M}_{ss}(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})(\mathbf{I}_N\otimes\mathbf{J}^{-1})\boldsymbol{\Delta}_N\\ \boldsymbol{\Psi}^{-1}\mathrm{M}_{ss}\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1})\boldsymbol{\Delta}_N\\ \mathrm{M}_{sr}'\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1})\boldsymbol{\Delta}_N\end{array}\right.$$
$$\left.\begin{array}{cc}\boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})\mathbf{E}_N\mathrm{M}_{ss}\boldsymbol{\Psi}^{-1} & \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})\mathbf{E}_N\mathrm{M}_{sr}\\ \boldsymbol{\Psi}^{-1}\mathrm{M}_{ss}\boldsymbol{\Psi}^{-1} & \boldsymbol{\Psi}^{-1}\mathrm{M}_{sr}\\ \mathrm{M}_{sr}'\boldsymbol{\Psi}^{-1} & \mathcal{M}_{rr}\end{array}\right].$$

Let us now obtain the asymptotic covariance matrix of the restricted ML estimators of $(\mathbf{j}, \boldsymbol{\psi})$ which fix $\boldsymbol{\varrho}$ to its true values. Lemmas 4 and 5 contain the inverses of $\mathcal{M}_{ss}$ and $\begin{bmatrix}(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1})(\mathbf{I}_N\otimes\mathbf{J}^{-1})\boldsymbol{\Delta}_N & \mathbf{E}_N\boldsymbol{\Psi}^{-1}\end{bmatrix}$, respectively. Thus, the asymptotic covariance matrix of $(\mathbf{j}, \boldsymbol{\psi})$ will be

$$\left\{\begin{array}{c}\boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})(\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Psi})[\mathbf{I}_{N^2}-\mathbf{E}_N\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})(\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Psi})]\\ \boldsymbol{\Psi}\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})\end{array}\right\}\mathcal{M}_{ss}^{-1}$$
$$\times\left\{[\mathbf{I}_{N^2}-(\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Psi})(\mathbf{I}_N\otimes\mathbf{J}')\mathbf{E}_N\mathbf{E}_N'](\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Psi})(\mathbf{I}_N\otimes\mathbf{J}')\boldsymbol{\Delta}_N \quad (\mathbf{I}_N\otimes\mathbf{J}')\mathbf{E}_N\boldsymbol{\Psi}\right\},$$

which does not have have any special structure, except in the unlikely event that $\mathbf{J}_0 = \mathbf{I}_N$, in which case the inverse in Lemma 5 would reduce to

$$\left\{\begin{array}{c}[\boldsymbol{\Delta}_N'(\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Psi})\boldsymbol{\Delta}_N]\boldsymbol{\Delta}_N'\\ \boldsymbol{\Psi}\mathbf{E}_N'\end{array}\right\},$$

where we have used the fact that $\mathbf{I}_{N^2} - \mathbf{E}_N\mathbf{E}_N' = \boldsymbol{\Delta}_N\boldsymbol{\Delta}_N'$ (see Proposition 4 in Magnus and Sentana (2020)). Tedious algebraic manipulations then show that the asymptotic covariance matrix of the restricted ML estimators of $(\mathbf{j}, \boldsymbol{\psi})$ which fix $\boldsymbol{\varrho}$ to its true values when $\mathbf{J}_0 = \mathbf{I}_N$ would be

$$\left\{\begin{array}{cc}[\boldsymbol{\Delta}_N'(\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Psi})\boldsymbol{\Delta}_N][\boldsymbol{\Delta}_N'(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N]^{-1}[\boldsymbol{\Delta}_N'(\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Psi})\boldsymbol{\Delta}_N] & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Psi}\mathrm{M}_{ss}^{-1}\boldsymbol{\Psi}\end{array}\right\}.$$

The matrix $\boldsymbol{\Delta}_N'(\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Psi})\boldsymbol{\Delta}_N$ is obviously diagonal. In turn, Proposition 5 in Magnus and Sentana (2020) implies that the matrix $\boldsymbol{\Delta}_N'(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N = \boldsymbol{\Delta}_N'\mathbf{K}_{NN}\boldsymbol{\Delta}_N + \boldsymbol{\Delta}_N'\boldsymbol{\Upsilon}\boldsymbol{\Delta}_N$ is the sum of a diagonal matrix $\mathbf{\Delta}_N'\mathbf{\Upsilon}\mathbf{\Delta}_N$ and a symmetric orthogonal matrix $\mathbf{\Delta}_N'\mathbf{K}_{NN}\mathbf{\Delta}_N$ whose only $N(N-1)$ non-zero elements are 1s in the positions corresponding to the ij and ji elements of $\mathbf{J}$ for $j > i$. Therefore, although the parameters in the different columns of $\mathbf{J}$ would not be asymptotically orthogonal when $\mathbf{J}_0 = \mathbf{I}_N$, the dependence seems to be limited to pairs of elements $\{\mathbf{J}\}_{ij}$ and $\{\mathbf{J}\}_{ji}$.

We can follow an analogous procedure to find the asymptotic covariance matrix of the unrestricted ML estimators of $(\mathbf{j}, \boldsymbol{\psi}, \boldsymbol{\varrho})$ for general $\mathbf{J}$, which will be

$$\left\{\begin{array}{cc} \mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})[\mathbf{I}_{N^2}-\mathbf{E}_N\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})] & \mathbf{0} \\ \mathbf{\Psi}\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{array}\right\}$$
$$\times\left[\begin{pmatrix} \mathcal{M}_{ss}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{pmatrix} + \begin{pmatrix} \mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathbf{E}_N' & -\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr} \\ -\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathbf{E}_N' & \mathcal{M}^{rr}\end{pmatrix}\right]$$
$$\times\left\{\begin{array}{ccc} [\mathbf{I}_{N^2}-(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})(\mathbf{I}_N\otimes\mathbf{J}')\mathbf{E}_N\mathbf{E}_N'](\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})(\mathbf{I}_N\otimes\mathbf{J}')\mathbf{\Delta}_N & (\mathbf{I}_N\otimes\mathbf{J}')\mathbf{E}_N\mathbf{\Psi} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_N \end{array}\right\}$$
$$=\left\{\begin{array}{c} \mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})[\mathbf{I}_{N^2}-\mathbf{E}_N\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})] \\ \mathbf{\Psi}\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J}) \\ \mathbf{0} \end{array}\right\}\mathcal{M}_{ss}^{-1}$$
$$\times\left\{[\mathbf{I}_{N^2}-(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})(\mathbf{I}_N\otimes\mathbf{J}')\mathbf{E}_N\mathbf{E}_N'](\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})(\mathbf{I}_N\otimes\mathbf{J}')\mathbf{\Delta}_N \quad (\mathbf{I}_N\otimes\mathbf{J}')\mathbf{E}_N\mathbf{\Psi} \quad \mathbf{0}\right\}$$
$$+\left\{\begin{array}{cc} \mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})[\mathbf{I}_{N^2}-\mathbf{E}_N\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})] & \mathbf{0} \\ \mathbf{\Psi}\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{array}\right\}$$
$$\times\begin{pmatrix} \mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathbf{E}_N' & -\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr} \\ -\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathbf{E}_N' & \mathcal{M}^{rr}\end{pmatrix}$$
$$\times\left\{\begin{array}{ccc} [\mathbf{I}_{N^2}-(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})(\mathbf{I}_N\otimes\mathbf{J}')\mathbf{E}_N\mathbf{E}_N'](\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})(\mathbf{I}_N\otimes\mathbf{J}')\mathbf{\Delta}_N & (\mathbf{I}_N\otimes\mathbf{J}')\mathbf{E}_N\mathbf{\Psi} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_N \end{array}\right\}.$$

Let us look at the second term in the sum. First of all, its northeastern block is

$$-\mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})[\mathbf{I}_{N^2}-\mathbf{E}_N\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})]\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}$$
$$= \mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}$$
$$+\mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})\mathbf{E}_N\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})]\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}$$
$$= \mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr} + \mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})\mathbf{E}_N\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}$$
$$= \mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr} + \mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})\mathbf{E}_N(\mathbf{I}_N\odot\mathbf{J})\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr} = \mathbf{0},$$

and the same applies to the southwestern one by symmetry.

Turning now to the eastern block, we get

$$-\mathbf{\Psi}\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr} = -\mathbf{\Psi}\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr},$$

a diagonal matrix, and by symmetry, the same applies to the southern block. The sourtheastern block is trivially $\mathcal{M}^{rr}$, which is also diagonal.

Let us now focus on the northwestern and western blocks, which are given by

$$\boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})(\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Psi})[\mathbf{I}_{N^2}-\mathbf{E}_N\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})(\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Psi})]\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathbf{E}_N'$$
$$\times[\mathbf{I}_{N^2}-(\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Psi})(\mathbf{I}_N\otimes\mathbf{J}')\mathbf{E}_N\mathbf{E}_N'](\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Psi})(\mathbf{I}_N\otimes\mathbf{J}')\boldsymbol{\Delta}_N \textit{ and}$$
$$\boldsymbol{\Psi}\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathbf{E}_N'$$
$$\times[\mathbf{I}_{N^2}-(\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Psi})(\mathbf{I}_N\otimes\mathbf{J}')\mathbf{E}_N\mathbf{E}_N'](\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Psi})(\mathbf{I}_N\otimes\mathbf{J}')\boldsymbol{\Delta}_N,$$

respectively. Given that the northeastern block is 0, these two blocks will be 0 too. Finally, given that the central block is

$$\boldsymbol{\Psi}\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J}')\mathbf{E}_N\boldsymbol{\Psi}=\boldsymbol{\Psi}\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\boldsymbol{\Psi},$$

the second term in the sum reduces to

$$\begin{pmatrix}\mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Psi}\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\boldsymbol{\Psi} & -\boldsymbol{\Psi}\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\\ \mathbf{0} & -\mathcal{M}^{rr}\mathrm{M}_{sr}\mathrm{M}_{ss}^{-1}\boldsymbol{\Psi} & \mathcal{M}^{rr}\end{pmatrix}. \tag{A27}$$

This expression confirms that the restricted and unrestricted ML estimators of $\mathbf{j}$ are equally efficient because the first term in the sum is a bordered version of the asymptotic covariance matrix of the restricted MLEs of $\mathbf{j}$ and $\boldsymbol{\psi}$.

Expression (A27) also implies that the unrestricted ML estimators of $\mathbf{j}$ and $\boldsymbol{\varrho}$ are asymptotically independent, and that the unrestricted MLEs of $\boldsymbol{\varrho}$ are as efficient as its restricted ML estimators which fix $\mathbf{j}$ to its true value and simultaneously estimate $\boldsymbol{\psi}$ and $\boldsymbol{\varrho}$. In fact, given that the asymptotic covariance matrix of those restricted estimators would be

$$\begin{pmatrix}\boldsymbol{\Psi}[\mathrm{M}_{ss}^{-1}+\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}]\boldsymbol{\Psi} & -\boldsymbol{\Psi}\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\\ -\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\boldsymbol{\Psi} & \mathcal{M}^{rr}\end{pmatrix}, \tag{A28}$$

and that all four blocks are diagonal matrices, it is tedious but otherwise straightforward to prove that each of the diagonal elements of $\mathcal{M}^{rr}$ coincides with the asymptotic variance of the MLE of η_i in a univariate Student t log-likelihood that only estimates this parameter and a scale parameter γ_i.

The comparison between (A27) and (A28) also indicates that the covariance between the ML estimators of $\boldsymbol{\psi}$ and $\boldsymbol{\varrho}$ is the same regardless of whether $\mathbf{j}$ is estimated or not. The same is true of the correction to the asymptotic covariance matrix of $\boldsymbol{\psi}$ resulting from estimating $\boldsymbol{\varrho}$. In contrast, $\boldsymbol{\Psi}\mathrm{M}_{ss}^{-1}\boldsymbol{\Psi}$ and $\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{C})\mathcal{M}_{ss}^{-1}(\mathbf{I}_N\otimes\mathbf{C}')\mathbf{E}_N=\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J}\boldsymbol{\Psi})\mathcal{M}_{ss}^{-1}(\mathbf{I}_N\otimes\boldsymbol{\Psi}\mathbf{J}')\mathbf{E}_N$ do not generally coincide unless $\mathbf{J}_0=\mathbf{I}_N$. □

TABLE 1: Univariate GARCH-M: Empirical rejection rates.

Student t_{12}

	RML=UML $\vartheta_i @(\bar{\boldsymbol{\theta}}_T, \bar{\eta})$				UML=PML $\vartheta_i @(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T)$			UML=PML $(\varphi_{im}, \varphi_{ic}) @(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T)$			RML=UML & UML=PML $\vartheta_i @(\bar{\boldsymbol{\theta}}_T, \bar{\eta}_T)$		
%	DWH1	DWH2	DWH3	LR	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3
1	9.64	14.50	0.95	1.01	1.65	0.81	1.68	1.94	1.25	1.90	8.96	13.87	1.96
5	15.56	18.73	4.82	5.15	4.98	4.32	5.65	6.12	5.56	6.57	14.37	18.98	4.95
10	20.08	21.55	9.93	10.32	9.45	8.68	9.92	11.35	10.71	11.77	18.65	22.71	8.85

Student t_8

	RML=UML $\vartheta_i @(\bar{\boldsymbol{\theta}}_T, \bar{\eta})$				UML=PML $\vartheta_i @(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T)$			UML=PML $(\varphi_{im}, \varphi_{ic}) @(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T)$			RML=UML & UML=PML $\vartheta_i @(\bar{\boldsymbol{\theta}}_T, \bar{\eta}_T)$		
%	DWH1	DWH2	DWH3	LR	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3
1	40.78	32.30	38.30	30.92	1.88	0.80	3.03	2.34	1.34	3.02	41.23	34.57	37.69
5	50.75	38.68	57.58	53.15	5.24	3.99	6.96	6.67	5.97	8.20	51.59	42.59	54.26
10	56.66	42.63	67.20	64.62	9.49	8.62	10.88	11.54	10.95	13.24	58.12	47.99	63.44

GC(0,3.2)

	RML=UML $\vartheta_i @(\bar{\boldsymbol{\theta}}_T, \bar{\eta})$				UML=PML $\vartheta_i @(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T)$			UML=PML $(\varphi_{im}, \varphi_{ic}) @(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T)$			RML=UML & UML=PML $\vartheta_i @(\bar{\boldsymbol{\theta}}_T, \bar{\eta}_T)$		
%	DWH1	DWH2	DWH3	LR	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3
1	99.70	100.0	100.0	100.0	27.82	10.58	92.46	41.09	41.83	92.98	99.98	100.0	100.0
5	99.77	100.0	100.0	100.0	41.82	20.71	94.59	55.53	54.57	95.13	99.98	100.0	100.0
10	99.80	100.0	100.0	100.0	50.20	28.25	95.50	63.33	61.89	96.18	99.98	100.0	100.0

GC(-0.9,3.2)

	RML=UML $\vartheta_i @(\bar{\boldsymbol{\theta}}_T, \bar{\eta})$				UML=PML $\vartheta_i @(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T)$			UML=PML $(\varphi_{im}, \varphi_{ic}) @(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T)$			RML=UML & UML=PML $\vartheta_i @(\bar{\boldsymbol{\theta}}_T, \bar{\eta}_T)$		
%	DWH1	DWH2	DWH3	LR	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3
1	99.81	100.0	100.0	100.0	47.69	50.44	98.83	100.0	100.0	100.0	99.98	100.0	100.0
5	99.84	100.0	100.0	100.0	61.40	64.23	99.17	100.0	100.0	100.0	100.0	100.0	100.0
10	99.87	100.0	100.0	100.0	68.67	71.13	99.28	100.0	100.0	100.0	100.0	100.0	100.0

Monte Carlo rejection percentages. DWH1: Wald-type Hausman test. DWH2: Hausman test based on UML (RML) score computed at PMLE (UMLE). DWH3: Hausman test based on PML (UML) score computed at MLE (RMLE). Expected Hessian and covariance matrices evaluated at RMLE $(\bar{\boldsymbol{\theta}}_T, \bar{\eta})$ or PMLE and sequential MM estimator $(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T)$. GC (Gram-Charlier expansion). Sample length=2,000. Replications=20,000.

TABLE 2: Multivariate market model: Empirical rejection rates.

Student t_{12}

	RML=UML $\vartheta_i@(\bar{\boldsymbol{\theta}}_T,\bar{\eta})$				UML=PML $\vartheta_i@(\hat{\boldsymbol{\theta}}_T,\hat{\eta}_T))$			UML=PML $(\mathbf{a},vech(\boldsymbol{\Omega}))@(\hat{\boldsymbol{\theta}}_T,\hat{\eta}_T)$			RML=UML & UML=PML $\vartheta_i@(\bar{\boldsymbol{\theta}}_T,\bar{\eta})$		
%	DWH1	DWH2	DWH3	LR	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3
1	1.31	1.07	0.98	1.06	5.04	0.09	2.31	5.38	0.46	3.17	3.32	1.12	2.64
5	5.10	5.51	4.89	5.64	10.92	1.29	5.71	12.77	3.11	10.05	6.43	3.71	5.90
10	10.09	10.68	9.77	10.68	15.76	4.23	9.29	19.57	7.18	16.68	9.64	7.15	8.96

Student t_8

	RML=UML $\vartheta_i@(\bar{\boldsymbol{\theta}}_T,\bar{\eta})$				UML=PML $\vartheta_i@(\hat{\boldsymbol{\theta}}_T,\hat{\eta}_T))$			UML=PML $(\mathbf{a},vech(\boldsymbol{\Omega}))@(\hat{\boldsymbol{\theta}}_T,\hat{\eta}_T)$			RML=UML & UML=PML $\vartheta_i@(\bar{\boldsymbol{\theta}}_T,\bar{\eta})$		
%	DWH1	DWH2	DWH3	LR	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3
1	41.07	34.46	35.29	27.92	6.21	0.09	3.05	5.99	0.31	3.98	46.78	32.57	40.52
5	57.39	53.69	53.66	49.13	12.76	1.62	7.19	14.11	2.71	11.66	60.04	50.02	55.13
10	66.37	63.48	63.10	60.29	17.61	4.50	11.16	20.91	6.35	18.40	67.06	59.15	62.89

DSMN(0.2,0.1)

	RML=UML $\vartheta_i@(\bar{\boldsymbol{\theta}}_T,\bar{\eta})$				UML=PML $\vartheta_i@(\tilde{\boldsymbol{\theta}}_T,\breve{\eta}_T)$			UML=PML $(\mathbf{a},vech(\boldsymbol{\Omega}))@(\tilde{\boldsymbol{\theta}}_T,\breve{\eta}_T)$			RML=UML & UML=PML $\vartheta_i@(\bar{\boldsymbol{\theta}}_T,\bar{\eta}_T)$		
%	DWH1	DWH2	DWH3	LR	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3
1	100.0	100.0	100.0	100.0	92.53	40.92	80.00	88.16	11.51	46.74	100.0	100.0	100.0
5	100.0	100.0	100.0	100.0	96.38	75.62	90.39	93.44	30.06	65.55	100.0	100.0	100.0
10	100.0	100.0	100.0	100.0	97.58	88.47	93.85	95.68	43.99	74.86	100.0	100.0	100.0

DSMN(0.2,0.1,0.5)

	RML=UML $\vartheta_i@(\bar{\boldsymbol{\theta}}_T,\bar{\eta})$				UML=PML $\vartheta_i@(\tilde{\boldsymbol{\theta}}_T,\breve{\eta}_T)$			UML=PML $(\mathbf{a},vech(\boldsymbol{\Omega}))@(\tilde{\boldsymbol{\theta}}_T,\breve{\eta}_T)$			RML=UML & UML=PML $\vartheta_i@(\bar{\boldsymbol{\theta}}_T,\bar{\eta}_T)$		
%	DWH1	DWH2	DWH3	LR	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3
1	100.0	100.0	100.0	100.0	96.25	43.98	86.72	99.79	97.45	98.11	100.0	100.0	100.0
5	100.0	100.0	100.0	100.0	98.30	78.15	93.84	99.94	99.27	99.42	100.0	100.0	100.0
10	100.0	100.0	100.0	100.0	98.95	89.58	96.20	99.99	99.67	99.71	100.0	100.0	100.0

Monte Carlo rejection percentages. DWH1: Wald-type Hausman test. DWH2: Hausman test based on UML (RML) score computed at PMLE (UMLE). DWH3: Hausman test based on PML (UML) score computed at MLE (RMLE). Expected Hessian and covariance matrices evaluated at RMLE $(\bar{\boldsymbol{\theta}}_T,\bar{\eta})$ or PMLE and sequential MM estimator $(\tilde{\boldsymbol{\theta}}_T,\breve{\eta}_T)$. DSMN (discrete scale mixture of two normals), DLSMN (discrete location-scale mixture of two normals). Sample length=500. Replications=20,000.

TABLE 3: Structural VAR(1): Empirical rejection rates.

Independent Stutent $t_{(\eta_1=0.15,\eta_2=0.10)}$ — Replications 20,000

	RML=UML $diag(\mathbf{C})@(\bar{\boldsymbol{\theta}}_T,\bar{\eta})$				UML=PML $vech(\boldsymbol{\Sigma})@(\hat{\boldsymbol{\theta}}_T,\hat{\eta}_T))$			UML=PML $\boldsymbol{\tau}@(\hat{\boldsymbol{\theta}}_T,\hat{\eta}_T)$			UML=PML $(\boldsymbol{\tau},vech(\boldsymbol{\Sigma}))@(\hat{\boldsymbol{\theta}}_T,\hat{\eta}_T))$			RML=UML& UML=PML $(vech(\boldsymbol{\Sigma}),diag(\mathbf{C}))@(\bar{\boldsymbol{\theta}}_T,\bar{\eta})$		
%	DWH1	DWH2	DWH3	LR	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3
1	2.70	2.22	1.11	1.06	3.39	2.06	3.04	1.65	1.16	1.51	3.64	2.09	3.31	4.61	2.76	3.28
5	7.26	6.86	4.98	5.11	6.65	4.17	5.86	5.77	4.87	5.42	7.39	4.85	6.63	8.74	6.18	6.54
10	12.40	11.88	9.76	9.97	9.75	6.51	8.81	10.61	9.78	10.26	11.18	7.87	10.13	12.29	9.47	9.70

Independent DLSMN$_{(0.52,0.06,0)(0.30,0.20,0.5)}$ — Replications 5,000

	RML=UML $diag(\mathbf{C})@(\bar{\boldsymbol{\theta}}_T,\bar{\eta})$				UML=PML $vech(\boldsymbol{\Sigma})@(\hat{\boldsymbol{\theta}}_T,\hat{\eta}_T))$			UML=PML $\boldsymbol{\tau}@(\hat{\boldsymbol{\theta}}_T,\hat{\eta}_T)$			UML=PML $(\boldsymbol{\tau},vech(\boldsymbol{\Sigma}))@(\hat{\boldsymbol{\theta}}_T,\hat{\eta}_T))$			RML=UML& UML=PML $(vech(\boldsymbol{\Sigma}),diag(\mathbf{C}))@(\bar{\boldsymbol{\theta}}_T,\bar{\eta})$		
%	DWH1	DWH2	DWH3	LR	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3
1	100	100	100	100	100	100	100	99.84	92.88	99.80	100	100	100	100	100	100
5	100	100	100	100	100	100	100	99.90	97.02	99.86	100	100	100	100	100	100
10	100	100	100	100	100	100	100	99.96	98.14	99.88	100	100	100	100	100	100

Monte Carlo rejection percentages. DWH1: Wald-type Hausman test. DWH2: Hausman test based on UML (RML) score computed at PMLE (UMLE). DWH3: Hausman test based on PML (UML) score computed at MLE (RMLE). Expected Hessian and covariance matrices evaluated at RMLE $(\bar{\boldsymbol{\theta}}_T,\bar{\eta})$ or PMLE and sequential MM estimator $(\tilde{\boldsymbol{\theta}}_T,\breve{\eta}_T)$. DLSMN (discrete location-scale mixture of two normals). Sample length=2000.

TABLE 4: Parameter estimates. Sample period 1960:08 - 2015:04

	PML			UML			RML		
τ'	0.013	1.261	0.013	0.008	1.045	-0.007	0.010	1.042	-0.002
				1.000	-0.006	0.069	1.000	-0.008	0.063
$\mathbf{J}$				14.045	1.000	0.771	21.354	1.000	0.968
				0.157	-0.001	1.000	0.208	-0.001	1.000
$\mathbf{\Psi}$				0.010	0.681	0.199	0.009	0.582	0.020
	0.001	-0.011	0.001	0.003	0.007	0.027	0.001	-0.009	0.000
$\mathbf{J\Psi^2J'} \times 10$	-0.011	4.329	0.007	0.007	5.063	0.305	-0.009	3.733	0.003
	0.001	0.007	0.007	0.027	0.305	0.397	0.000	0.003	0.004

TABLE 5: DHW test statistics. Sample period 1960:08 - 2015:04

Test	d.f.	Statistic	p-value
PML vs. UML			
$\boldsymbol{\tau}@(\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T)$	3	13.90	0.003
$vech(\mathbf{\Sigma})@(\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T)$	6	28.66	7×10^{-5}
$(\boldsymbol{\tau}, vech(\mathbf{\Sigma}))@(\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T)$	9	42.57	0.0
UML vs. RML			
$diag(\mathbf{C})@(\bar{\boldsymbol{\theta}}_T, \bar{\eta})$	3	343.93	0.0
$\boldsymbol{\eta} = \bar{\boldsymbol{\eta}}$	3	143.55	0.0

PML vs. UML tests are based on the UML score computed at the PMLE. In turn, UML vs. RML tests correspond to the UML score computed at the RMLE, and the LR test, respectively.

TABLE 6: DHW test statistics. Sample period 1988:05 - 2015:04

Test	d.f.	Statistic	p-value
PML vs. UML			
$\boldsymbol{\tau}@(\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T)$	3	5.650	0.130
$vech(\mathbf{\Sigma})@(\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T)$	6	14.57	0.024
$(\boldsymbol{\tau}, vech(\mathbf{\Sigma}))@(\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T)$	9	20.22	0.017
UML vs. RML			
$diag(\mathbf{C})@(\bar{\boldsymbol{\theta}}_T, \bar{\eta})$	3	69.69	0.0
$\boldsymbol{\eta} = \bar{\boldsymbol{\eta}}$	3	37.82	0.0

PML vs. UML tests are based on the UML score computed at the PMLE. In turn, UML vs. RML tests correspond to the UML score computed at the RMLE, and the LR test, respectively.

FIGURE 1: IRF and FVED. DGP: Independent $DLSMN_{\lambda,\kappa,\delta}$ (0,0.52,0.06)(0.3,0.2,0.2)

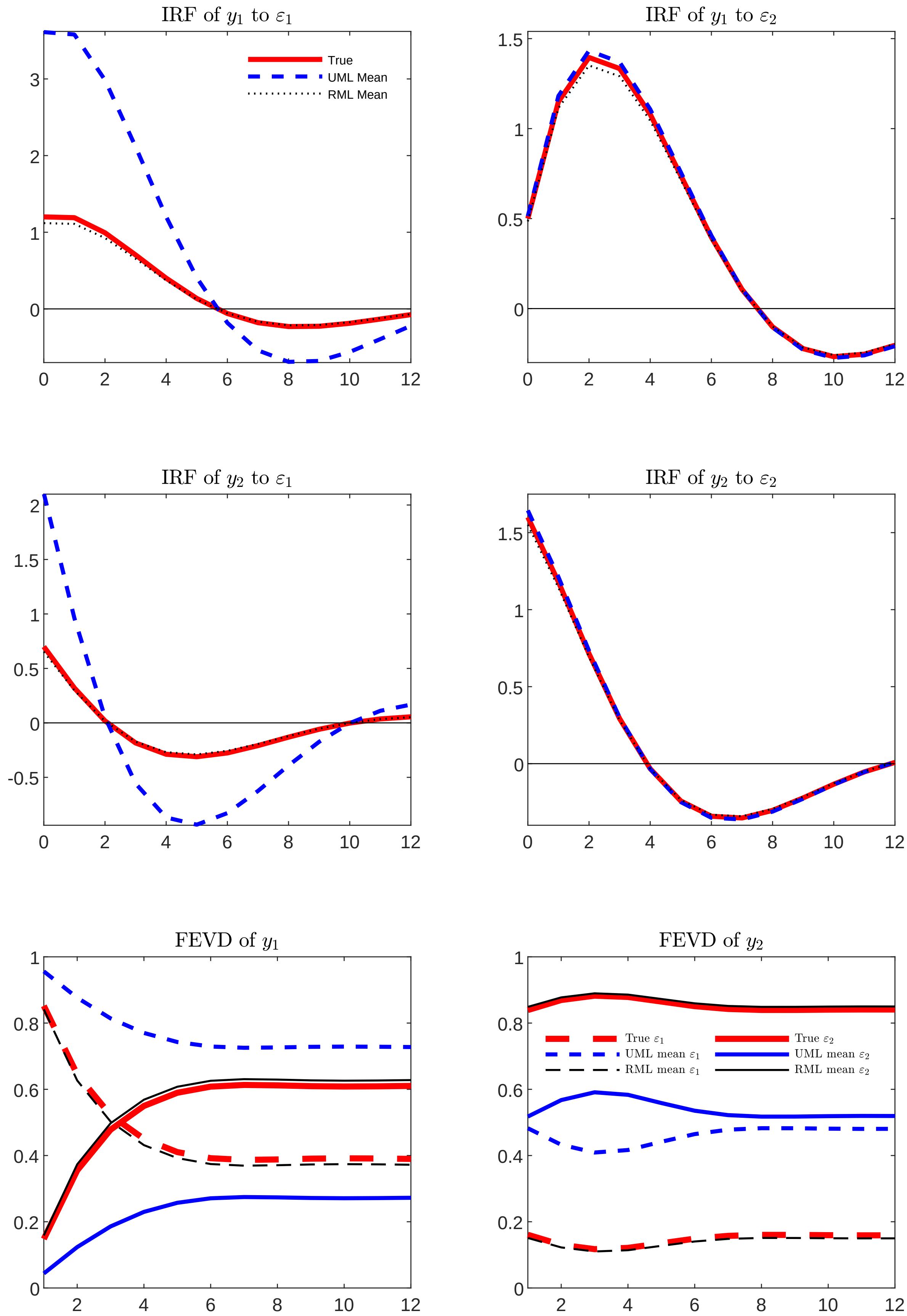
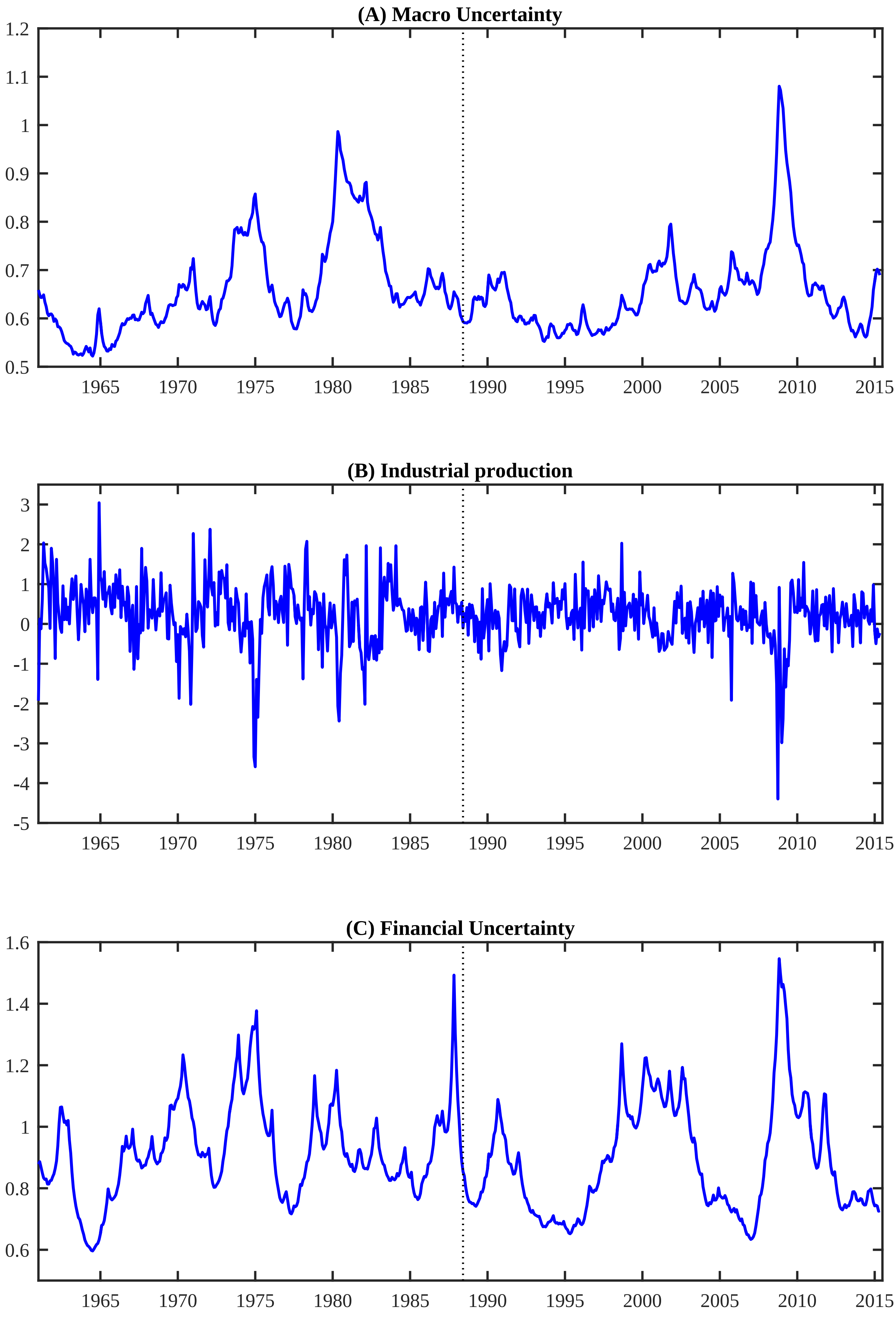


FIGURE 2: Data



Supplemental Appendices for

Specification tests for non-Gaussian maximum likelihood estimators

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>

August 2019

Revised: November 2020

B Auxiliary results

Lemma 1 *Let* $\hat{\boldsymbol{\theta}}_T = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \bar{\mathbf{m}}'_T(\boldsymbol{\theta})\tilde{\mathcal{S}}_{mT}\bar{\mathbf{m}}_T(\boldsymbol{\theta})$ *denote the GMM estimator of* $\boldsymbol{\theta}$ *over the parameter space* $\boldsymbol{\Theta}$ *based on the average influence functions* $\bar{\mathbf{m}}_T(\boldsymbol{\theta})$ *and weighting matrix* $\tilde{\mathcal{S}}_{mT}$*, and consider a homeomorphic and continuously differentiable transformation* $\boldsymbol{\pi}(.)$ *from the original parameters* $\boldsymbol{\theta}$ *to a new set of parameters* $\boldsymbol{\pi}$*, with* $\text{rank}\left[\partial\boldsymbol{\pi}'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\right]$ *evaluated at* $\hat{\boldsymbol{\theta}}_T$ *equal to* $p = \dim(\boldsymbol{\theta})$*. If* $\hat{\boldsymbol{\theta}}_T \in int(\boldsymbol{\Theta})$*, then*

$$\begin{aligned}\hat{\boldsymbol{\theta}}_T &= \boldsymbol{\theta}(\hat{\boldsymbol{\pi}}_T),\\ \hat{\boldsymbol{\pi}}_T &= \boldsymbol{\pi}(\hat{\boldsymbol{\theta}}_T),\end{aligned}$$

and

$$\bar{\mathbb{m}}'_T(\hat{\boldsymbol{\pi}}_T)\tilde{\mathcal{S}}_{mT}\bar{\mathbb{m}}_T(\hat{\boldsymbol{\pi}}_T) = \bar{\mathbf{m}}'_T(\hat{\boldsymbol{\theta}}_T)\tilde{\mathcal{S}}_{mT}\bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T),$$

where $\boldsymbol{\theta}(\boldsymbol{\pi})$ *is the inverse mapping such that* $\boldsymbol{\pi}[\boldsymbol{\theta}(\boldsymbol{\pi})] = \boldsymbol{\pi}$*,* $\bar{\mathbb{m}}_T(\boldsymbol{\pi}) = \bar{\mathbf{m}}_T[\boldsymbol{\theta}(\boldsymbol{\pi})]$ *are the average influence functions written in terms of* $\boldsymbol{\pi}$*, and* $\hat{\boldsymbol{\pi}}_T = \arg\min_{\boldsymbol{\pi}\in\boldsymbol{\Pi}} \bar{\mathbb{m}}'_T(\boldsymbol{\pi})\tilde{\mathcal{S}}_{mT}\bar{\mathbb{m}}_T(\boldsymbol{\pi})$*.*

Proof. The interior solution assumption implies that the sample first-order condition characterising $\hat{\boldsymbol{\theta}}_T$ is

$$\frac{\partial\bar{\mathbf{m}}'_T(\hat{\boldsymbol{\theta}}_T)}{\partial\boldsymbol{\theta}}\tilde{\mathcal{S}}_{\mathrm{m}T}\bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T) = \mathbf{0}, \tag{B1}$$

while the corresponding condition for $\hat{\boldsymbol{\pi}}_T$ will be

$$\frac{\partial\bar{\mathbb{m}}'_T(\hat{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\tilde{\mathcal{S}}_{\mathrm{m}T}\bar{\mathbb{m}}_T(\hat{\boldsymbol{\pi}}_T) = \frac{\partial\boldsymbol{\theta}'(\hat{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\frac{\partial\bar{\mathbf{m}}'_T[\boldsymbol{\theta}(\hat{\boldsymbol{\pi}}_T)]}{\partial\boldsymbol{\theta}}\tilde{\mathcal{S}}_{\mathrm{m}T}\bar{\mathbf{m}}_T[\boldsymbol{\theta}(\hat{\boldsymbol{\pi}}_T)] = \mathbf{0} \tag{B2}$$

by the chain rule for derivatives. Given that $\text{rank}\left[\partial\boldsymbol{\theta}'(\boldsymbol{\pi})/\partial\boldsymbol{\pi}\right]$ evaluated at $\boldsymbol{\pi}(\hat{\boldsymbol{\theta}}_T)$ is p in view of our assumption on the rank of the direct Jacobian $\partial\boldsymbol{\pi}'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ by virtue of the inverse mapping theorem, the above equations imply that $\hat{\boldsymbol{\theta}}_T = \boldsymbol{\theta}(\hat{\boldsymbol{\pi}}_T)$, whence the other two results trivially follow. □

This result confirms the numerical invariance of the GMM criterion to reparametrisations when the weighting matrix remains the same, a condition satisfied by the most popular choices, including the identity matrix, as well as the unconditional sample variance of the influence functions and its long-run counterpart when the initial estimators at which those matrices are evaluated satisfy $\boldsymbol{\pi}^i = \boldsymbol{\pi}(\boldsymbol{\theta}^i)$. Obviously, in exactly identified contexts, such as the one implicitly arising in maximum likelihood estimation, in which the usual sufficient identification condition $\text{rank}\{E[\partial\mathbf{m}_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}']\} = p$ holds, the weighting matrix becomes irrelevant, at least in large samples, which allows us to replace the first order conditions (B1) and (B2) by $\bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T) = \mathbf{0}$, and $\bar{\mathbb{m}}_T(\hat{\boldsymbol{\pi}}_T) = \mathbf{0}$, respectively. Aside from this change, the results of the lemma continue to hold.

Lemma 2 *Let* ς *denote a scalar random variable with continuously differentiable density function* $h(\varsigma;\boldsymbol{\eta})$ *over the possibly infinite domain* $[a,b]$*, and let* $m(\varsigma)$ *denote a continuously differentiable function over the same domain such that* $E[m(\varsigma)|\boldsymbol{\eta}] = k(\boldsymbol{\eta}) < \infty$*. Then*

$$E\left[\partial m(\varsigma)/\partial\varsigma|\,\boldsymbol{\eta}\right] = -E\left[m(\varsigma)\partial\ln h(\varsigma;\boldsymbol{\eta})/\partial\varsigma|\,\boldsymbol{\eta}\right],$$

as long as the required expectations are defined and bounded.

Proof. If we differentiate

$$k(\boldsymbol{\eta}) = E\left[m(\varsigma)|\,\boldsymbol{\eta}\right] = \int_a^b m(\varsigma)h(\varsigma;\boldsymbol{\eta})d\varsigma$$

with respect to ς, we get

$$0=\int_a^b \frac{\partial m(\varsigma)}{\partial\varsigma}h(\varsigma;\boldsymbol{\eta})d\varsigma+\int_a^b m(\varsigma)\frac{\partial h(\varsigma;\boldsymbol{\eta})}{\partial\varsigma}d\varsigma=\int_a^b \frac{\partial m(\varsigma)}{\partial\varsigma}h(\varsigma;\boldsymbol{\eta})d\varsigma+\int_a^b m(\varsigma)h(\varsigma;\boldsymbol{\eta})\frac{\partial \ln h(\varsigma;\boldsymbol{\eta})}{\partial\varsigma}d\varsigma,$$

as required. □

Lemma 3 *If* $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0$ *is i.i.d.* $D(\mathbf{0},\mathbf{I}_N,\boldsymbol{\varrho})$ *with density function* $f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})$, *where* $\boldsymbol{\varrho}=\mathbf{0}$ *denotes normality, then*

$$E\left\{\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\left[\mathbf{e}'_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}),\mathbf{e}'_{rt}(\boldsymbol{\theta},\boldsymbol{\varrho})\right]\Big|\, I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}\right\} = [\mathcal{K}\left(\mathbf{0}\right)|\mathbf{0}]. \tag{B3}$$

Proof. We can use the conditional analogue to the generalised information matrix equality (see e.g. Newey and McFadden (1994)) to show that

$$\begin{aligned} E\left\{\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})\left[\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\varrho}),\mathbf{s}'_{\boldsymbol{\varrho}t}(\boldsymbol{\theta},\boldsymbol{\varrho})\right]\Big|\, I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}\right\} &= -E\left\{\left[\frac{\partial \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})}{\partial\boldsymbol{\theta}'}\middle|\frac{\partial \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})}{\partial\boldsymbol{\varrho}'}\right]\middle| I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}\right\} \\ &= -E\left\{\left[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta};\mathbf{0})|\mathbf{0}\right]\right| I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}\} = [\mathcal{A}_t(\phi)|\mathbf{0}] \end{aligned}$$

irrespective of the conditional distribution of $\boldsymbol{\varepsilon}_t^*$, where we have used the fact that $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})$ does not vary with $\boldsymbol{\varrho}$ when regarded as the influence function for $\tilde{\boldsymbol{\theta}}_T$. Then, the required result follows from the martingale difference nature of both $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$ and $\mathbf{e}_t(\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0)$. □

Lemma 4

$$\begin{pmatrix} \mathcal{M}_{ss} & \mathcal{M}_{sr} \\ \mathcal{M}'_{sr} & \mathcal{M}_{rr} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{K}_{NN}+\boldsymbol{\Upsilon} & \mathbf{E}_N\mathrm{M}_{sr} \\ \mathrm{M}'_{sr}\mathbf{E}_N & \mathcal{M}_{rr} \end{pmatrix}^{-1}$$

$$\begin{pmatrix} \boldsymbol{\Delta}_N[\boldsymbol{\Delta}'_N(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N]^{-1}\boldsymbol{\Delta}'_N & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{E}_N & 0 \\ 0 & \mathbf{I}_N \end{pmatrix}\begin{pmatrix} \mathrm{M}_{ss} & \mathrm{M}_{sr} \\ \mathrm{M}'_{sr} & \mathcal{M}_{rr} \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{E}'_N & 0 \\ 0 & \mathbf{I}_N \end{pmatrix}, \tag{B4}$$

where $\mathcal{M}_{ss}$, $\mathcal{M}_{sr}$, $\mathcal{M}_{rr}$, $\boldsymbol{\Upsilon}$ *and* M_{sr} *are defined in Proposition D2, and* $\mathrm{M}_{ss}=(\mathbf{I}_N+\mathbf{E}'_N\boldsymbol{\Upsilon}\mathbf{E}_N)$ *is a diagonal matrix of order* N *with typical element* $\mathrm{M}_{ss}(\boldsymbol{\varrho}_i)$.

Proof. Using the partitioned inverse formula, we get

$$\begin{pmatrix} \mathcal{M}_{ss} & \mathcal{M}_{sr} \\ \mathcal{M}'_{sr} & \mathcal{M}_{rr} \end{pmatrix}^{-1} = \begin{bmatrix} \mathcal{M}_{ss}^{-1}+\mathcal{M}_{ss}^{-1}\mathcal{M}_{sr}\mathcal{M}^{rr}\mathcal{M}'_{sr}\mathcal{M}_{ss}^{-1} & -\mathcal{M}_{ss}^{-1}\mathcal{M}_{sr}\mathcal{M}^{rr} \\ -\mathcal{M}^{rr}\mathcal{M}'_{sr}\mathcal{M}_{ss}^{-1} & (\mathcal{M}_{rr}-\mathcal{M}'_{sr}\mathcal{M}_{ss}^{-1}\mathcal{M}_{sr})^{-1} \end{bmatrix}.$$

Given that $\boldsymbol{\Upsilon}$ is diagonal, we can use Proposition 7 in Magnus and Sentana (2020), which yields

$$\begin{aligned} \mathcal{M}_{ss}^{-1} &= (\mathbf{K}_{NN}+\boldsymbol{\Upsilon})^{-1} = \boldsymbol{\Delta}_N[\boldsymbol{\Delta}'_N(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N]^{-1}\boldsymbol{\Delta}'_N + \mathbf{E}_N(\mathbf{I}_N+\mathbf{E}'_N\boldsymbol{\Upsilon}\mathbf{E}_N)^{-1}\mathbf{E}'_N \\ &= \boldsymbol{\Delta}_N[\boldsymbol{\Delta}'_N(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N]^{-1}\boldsymbol{\Delta}'_N + \mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathbf{E}'_N. \end{aligned}$$

In turn, Theorem 7.4(i) in Magnus (1988) states that $\mathbf{K}_{NN}\mathbf{E}_N = \mathbf{E}_N$, which implies that $\mathcal{M}_{ss}\mathbf{E}_N = (\mathbf{K}_{NN} + \boldsymbol{\Upsilon})\mathbf{E}_N = (\mathbf{I}_{N^2} + \boldsymbol{\Upsilon})\mathbf{E}_N = \mathbf{E}_N(\mathbf{I}_N + \mathbf{E}_N'\boldsymbol{\Upsilon}\mathbf{E}_N) = \mathbf{E}_N\mathrm{M}_{ss}$ by virtue of Proposition 3 in Magnus and Sentana (2020). Then, if we premultiply both sides by $\mathcal{M}_{ss}^{-1} = (\mathbf{K}_{NN} + \boldsymbol{\Upsilon})^{-1}$, we end up with $\mathbf{E}_N = \mathcal{M}_{ss}^{-1}\mathbf{E}_N\mathrm{M}_{ss}$, whence we finally obtain that $\mathcal{M}_{ss}^{-1}\mathbf{E}_N = \mathbf{E}_N\mathrm{M}_{ss}^{-1}$. Thus, $\mathcal{M}_{ss}^{-1}\mathcal{M}_{sr} = \mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}$, where $\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}$ is a diagonal matrix with typical element $\mathrm{M}_{sr}(\boldsymbol{\varrho}_i)/\mathrm{M}_{ss}(\boldsymbol{\varrho}_i)$. Therefore $\mathcal{M}_{sr}'\mathcal{M}_{ss}^{-1}\mathcal{M}_{sr} = \mathrm{M}_{sr}'\mathbf{E}_N'\mathcal{M}_{ss}^{-1}\mathbf{E}_N\mathrm{M}_{sr} = \mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}$ will be a diagonal $N \times N$ matrix with typical diagonal element $\mathrm{M}_{sr}^2(\boldsymbol{\varrho}_i)/\mathrm{M}_{ss}(\boldsymbol{\varrho}_i)$. In turn, this implies that $\mathcal{M}_{rr} - \mathcal{M}_{sr}'\mathcal{M}_{ss}^{-1}\mathcal{M}_{sr} = \mathcal{M}_{rr} - \mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}$ is a diagonal matrix with typical element $\mathrm{M}_{rr}(\nu_i) - \mathrm{M}_{sr}^2(\boldsymbol{\varrho}_i)/\mathrm{M}_{ss}(\boldsymbol{\varrho}_i)$, so that $\mathcal{M}^{rr} = (\mathcal{M}_{rr} - \mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr})^{-1}$ is also diagonal. Moreover, $\mathcal{M}_{ss}^{-1}\mathcal{M}_{sr}\mathcal{M}^{rr} = \mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}$, where $\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}$ is once again diagonal with typical element $[\mathrm{M}_{sr}(\boldsymbol{\varrho}_i)/\mathrm{M}_{ss}(\boldsymbol{\varrho}_i)]/[\mathrm{M}_{rr}(\nu_i) - \mathrm{M}_{sr}^2(\boldsymbol{\varrho}_i)/\mathrm{M}_{ss}(\boldsymbol{\varrho}_i)]$.

If we put all these pieces together, we end up with

$$\begin{pmatrix} \mathcal{M}_{ss} & \mathcal{M}_{sr} \\ \mathcal{M}_{sr}' & \mathcal{M}_{rr} \end{pmatrix}^{-1} = \begin{pmatrix} \mathcal{M}_{ss}^{-1} + \mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathbf{E}_N' & -\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr} \\ -\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathbf{E}_N' & \mathcal{M}^{rr} \end{pmatrix}$$

$$= \left\{ \begin{array}{cc} \boldsymbol{\Delta}_N[\boldsymbol{\Delta}_N'(\mathbf{K}_{NN} + \boldsymbol{\Upsilon})\boldsymbol{\Delta}_N]^{-1}\boldsymbol{\Delta}_N' + \mathbf{E}_N(\mathrm{M}_{ss}^{-1} + \mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1})\mathbf{E}_N' & -\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr} \\ -\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathbf{E}_N' & \mathcal{M}^{rr} \end{array} \right\}$$

$$= \begin{pmatrix} \boldsymbol{\Delta}_N[\boldsymbol{\Delta}_N'(\mathbf{K}_{NN} + \boldsymbol{\Upsilon})\boldsymbol{\Delta}_N]^{-1}\boldsymbol{\Delta}_N' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

$$+ \begin{pmatrix} \mathbf{E}_N & 0 \\ 0 & \mathbf{I}_N \end{pmatrix} \begin{pmatrix} \mathrm{M}_{ss}^{-1} + \mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1} & -\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr} \\ -\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1} & \mathcal{M}^{rr} \end{pmatrix} \begin{pmatrix} \mathbf{E}_N' & 0 \\ 0 & \mathbf{I}_N \end{pmatrix}$$

$$= \begin{pmatrix} \boldsymbol{\Delta}_N[\boldsymbol{\Delta}_N'(\mathbf{K}_{NN} + \boldsymbol{\Upsilon})\boldsymbol{\Delta}_N]^{-1}\boldsymbol{\Delta}_N' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{E}_N & 0 \\ 0 & \mathbf{I}_N \end{pmatrix} \begin{pmatrix} \mathrm{M}_{ss} & \mathrm{M}_{sr} \\ \mathrm{M}_{sr}' & \mathcal{M}_{rr} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{E}_N' & 0 \\ 0 & \mathbf{I}_N \end{pmatrix},$$

as claimed. □

Proposition B1 *If model (18) with cross-sectionally independent symmetric structural shocks generates a covariance stationary process, then:*

1. *Its information matrix is block diagonal between* $(\boldsymbol{\tau}', \mathbf{a}')'$ *and* $(\mathbf{c}', \boldsymbol{\varrho}')'$
2. *The asymptotic covariance matrix of the restricted and unrestricted ML estimators of* $(\boldsymbol{\tau}', \mathbf{a}')'$ *will be given by*

$$\begin{bmatrix} 1 & \boldsymbol{\mu}' & \cdots & \boldsymbol{\mu}' \\ \boldsymbol{\mu} & (\boldsymbol{\Gamma}(0) + \boldsymbol{\mu}\boldsymbol{\mu}') & \cdots & (\boldsymbol{\Gamma}(p-1) + \boldsymbol{\mu}\boldsymbol{\mu}') \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mu} & (\boldsymbol{\Gamma}'(p-1) + \boldsymbol{\mu}\boldsymbol{\mu}') & \cdots & (\boldsymbol{\Gamma}(0) + \boldsymbol{\mu}\boldsymbol{\mu}') \end{bmatrix}^{-1} \otimes \mathbf{C}\mathcal{M}_{ll}^{-1}\mathbf{C}',$$

   where $\boldsymbol{\Gamma}(p)$ *is the* p^{th} *autocovariance matrix of* $\mathbf{y}_t$ *and* $\mathcal{M}_{ll}$ *is defined in Proposition D2.*
3. *The asymptotic covariance matrices of the restricted and unrestricted ML estimators of* $\mathbf{c}$ *and* $\boldsymbol{\varrho}$ *are given by*

$$(\mathbf{I}_N \otimes \mathbf{C})\mathcal{M}_{ss}^{-1}(\mathbf{I}_N \otimes \mathbf{C}') \text{ and}$$

$$\begin{bmatrix} (\mathbf{I}_N \otimes \mathbf{C}'^{-1})\mathcal{M}_{ss}(\mathbf{I}_N \otimes \mathbf{C}^{-1}) & (\mathbf{I}_N \otimes \mathbf{C}'^{-1})\mathcal{M}_{sr} \\ \mathcal{M}_{sr}'(\mathbf{I}_N \otimes \mathbf{C}^{-1}) & \mathcal{M}_{rr} \end{bmatrix}^{-1},$$

   respectively, where $\mathcal{M}_{ss}$, $\mathcal{M}_{sr}$ *and* $\mathcal{M}_{rr}$ *are also defined in Proposition D2 and the rank of the difference between the asymptotic variances of these two estimators of* $\mathbf{c}$ *is* N.

Proof. Given the mapping between the structural and reduced form parameters, the contribution to the conditional log-likelihood function from observation t $(t = 1, \ldots, T)$ will be

$$l_t(\mathbf{y}_t; \boldsymbol{\phi}) = -\ln|\mathbf{C}| + l[\varepsilon^*_{1t}(\boldsymbol{\theta}); \boldsymbol{\varrho}_1] + \ldots + l[\varepsilon^*_{Nt}(\boldsymbol{\theta}); \boldsymbol{\varrho}_N],$$

where $l[\varepsilon^*_{it}(\boldsymbol{\theta}); \boldsymbol{\varrho}_i]$ is the univariate log-likelihood function for the i^{th} structural shock $\varepsilon^*_{it}(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}) = \mathbf{C}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$, and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = (\mathbf{y}_t - \boldsymbol{\tau} - \boldsymbol{\Phi}_1\mathbf{y}_{t-1} - \ldots - \boldsymbol{\Phi}_p\mathbf{y}_{t-p})$. To compute the gradient and information matrix, we rely on the expressions in Supplemental Appendix D.3 because the assumed multivariate distribution for $\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta})$ is not elliptically symmetric despite the marginal distributions of its components being symmetric. Given that the conditional mean vector and covariance matrix of (18) are given by

$$\begin{aligned}
\boldsymbol{\mu}_t(\boldsymbol{\theta}) &= \boldsymbol{\tau} + \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p}, \\
\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) &= \mathbf{C}\mathbf{C}',
\end{aligned}$$

respectively, straightforward algebra shows that

$$\begin{aligned}
\mathbf{Z}_{lt}(\boldsymbol{\theta}) &= \frac{\partial \boldsymbol{\mu}'_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}_N \\ \mathbf{y}_{t-1} \otimes \mathbf{I}_N \\ \vdots \\ \mathbf{y}_{t-p} \otimes \mathbf{I}_N \\ \mathbf{0}_{N^2 \times N} \end{pmatrix} \mathbf{C}^{-1\prime}, \\
\mathbf{Z}_{st}(\boldsymbol{\theta}) &= \frac{\partial vec'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} [\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})] = \begin{pmatrix} \mathbf{0}_{N \times N^2} \\ \mathbf{0}_{N^2 \times N^2} \\ \vdots \\ \mathbf{0}_{N^2 \times N^2} \\ \mathbf{I}_{N^2} \end{pmatrix} (\mathbf{I}_N \otimes \mathbf{C}^{-1\prime}),
\end{aligned}$$

which means that the conditional mean and variance parameters are variation free. This fact, combined with the symmetry of the Student t and the formulas in Proposition D2, immediately implies that the information matrix will be block diagonal. Specifically, the block of the information matrix corresponding to the $N + pN^2$ conditional mean parameters $(\boldsymbol{\tau}, \mathbf{a})$ will be

$$E[\mathbf{Z}_{lt}(\boldsymbol{\theta})\mathcal{M}_{ll}\mathbf{Z}'_{lt}(\boldsymbol{\theta})] = E \begin{bmatrix} 1 & \mathbf{y}'_{t-1} & \cdots & \mathbf{y}'_{t-p} \\ \mathbf{y}_{t-1} & \mathbf{y}_{t-1}\mathbf{y}'_{t-1} & \cdots & \mathbf{y}_{t-1}\mathbf{y}'_{t-p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_{t-p} & \mathbf{y}_{t-p}\mathbf{y}'_{t-1} & \cdots & \mathbf{y}_{t-p}\mathbf{y}'_{t-p} \end{bmatrix} \otimes \mathbf{C}^{-1\prime}\mathcal{M}_{ll}\mathbf{C}^{-1} \quad \text{(B5)}$$

$$= \begin{bmatrix} 1 & \boldsymbol{\mu}' & \cdots & \boldsymbol{\mu}' \\ \boldsymbol{\mu} & (\boldsymbol{\Gamma}(0) + \boldsymbol{\mu}\boldsymbol{\mu}') & \cdots & (\boldsymbol{\Gamma}(p-1) + \boldsymbol{\mu}\boldsymbol{\mu}') \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mu} & (\boldsymbol{\Gamma}'(p-1) + \boldsymbol{\mu}\boldsymbol{\mu}') & \cdots & (\boldsymbol{\Gamma}(0) + \boldsymbol{\mu}\boldsymbol{\mu}') \end{bmatrix} \otimes \mathbf{C}^{-1\prime}\mathcal{M}_{ll}\mathbf{C}^{-1}. \quad \text{(B6)}$$

In turn, the (conditional) information matrix for the unrestricted ML estimators of the N^2 structural shock coefficients $\mathbf{c}$ and the N shape parameters $\boldsymbol{\varrho}$ will be:

$$\begin{pmatrix} \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{pmatrix} \begin{pmatrix} \mathcal{M}_{ss} & \mathcal{M}_{sr} \\ \mathcal{M}'_{sr} & \mathcal{M}_{rr} \end{pmatrix} \begin{pmatrix} \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{pmatrix}.$$

In this respect, we can use the results in Proposition D2 to prove that

$$\begin{pmatrix} \mathcal{M}_{ss} & \mathcal{M}_{sr} \\ \mathcal{M}'_{sr} & \mathcal{M}_{rr} \end{pmatrix} = \begin{pmatrix} \mathbf{K}_{NN}+\mathbf{\Upsilon} & \mathbf{E}_N \mathrm{M}_{sr} \\ \mathrm{M}'_{sr}\mathbf{E}_N & \mathcal{M}_{rr} \end{pmatrix}.$$

Hence, the information matrix will be

$$\begin{aligned} &\begin{pmatrix} \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{pmatrix}\begin{pmatrix} \mathcal{M}_{ss} & \mathcal{M}_{sr} \\ \mathcal{M}'_{sr} & \mathcal{M}_{rr} \end{pmatrix}\begin{pmatrix} \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{pmatrix} \\ = &\begin{bmatrix} (\mathbf{I}_N\otimes\mathbf{C}'^{-1})(\mathbf{K}_{NN}+\mathbf{\Upsilon})(\mathbf{I}_N\otimes\mathbf{C}^{-1}) & (\mathbf{I}_N\otimes\mathbf{C}'^{-1})\mathbf{E}_N\mathrm{M}_{sr} \\ \mathrm{M}'_{sr}\mathbf{E}_N(\mathbf{I}_N\otimes\mathbf{C}^{-1}) & \mathcal{M}_{rr} \end{bmatrix}. \end{aligned}$$

If we then use the expressions in Lemma 4, we can easily show that the inverse of the information matrix will be

$$\begin{bmatrix} (\mathbf{I}_N\otimes\mathbf{C})\{\mathbf{\Delta}_N[\mathbf{\Delta}'_N(\mathbf{K}_{NN}+\mathbf{\Upsilon})\mathbf{\Delta}_N]^{-1}\mathbf{\Delta}'_N+\mathbf{E}_N\mathrm{M}^{ss}\mathbf{E}'_N\}(\mathbf{I}_N\otimes\mathbf{C}) & -(\mathbf{I}_N\otimes\mathbf{C})\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr} \\ -\mathcal{M}^{rr}\mathrm{M}'_{sr}\mathrm{M}_{ss}^{-1}\mathbf{E}'_N(\mathbf{I}_N\otimes\mathbf{C}) & \mathcal{M}^{rr} \end{bmatrix},$$

where $\mathrm{M}^{ss} = \mathrm{M}_{ss}^{-1}+\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}'_{sr}\mathrm{M}_{ss}^{-1}$.

In contrast, if we assume that the shape parameters are fixed at their true values, the asymptotic covariance matrix of the restricted ML estimators of $\mathbf{c}$ will be

$$\begin{aligned} (\mathbf{I}_N\otimes\mathbf{C})\mathcal{M}_{ss}^{-1}(\mathbf{I}_N\otimes\mathbf{C}') &= (\mathbf{I}_N\otimes\mathbf{C})\mathbf{\Delta}_N[\mathbf{\Delta}'_N(\mathbf{K}_{NN}+\mathbf{\Upsilon})\mathbf{\Delta}_N]^{-1}\mathbf{\Delta}'_N(\mathbf{I}_N\otimes\mathbf{C}') \\ &\quad +(\mathbf{I}_N\otimes\mathbf{C})\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathbf{E}'_N(\mathbf{I}_N\otimes\mathbf{C}'). \end{aligned}$$

Therefore, the efficiency loss from simultaneously estimating the N shape parameters $\boldsymbol{\varrho}$ will be

$$(\mathbf{I}_N\otimes\mathbf{C})\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}'_{sr}\mathrm{M}_{ss}^{-1}\mathbf{E}'_N(\mathbf{I}_N\otimes\mathbf{C}'),$$

which has rank N rather than N^2 because $\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}'_{sr}\mathrm{M}_{ss}^{-1}\mathbf{E}'_N$ is a diagonal matrix of rank N in which the non-zero diagonal elements are

$$\frac{1}{\psi_i^2}\frac{\mathrm{M}_{sr}^2(\boldsymbol{\varrho}_i)}{\mathrm{M}_{ss}^2(\boldsymbol{\varrho}_i)}\left[\mathrm{M}_{rr}(\boldsymbol{\varrho}_i)-\frac{\mathrm{M}_{sr}^2(\boldsymbol{\varrho}_i)}{\mathrm{M}_{ss}(\boldsymbol{\varrho}_i)}\right]^{-1}.$$

Finally, note that since the ranks of $(\mathbf{I}_N\otimes\mathbf{C}'^{-1})$ and $\mathcal{M}_{sr}=\mathbf{E}_N\mathrm{M}_{sr}$ are N^2 and N, respectively, Sylvester's rank inequality implies that

$$rank[(\mathbf{I}_N\otimes\mathbf{C})\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}] = N,$$

so that Holly's (1982) condition holds. □

Proposition B2 *If model (18) with cross-sectionally independent symmetric structural shocks generates a covariance stationary process, then the asymptotic covariance matrix of the Gaussian PML estimators is block diagonal between $(\boldsymbol{\tau}', \mathbf{a}')'$ and $\boldsymbol{\sigma}$, with the first block given by*

$$\begin{bmatrix} 1 & \boldsymbol{\mu}' & \cdots & \boldsymbol{\mu}' \\ \boldsymbol{\mu} & (\mathbf{\Gamma}(0)+\boldsymbol{\mu}\boldsymbol{\mu}') & \cdots & (\mathbf{\Gamma}(p-1)+\boldsymbol{\mu}\boldsymbol{\mu}') \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mu} & (\mathbf{\Gamma}'(p-1)+\boldsymbol{\mu}\boldsymbol{\mu}') & \cdots & (\mathbf{\Gamma}(0)+\boldsymbol{\mu}\boldsymbol{\mu}') \end{bmatrix}^{-1}\otimes\mathbf{\Sigma}$$

and the second block by

$$[\mathbf{D}_N'(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{D}_N]^{-1}\mathbf{D}_N'(\mathbf{C}^{-1}\otimes\mathbf{C}^{-1})\mathcal{K}(\mathbf{C}^{-\prime}\otimes\mathbf{C}^{-1\prime})\mathbf{D}_N[\mathbf{D}_N'(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{D}_N]^{-1},$$

where $\mathcal{K}=E[vec(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^*-\mathbf{I}_N)vec(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^*-\mathbf{I}_N)']$ *is the* $N^2\times N^2$ *matrix of fourth-order moments of the structural shocks.*

Proof. The information matrix equality implies that the expected value of the (minus) Hessian of the Gaussian pseudo log-likelihood usually coincides with the value of the true information matrix under normality. Therefore, we could exploit the fact that $\mathcal{M}_{ll}=\mathbf{I}_N$ and $\mathbf{C}^{-1\prime}\mathcal{M}_{ll}\mathbf{C}^{-1}=\mathbf{\Sigma}^{-1}$ under normality to simplify the expressions we have already derived for $\boldsymbol{\tau}$ and $\mathbf{a}$ in Proposition B1. However, the situation is slightly more complicated for $\boldsymbol{\sigma}$ because the number of parameters that can be identified by the Gaussian and non-Gaussian PMLs is different. For that reason, we use the expressions in Proposition C2 to prove that the bottom block of the (minus) expected value of the Hessian will be given by

$$\mathcal{A}_{\boldsymbol{\sigma}\boldsymbol{\sigma}}=\frac{1}{4}\mathbf{D}_N'(\mathbf{\Sigma}^{-\frac{1}{2}}\otimes\mathbf{\Sigma}^{-\frac{1}{2}})(\mathbf{I}_{N^2}+\mathbf{K}_{NN})(\mathbf{\Sigma}^{-\frac{1}{2}\prime}\otimes\mathbf{\Sigma}^{-\frac{1}{2}\prime})\mathbf{D}_N=\frac{1}{2}\mathbf{D}_N'(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{D}_N$$

regardless of the choice of square root matrix in view of the properties of the duplication and commutation matrix in Magnus and Neudecker (2919).

As for the matrix $\mathcal{B}$, which contains the asymptotic variance of the Gaussian scores, the symmetry of the marginal distributions of the structural shocks together with the cross-sectional independence across shocks imply that we will also obtain a block diagonal expression with the same block for the conditional mean parameters as $\mathcal{A}$. In contrast, the block for the conditional variance parameters $\boldsymbol{\sigma}$ will be different. To obtain it, we can use the expressions in Proposition C2 with $\mathbf{C}$ playing the role of $\mathbf{\Sigma}^{\frac{1}{2}}$ to exploit the cross-sectional independence of the structural shocks, which leads to

$$\mathcal{B}_{\boldsymbol{\sigma}\boldsymbol{\sigma}}=\frac{1}{4}\mathbf{D}_N'(\mathbf{C}^{-1}\otimes\mathbf{C}^{-1})\mathcal{K}(\mathbf{C}^{-\prime}\otimes\mathbf{C}^{-1\prime})\mathbf{D}_N,$$

where $\mathcal{K}$ is equal to $\mathbf{K}_{NN}$ plus a a block diagonal matrix in which each of the N blocks is diagonal of size $N\times N$ with the following structure:

$$\begin{bmatrix}1&0&0&0&0&0&0\\0&\ddots&0&0&0&0&0\\0&0&1&0&0&0&0\\0&0&0&\kappa_{ii}(\boldsymbol{\varrho}_i)&0&0&0\\0&0&0&0&1&0&0\\0&0&0&0&0&\ddots&0\\0&0&0&0&0&0&1\end{bmatrix}.$$

In the Student t case, $\kappa_{ii}(\boldsymbol{\varrho}_i)=(\nu_i+2)/(\nu_i-4)$. □

Proposition B3 *If model (18) with cross-sectionally independent symmetric structural shocks generates a covariance stationary process, then the scores and information matrix of* $\boldsymbol{\sigma}_L$ *and* $\boldsymbol{\omega}$ *are given by*

$$\begin{bmatrix}s_{\boldsymbol{\sigma}_L}(\boldsymbol{\theta};\boldsymbol{\varrho})\\s_{\boldsymbol{\omega}}(\boldsymbol{\theta};\boldsymbol{\varrho})\end{bmatrix}=\begin{bmatrix}\mathbf{L}_N(\mathbf{I}_N\otimes\mathbf{\Sigma}_L^{-1\prime})(\mathbf{Q}\otimes\mathbf{Q})\\\partial vec'(\mathbf{Q})/\partial\boldsymbol{\omega}\cdot(\mathbf{I}_N\otimes\mathbf{Q})\end{bmatrix}\mathbf{e}_{st}(\boldsymbol{\phi})$$

and

$$\left[\begin{array}{c}\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L^{-1\prime})(\mathbf{Q}\otimes\mathbf{Q})\\ \partial vec'(\mathbf{Q})/\partial\boldsymbol{\omega}\cdot(\mathbf{I}_N\otimes\mathbf{Q})\end{array}\right]\mathcal{M}_{ss}\left[\begin{array}{cc}(\mathbf{Q}'\otimes\mathbf{Q}')(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L^{-1})\mathbf{L}_N & (\mathbf{I}_N\otimes\mathbf{Q})\cdot\partial vec(\mathbf{Q})/\partial\boldsymbol{\omega}'\end{array}\right].$$

Proof. As in Proposition 14, the proof builds up on Proposition B1. Specifically, given that $vec(\mathbf{C})=(\mathbf{Q}'\otimes\mathbf{I}_N)vec(\boldsymbol{\Sigma}_L)=(\mathbf{Q}'\otimes\mathbf{I}_N)\mathbf{L}_N'vech(\boldsymbol{\Sigma}_L)$, straightforward algebra shows that

$$\frac{\partial\mathbf{c}}{\partial\boldsymbol{\sigma}_L'}=(\mathbf{Q}'\otimes\mathbf{I}_N)\mathbf{L}_N'.$$

Similarly, given that we can also write $vec(\mathbf{C})=(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L)vec(\mathbf{Q})$, we will have that

$$\frac{\partial\mathbf{c}}{\partial\boldsymbol{\omega}'}=(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L)\frac{\partial vec(\mathbf{Q})}{\partial\boldsymbol{\omega}'},$$

where $\partial vec(\mathbf{Q})/\partial\boldsymbol{\omega}'$ depends on the particular parametrisation of orthogonal matrices chosen (see Magnus, Pijls and Sentana (2020)). Given that

$$s_{\mathbf{c}}(\boldsymbol{\theta};\boldsymbol{\varrho})=(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathbf{e}_{st}(\boldsymbol{\phi}),$$

this direct approach allows us to obtain the scores for $\boldsymbol{\sigma}_L$ and $\boldsymbol{\omega}$ as

$$\left[\begin{array}{c}s_{\boldsymbol{\sigma}_L}(\boldsymbol{\theta};\boldsymbol{\varrho})\\ s_{\boldsymbol{\omega}}(\boldsymbol{\theta};\boldsymbol{\varrho})\end{array}\right]=\left(\begin{array}{c}\partial\mathbf{c}'/\partial\boldsymbol{\sigma}_L\\ \partial\mathbf{c}'/\partial\boldsymbol{\omega}\end{array}\right)s_{\mathbf{c}}(\boldsymbol{\theta};\boldsymbol{\varrho})=\left[\begin{array}{c}\mathbf{L}_N(\mathbf{Q}\otimes\mathbf{I}_N)\\ \partial vec'(\mathbf{Q})/\partial\boldsymbol{\omega}\cdot(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L')\end{array}\right]s_{\mathbf{c}}(\boldsymbol{\theta};\boldsymbol{\varrho}).$$

But since $\mathbf{C}=\boldsymbol{\Sigma}_L\mathbf{Q}$ so $\mathbf{C}^{-1}=\mathbf{Q}'\boldsymbol{\Sigma}_L^{-1}$ and $\mathbf{C}^{-1\prime}=\boldsymbol{\Sigma}_L^{-1\prime}\mathbf{Q}$, we have that

$$\begin{aligned}\left[\begin{array}{c}\mathbf{L}_N(\mathbf{Q}\otimes\mathbf{I}_N)\\ \partial vec'(\mathbf{Q})/\partial\boldsymbol{\omega}\cdot(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L')\end{array}\right](\mathbf{I}_N\otimes\mathbf{C}^{-1\prime}) &= \left[\begin{array}{c}\mathbf{L}_N(\mathbf{Q}\otimes\mathbf{I}_N)(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L^{-1\prime}\mathbf{Q})\\ \partial vec'(\mathbf{Q})/\partial\boldsymbol{\omega}\cdot(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L')(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L^{-1\prime}\mathbf{Q})\end{array}\right]\\ &= \left[\begin{array}{c}\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L^{-1\prime})(\mathbf{Q}\otimes\mathbf{Q})\\ \partial vec'(\mathbf{Q})/\partial\boldsymbol{\omega}\cdot(\mathbf{I}_N\otimes\mathbf{Q})\end{array}\right],\end{aligned}$$

whence the expression for the scores and information matrix immediately follows. The dependence of the scores $s_{\boldsymbol{\sigma}_L}(\boldsymbol{\theta};\boldsymbol{\varrho})$ on $\mathbf{Q}$ simply reflects the fact that we have defined $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})=\mathbf{C}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ in terms of the true underlying independent shocks. We explain how to compute $\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L^{-1\prime})$ efficiently at the end of Appendix D.1. □

To obtain the asymptotic variances of $\boldsymbol{\sigma}_L$, we can alternatively use the following two-step procedure. First, we go from the structural loading matrix $\mathbf{C}$ to $\boldsymbol{\Sigma}$. Given that $d\boldsymbol{\Sigma}=(d\mathbf{C})\mathbf{C}'+\mathbf{C}(d\mathbf{C}')$, it immediately follows that

$$\begin{aligned}dvec(\boldsymbol{\Sigma}) &= (\mathbf{C}\otimes\mathbf{I}_N)dvec(\mathbf{C})+(\mathbf{I}_N\otimes\mathbf{C})dvec(\mathbf{C}')\\ &= (\mathbf{C}\otimes\mathbf{I}_N)dvec(\mathbf{C})+(\mathbf{I}_N\otimes\mathbf{C})\mathbf{K}_{NN}dvec(\mathbf{C})=(\mathbf{I}_{N^2}+\mathbf{K}_{NN})(\mathbf{C}\otimes\mathbf{I}_N)dvec(\mathbf{C}),\end{aligned}$$

so that

$$\frac{\partial\boldsymbol{\sigma}}{\partial\mathbf{c}'}=\mathbf{D}_N^+(\mathbf{I}_{N^2}+\mathbf{K}_{NN})(\mathbf{C}\otimes\mathbf{I}_N),$$

where $\mathbf{D}_N^+$ is the Moore-Penrose inverse of the duplication matrix (see Magnus, 1988). Using this Jacobian, the delta method allows us to obtain the asymptotic covariance matrix of the restricted and unrestricted MLEs of the reduced form parameters $\boldsymbol{\sigma}$, but not their scores because $\text{rank}(\partial\boldsymbol{\sigma}/\partial\mathbf{c}')=N(N+1)/2$, so we cannot invert it. Then, we can go from $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma}_L$ by exploiting expression (E13) in Appendix D.1.

Lemma 5

$$[(\mathbf{\Psi}\otimes\mathbf{\Psi}^{-1})(\mathbf{I}_N\otimes\mathbf{J}^{-1})\mathbf{\Delta}_N \vdots \mathbf{E}_N\mathbf{\Psi}^{-1}]^{-1} = \left\{ \begin{array}{c} \mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})[\mathbf{I}_{N^2}-\mathbf{E}_N\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})] \\ \mathbf{\Psi}\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J}) \end{array} \right\}.$$

Proof. Let us look at the four blocks of

$$\left\{ \begin{array}{c} \mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})[\mathbf{I}_{N^2}-\mathbf{E}_N\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})] \\ \mathbf{\Psi}\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J}) \end{array} \right\} [\ (\mathbf{\Psi}\otimes\mathbf{\Psi}^{-1})(\mathbf{I}_N\otimes\mathbf{J}^{-1})\mathbf{\Delta}_N \quad \mathbf{E}_N\mathbf{\Psi}^{-1}\ ].$$

The northwestern block is

$$\begin{aligned} & \mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})(\mathbf{\Psi}\otimes\mathbf{\Psi}^{-1})(\mathbf{I}_N\otimes\mathbf{J}^{-1})\mathbf{\Delta}_N \\ & -\mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})\mathbf{E}_N\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})(\mathbf{\Psi}\otimes\mathbf{\Psi}^{-1})(\mathbf{I}_N\otimes\mathbf{J}^{-1})\mathbf{\Delta}_N \\ = \ & \mathbf{\Delta}_N'\mathbf{\Delta}_N - \mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})\mathbf{E}_N\mathbf{E}_N'\mathbf{\Delta}_N = \mathbf{I}_{N(N-1)} \end{aligned}$$

by virtue of Proposition 4 in Magnus and Sentana (2020). Similarly, the northeastern block is

$$\begin{aligned} & \mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})\mathbf{E}_N\mathbf{\Psi}^{-1} - \mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})\mathbf{E}_N\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})\mathbf{E}_N\mathbf{\Psi}^{-1} \\ = \ & \mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})\mathbf{E}_N\mathbf{\Psi}^{-1} - \mathbf{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J})\mathbf{E}_N\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})\mathbf{E}_N\mathbf{\Psi}^{-1} = \mathbf{0} \end{aligned}$$

thanks to Propositions 2 and 3 in Magnus and Sentana (2020), together with the fact that the diagonal elements of $\mathbf{J}$ are normalised to 1. The same propositions also imply that the southwestern block will be

$$\mathbf{\Psi}\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}\otimes\mathbf{\Psi}^{-1})(\mathbf{I}_N\otimes\mathbf{J}^{-1})\mathbf{\Delta}_N = \mathbf{\Psi}\mathbf{E}_N'\mathbf{\Delta}_N = \mathbf{0},$$

while the sourtheastern one

$$\mathbf{\Psi}\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J})\mathbf{E}_N\mathbf{\Psi}^{-1} = \mathbf{\Psi}(\mathbf{I}_N\odot\mathbf{J})\mathbf{\Psi}^{-1} = \mathbf{I}_N,$$

as claimed. □

C The special case of spherical distributions

C.1 Some useful distribution results

A spherically symmetric random vector of dimension N, $\boldsymbol{\varepsilon}_t^\bullet$, is fully characterised in Theorem 2.5 (iii) of Fang, Kotz and Ng (1990) as $\boldsymbol{\varepsilon}_t^\bullet = e_t\mathbf{u}_t$, where $\mathbf{u}_t$ is uniformly distributed on the unit sphere surface in $\mathbb{R}^N$, and e_t is a non-negative random variable independent of $\mathbf{u}_t$, whose distribution determines the distribution of $\boldsymbol{\varepsilon}_t^\bullet$. The variables e_t and $\mathbf{u}_t$ are referred to as the generating variate and the uniform base of the spherical distribution. Assuming that $E(e_t^2) < \infty$, we can standardise $\boldsymbol{\varepsilon}_t^\bullet$ by setting $E(e_t^2) = N$, so that $E(\boldsymbol{\varepsilon}_t^\bullet) = \mathbf{0}$, $V(\boldsymbol{\varepsilon}_t^\bullet) = \mathbf{I}_N$. Specifically, if $\boldsymbol{\varepsilon}_t^\bullet$ is distributed as a standardised multivariate Student t random vector of dimension N with ν_0 degrees of freedom, then $e_t = \sqrt{(\nu_0-2)\zeta_t/\xi_t}$, where ζ_t is a chi-square random variable with N degrees of freedom, and ξ_t is an independent Gamma variate with mean $\nu_0 > 2$ and variance

$2\nu_0$. If we further assume that $E(e_t^4) < \infty$, then the coefficient of multivariate excess kurtosis κ_0, which is given by $E(e_t^4)/[N(N+2)] - 1$, will also be bounded. For instance, $\kappa_0 = 2/(\nu_0 - 4)$ in the Student t case with $\nu_0 > 4$, and $\kappa_0 = 0$ under normality. In this respect, note that since $E(e_t^4) \geq E^2(e_t^2) = N^2$ by the Cauchy-Schwarz inequality, with equality if and only if $e_t = \sqrt{N}$ so that $\boldsymbol{\varepsilon}_t^\bullet$ is proportional to $\mathbf{u}_t$, then $\kappa_0 \geq -2/(N+2)$, the minimum value being achieved in the uniformly distributed case.

Then, it is easy to combine the representation of spherical distributions above with the higher order moments of a multivariate normal vector in Balestra and Holly (1990) to prove that the third and fourth moments of a spherically symmetric distribution with $V(\boldsymbol{\varepsilon}_t^\bullet) = \mathbf{I}_N$ are given by

$$E(\boldsymbol{\varepsilon}_t^\bullet \boldsymbol{\varepsilon}_t^{\bullet\prime} \otimes \boldsymbol{\varepsilon}_t^\bullet) = \mathbf{0}, \tag{C1}$$

$$E(\boldsymbol{\varepsilon}_t^\bullet \boldsymbol{\varepsilon}_t^{\bullet\prime} \otimes \boldsymbol{\varepsilon}_t^\bullet \boldsymbol{\varepsilon}_t^{\bullet\prime}) = E[vec(\boldsymbol{\varepsilon}_t^\bullet \boldsymbol{\varepsilon}_t^{\bullet\prime}) vec'(\boldsymbol{\varepsilon}_t^\bullet \boldsymbol{\varepsilon}_t^\bullet)] = (\kappa_0 + 1)[(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + vec(\mathbf{I}_N)\, vec'(\mathbf{I}_N)], \tag{C2}$$

where $\mathbf{K}_{mn}$ is the commutation matrix of orders m and n (see e.g. Magnus and Neudecker (2019)).

C.2 Likelihood, score and Hessian for spherically symmetric distributions

Let $\exp[c(\boldsymbol{\eta}) + g(\varsigma_t, \boldsymbol{\eta})]$ denote the assumed conditional density of $\boldsymbol{\varepsilon}_t^*$ given I_{t-1} and the shape parameters, where $c(\boldsymbol{\eta})$ corresponds to the constant of integration, $g(\varsigma_t, \boldsymbol{\eta})$ to its kernel and $\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$. Ignoring initial conditions, the log-likelihood function of a sample of size T for those values of $\boldsymbol{\theta}$ for which $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ has full rank will take the form $L_T(\boldsymbol{\phi}) = \sum_{t=1}^T l_t(\boldsymbol{\phi})$, where $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + c(\boldsymbol{\eta}) + g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]$, $d_t(\boldsymbol{\theta}) = \ln|\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})|$ is the Jacobian, $\varsigma_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta})$.

Let $\mathbf{s}_t(\boldsymbol{\phi})$ denote the score function $\partial l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$, and partition it into two blocks, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ and $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})$, whose dimensions conform to those of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, respectively. If $\boldsymbol{\mu}_t(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, $c(\boldsymbol{\eta})$ and $g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]$ are differentiable, then

$$\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) = \partial c(\boldsymbol{\eta})/\partial\boldsymbol{\eta} + \partial g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]/\partial\boldsymbol{\eta} = \mathbf{e}_{rt}(\boldsymbol{\phi}), \tag{C3}$$

while

$$\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) = \frac{\partial d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} + \frac{\partial g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]}{\partial\varsigma}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = [\mathbf{Z}_{lt}(\boldsymbol{\theta}), \mathbf{Z}_{st}(\boldsymbol{\theta})]\begin{bmatrix}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{e}_{st}(\boldsymbol{\phi})\end{bmatrix} = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}), \tag{C4}$$

where

$$\begin{aligned}
\partial d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} &= -\mathbf{Z}_{st}(\boldsymbol{\theta})vec(\mathbf{I}_N),\\
\partial\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} &= -2\{\mathbf{Z}_{lt}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) + \mathbf{Z}_{st}(\boldsymbol{\theta})vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right]\}, && \text{(C5)}\\
\mathbf{Z}_{lt}(\boldsymbol{\theta}) &= \partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\cdot\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}), && \text{(C6)}\\
\mathbf{Z}_{st}(\boldsymbol{\theta}) &= \frac{1}{2}\partial vec'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}\cdot[\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})], && \text{(C7)}\\
\mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\eta}) &= \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}), && \text{(C8)}\\
\mathbf{e}_{st}(\boldsymbol{\theta}, \boldsymbol{\eta}) &= vec\left\{\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) - \mathbf{I}_N\right\}, && \text{(C9)}
\end{aligned}$$

and

$$\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] = -2\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]/\partial\varsigma \tag{C10}$$

is a damping factor that reflects the tail-thickness of the distribution assumed for estimation purposes. Importantly, while both $\mathbf{Z}_{dt}(\boldsymbol{\theta})$ and $\mathbf{e}_{dt}(\boldsymbol{\phi})$ depend on the specific choice of square root matrix $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ does not, a property that inherits from $l_t(\boldsymbol{\phi})$. As we shall see in Supplemental Appendix D, this result is not generally true for non-spherical distributions.

Obviously, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})$ reduces to the multivariate normal expression in Bollerslev and Wooldridge (1992), in which case:

$$\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0}) = \begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0}) \\ \mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0}) \end{bmatrix} = \left\{ \begin{array}{c} \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \\ vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})-\mathbf{I}_N\right] \end{array} \right\}.$$

Assuming further twice differentiability of the different functions involved, we will have that the Hessian function $\mathbf{h}_t(\boldsymbol{\phi}) = \partial\mathbf{s}_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}' = \partial^2 l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'$ will be

$$\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) = \frac{\partial^2 d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + \frac{\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{(\partial\varsigma)^2}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + \frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\varsigma}\frac{\partial^2\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}, \tag{C11}$$

$$\mathbf{h}_{\boldsymbol{\theta\eta}t}(\boldsymbol{\phi}) = \partial\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\cdot\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\varsigma\partial\boldsymbol{\eta}', \tag{C12}$$

$$\mathbf{h}_{\boldsymbol{\eta\eta}t}(\boldsymbol{\phi}) = \partial^2 c(\boldsymbol{\eta})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}' + \partial^2 g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}',$$

where

$$\partial^2 d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' = 2\mathbf{Z}_{st}(\boldsymbol{\theta})\mathbf{Z}_{st}'(\boldsymbol{\theta}) - \frac{1}{2}\left\{vec'\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right]\otimes\mathbf{I}_p\right\}\partial vec\left\{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}\right\}/\partial\boldsymbol{\theta}', \tag{C13}$$

$$\begin{aligned}\partial^2\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' &= 2\mathbf{Z}_{lt}(\boldsymbol{\theta})\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + 8\mathbf{Z}_{st}(\boldsymbol{\theta})[\mathbf{I}_N\otimes\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})]\mathbf{Z}_{st}'(\boldsymbol{\theta}) + 4\mathbf{Z}_{lt}(\boldsymbol{\theta})[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta}) \\ &+4\mathbf{Z}_{st}(\boldsymbol{\theta})[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{lt}'(\boldsymbol{\theta}) - 2[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_p]\partial vec[\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}]\partial\boldsymbol{\theta}' \\ &-\{vec'[\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})]\otimes\mathbf{I}_p\}\partial vec\{\partial vec'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}\}/\partial\boldsymbol{\theta}'.\end{aligned}$$

Note that $\partial\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, $\partial^2 d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ and $\partial^2\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ depend on the dynamic model specification, while $\partial^2 g(\varsigma,\eta)/(\partial\varsigma)^2$, $\partial^2 g(\varsigma,\eta)/\partial\varsigma\partial\boldsymbol{\eta}'$ and $\partial g(\varsigma,\eta)/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'$ depend on the specific spherical distribution assumed for estimation purposes (see Fiorentini, Sentana and Calzolari (2003) for expressions for $\delta(\varsigma_t,\boldsymbol{\eta})$, $c(\boldsymbol{\eta})$, $g(\varsigma_t,\boldsymbol{\eta})$ and its derivatives in the multivariate Student t case, Amengual and Sentana (2010) for the Kotz distribution and discrete scale mixture of normals, and Amengual, Fiorentini and Sentana (2013) for polynomial expansions).

C.3 Asymptotic distribution

Given correct specification, the results in Crowder (1976) imply that $\mathbf{e}_t(\boldsymbol{\phi}) = [\mathbf{e}_{dt}'(\boldsymbol{\phi}), \mathbf{e}_{rt}(\boldsymbol{\phi})]'$ evaluated at $\boldsymbol{\phi}_0$ follows a vector martingale difference, and therefore, the same is true of the score vector $\mathbf{s}_t(\boldsymbol{\phi})$. His results also imply that, under suitable regularity conditions, the asymptotic distribution of the joint ML estimator will be $\sqrt{T}(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0) \to N\left[\mathbf{0}, \mathcal{I}^{-1}(\boldsymbol{\phi}_0)\right]$, where $\mathcal{I}(\boldsymbol{\phi}_0) = E[\mathcal{I}_t(\boldsymbol{\phi}_0)|\boldsymbol{\phi}_0]$,

$$\mathcal{I}_t(\boldsymbol{\phi}) = V\left[\mathbf{s}_t(\boldsymbol{\phi})|I_{t-1};\boldsymbol{\phi}\right] = \mathbf{Z}_t(\boldsymbol{\theta})\mathcal{M}(\boldsymbol{\phi})\mathbf{Z}_t'(\boldsymbol{\theta}) = -E\left[\mathbf{h}_t(\boldsymbol{\phi})|I_{t-1};\boldsymbol{\phi}\right],$$

$$\mathbf{Z}_t(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{Z}_{dt}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{pmatrix}, \tag{C14}$$

and $\mathcal{M}(\boldsymbol{\phi}) = V[\mathbf{e}_t(\boldsymbol{\phi})|\boldsymbol{\phi}]$. In particular, Crowder (1976) requires: (i) $\boldsymbol{\phi}_0$ is locally identified and belongs to the interior of the admissible parameter space, which is a compact subset of $\mathbb{R}^{p+q}$; (ii) the Hessian matrix is non-singular and continuous throughout some neighbourhood of $\boldsymbol{\phi}_0$; (iii) there is uniform convergence to the integrals involved in the computation of the mean vector and covariance matrix of $\mathbf{s}_t(\boldsymbol{\phi})$; and (iv) $-E^{-1}\left[-T^{-1}\sum_t \mathbf{h}_t(\boldsymbol{\phi})\right] T^{-1}\sum_t \mathbf{h}_t(\boldsymbol{\phi}) \xrightarrow{p} \mathbf{I}_{p+q}$, where $E^{-1}\left[-T^{-1}\sum_t \mathbf{h}_t(\boldsymbol{\phi})\right]$ is positive definite on a neighbourhood of $\boldsymbol{\phi}_0$.

As for $\tilde{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}})$, assuming that $\bar{\boldsymbol{\eta}}$ coincides with the true value of this parameter vector, the same arguments imply that $\sqrt{T}[\tilde{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}}) - \boldsymbol{\theta}_0] \to N\left[\mathbf{0}, \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\right]$, where $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)$ is the relevant block of the information matrix.

The next proposition, which originally appeared as Proposition 1 in Fiorentini and Sentana (2007), generalises Propositions 3 in Lange, Little and Taylor (1989), 1 in Fiorentini, Sentana and Calzolari (2003) and 5.2 in Hafner and Rombouts (2007), providing detailed expressions for $\mathcal{M}(\boldsymbol{\phi})$ in models with non-zero conditional means:

Proposition C1 *If $\boldsymbol{\varepsilon}^*_t|I_{t-1};\boldsymbol{\phi}$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta})$ with density* $\exp[c(\boldsymbol{\eta}) + g(\varsigma_t, \boldsymbol{\eta})]$, *then*

$$\mathcal{M}(\boldsymbol{\eta}) = \begin{pmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{M}_{ss}(\boldsymbol{\eta}) & \mathcal{M}_{sr}(\boldsymbol{\eta}) \\ \mathbf{0} & \mathcal{M}'_{sr}(\boldsymbol{\eta}) & \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{pmatrix}, \tag{C15}$$

$$\mathcal{M}_{ll}(\boldsymbol{\eta}) = \mathrm{M}_{ll}(\boldsymbol{\eta})\mathbf{I}_N, \tag{C16}$$

$$\mathcal{M}_{ss}(\boldsymbol{\eta}) = \mathrm{M}_{ss}(\boldsymbol{\eta})\left(\mathbf{I}_{N^2} + \mathbf{K}_{NN}\right) + [\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1]vec(\mathbf{I}_N)vec'(\mathbf{I}_N), \tag{C17}$$

$$\mathcal{M}_{sr}(\boldsymbol{\eta}) = vec(\mathbf{I}_N)\mathrm{M}_{sr}(\boldsymbol{\eta}), \tag{C18}$$

$$\mathrm{M}_{ll}(\boldsymbol{\eta}) = E\left[\delta^2(\varsigma_t,\boldsymbol{\eta})\frac{\varsigma_t}{N}\bigg|\boldsymbol{\eta}\right] = E\left[\frac{2\partial\delta(\varsigma_t,\boldsymbol{\eta})}{\partial\varsigma}\frac{\varsigma_t}{N} + \delta(\varsigma_t,\boldsymbol{\eta})\bigg|\boldsymbol{\eta}\right],$$

$$\mathrm{M}_{ss}(\boldsymbol{\eta}) = \frac{N}{N+2}\left\{1 + V\left[\delta(\varsigma_t,\boldsymbol{\eta})\frac{\varsigma_t}{N}\bigg|\boldsymbol{\eta}\right]\right\} = \frac{N}{N+2}E\left[\frac{2\partial\delta(\varsigma_t,\boldsymbol{\eta})}{\partial\varsigma}\left(\frac{\varsigma_t}{N}\right)^2\bigg|\boldsymbol{\eta}\right] + 1,$$

$$\mathrm{M}_{sr}(\boldsymbol{\eta}) = E\left\{\left[\delta(\varsigma_t,\boldsymbol{\eta})\frac{\varsigma_t}{N} - 1\right]\mathbf{e}'_{rt}(\boldsymbol{\phi})\bigg|\boldsymbol{\phi}\right\} = -E\left[\frac{\varsigma_t}{N}\frac{\partial\delta(\varsigma_t,\boldsymbol{\eta})}{\partial\boldsymbol{\eta}'}\bigg|\boldsymbol{\eta}\right].$$

Proof. For our purposes it is convenient to rewrite $\mathbf{e}_{dt}(\boldsymbol{\phi}_0)$ as

$$\begin{aligned} \mathbf{e}_{lt}(\boldsymbol{\phi}_0) &= \delta[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0]\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0) = \delta(\varsigma_t,\boldsymbol{\eta}_0)\sqrt{\varsigma_t}\mathbf{u}_t, \\ \mathbf{e}_{st}(\boldsymbol{\phi}_0) &= vec\left\{\delta[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0]\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta}_0) - \mathbf{I}_N\right\} = vec\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\varsigma_t\mathbf{u}_t\mathbf{u}'_t - \mathbf{I}_N\right], \end{aligned}$$

where ς_t and $\mathbf{u}_t$ are mutually independent for any standardised spherical distribution, with $E(\mathbf{u}_t) = \mathbf{0}$, $E(\mathbf{u}_t\mathbf{u}'_t) = N^{-1}\mathbf{I}_N$, $E(\varsigma_t) = N$ and $E(\varsigma_t^2) = N(N+2)(\kappa_0 + 1)$. Importantly, we only need to compute unconditional moments because ς_t and $\mathbf{u}_t$ are independent of $\mathbf{z}_t$ and I_{t-1} by assumption. Then, it easy to see that

$$E[\mathbf{e}_{lt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = E[\delta(\varsigma_t,\boldsymbol{\eta})\sqrt{\varsigma_t}|\boldsymbol{\eta}]\cdot E(\mathbf{u}_t) = \mathbf{0},$$

and that

$$E[\mathbf{e}_{st}(\boldsymbol{\phi})|\boldsymbol{\phi}] = vec\left\{E\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\varsigma_t|\boldsymbol{\eta}\right]\cdot E(\mathbf{u}_t\mathbf{u}'_t) - \mathbf{I}_N\right\} = vec(\mathbf{I}_N)\left\{E\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)(\varsigma_t/N)|\boldsymbol{\eta}\right] - 1\right\}.$$

In this context, we can use expression (2.21) in Fang, Kotz and Ng (1990) to write the density function of ς_t as

$$h(\varsigma_t;\boldsymbol{\eta}) = \frac{\pi^{N/2}}{\Gamma(N/2)}\varsigma_t^{N/2-1}\exp[c(\boldsymbol{\eta}) + g(\varsigma_t,\boldsymbol{\eta})], \tag{C19}$$

whence

$$[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N) - 1] = -\frac{2}{N}\left[1 + \varsigma_t \cdot \partial \ln h(\varsigma_t;\boldsymbol{\eta})/\partial\varsigma\right]. \tag{C20}$$

On this basis, we can use Lemma 2 in Supplemental Appendix B to show that $E(\varsigma_t) = N < \infty$ implies

$$E\left[\varsigma_t \cdot \partial \ln h(\varsigma_t;\boldsymbol{\eta})/\partial\varsigma|\boldsymbol{\eta}\right] = -E\left[1\right] = -1,$$

which in turn implies that

$$E\left[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N) - 1|\boldsymbol{\eta}\right] = 0 \tag{C21}$$

in view of (C20). Consequently, $E[\mathbf{e}_{st}(\boldsymbol{\phi})|\boldsymbol{\phi}] = \mathbf{0}$, as required.

Similarly, we can also show that

$$\begin{aligned}
E[\mathbf{e}_{lt}(\boldsymbol{\phi})\mathbf{e}'_{lt}(\boldsymbol{\phi})|\boldsymbol{\phi}] &= E\left\{\delta^2(\varsigma_t,\boldsymbol{\eta})\varsigma_t\mathbf{u}_t\mathbf{u}'_t|\boldsymbol{\eta}\right\} = \mathbf{I}_N \cdot E[\delta^2(\varsigma_t,\boldsymbol{\eta}_0)(\varsigma_t/N)|\boldsymbol{\eta}],\\
E[\mathbf{e}_{lt}(\boldsymbol{\phi})\mathbf{e}'_{st}(\boldsymbol{\phi})|\boldsymbol{\phi}] &= E\left\{\delta(\varsigma_t,\boldsymbol{\eta})\sqrt{\varsigma_t}\mathbf{u}_t vec'\left[\delta(\varsigma_t,\boldsymbol{\eta})\varsigma_t\mathbf{u}_t\mathbf{u}'_t - \mathbf{I}_N\right]|\boldsymbol{\eta}\right\} = \mathbf{0}
\end{aligned}$$

by virtue of (C1), and

$$\begin{aligned}
E[\mathbf{e}_{st}(\boldsymbol{\phi}_0)\mathbf{e}'_{st}(\boldsymbol{\phi}_0)|\boldsymbol{\phi}] &= E\left\{vec\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\varsigma_t\mathbf{u}_t\mathbf{u}'_t - \mathbf{I}_N\right] vec'\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\varsigma_t\mathbf{u}_t\mathbf{u}'_t - \mathbf{I}_N\right]|\boldsymbol{\eta}\right\}\\
&= E\left[\delta(\varsigma_t,\boldsymbol{\eta})\varsigma_t|\boldsymbol{\eta}\right]^2 \frac{1}{N(N+2)}[(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + vec\left(\mathbf{I}_N\right) vec'\left(\mathbf{I}_N\right)]\\
&\quad -2E\left[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N)|\boldsymbol{\eta}\right] vec\left(\mathbf{I}_N\right) vec'\left(\mathbf{I}_N\right) + vec\left(\mathbf{I}_N\right) vec'\left(\mathbf{I}_N\right)\\
&= \frac{N}{(N+2)}E\left[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N)|\boldsymbol{\eta}\right]^2(\mathbf{I}_{N^2} + \mathbf{K}_{NN})\\
&\quad + \left\{\frac{N}{(N+2)}E\left[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N)|\boldsymbol{\eta}\right]^2 - 1\right\} vec\left(\mathbf{I}_N\right) vec'\left(\mathbf{I}_N\right)]
\end{aligned}$$

by virtue of (C2), (C20) and (C21).

Finally, it is clear from (C3) that $\mathbf{e}_{rt}(\boldsymbol{\phi}_0)$ will be a function of ς_t but not of $\mathbf{u}_t$, which immediately implies that $E[\mathbf{e}_{lt}(\boldsymbol{\phi})\mathbf{e}'_{rt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = \mathbf{0}$, and that

$$\begin{aligned}
E[\mathbf{e}_{st}(\boldsymbol{\phi})\mathbf{e}'_{rt}(\boldsymbol{\phi})|\boldsymbol{\phi}] &= E\left\{vec\left[\delta(\varsigma_t,\boldsymbol{\eta})\varsigma_t \cdot \mathbf{u}_t\mathbf{u}'_t - \mathbf{I}_N\right]\mathbf{e}'_{rt}(\boldsymbol{\phi})\right\}\\
&= vec(\mathbf{I}_N)E\left\{\left[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N) - 1\right]\mathbf{e}'_{rt}(\boldsymbol{\phi})\right\}.
\end{aligned}$$

To obtain the expected value of the Hessian, it is also convenient to write $\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$ in (C11) as

$$
\begin{aligned}
&-4\mathbf{Z}_{st}(\boldsymbol{\theta}_0)[\mathbf{I}_N \otimes \{\delta[\varsigma_t(\boldsymbol{\theta}_0),\eta_0]\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)-\mathbf{I}_N\}]\mathbf{Z}_{st}'(\boldsymbol{\theta}_0)\\
&+[\mathbf{e}_{lt}'(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_p]\frac{\partial vec}{\partial\boldsymbol{\theta}'}\left[\frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]\\
&+\frac{1}{2}\{\mathbf{e}_{st}'(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)[\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\otimes\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)]\otimes\mathbf{I}_p\}\frac{\partial vec}{\partial\boldsymbol{\theta}'}\left\{\frac{\partial vec'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}}\right\}\\
&-2\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)[\mathbf{e}_{lt}'(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)\otimes\mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta}_0)-2\mathbf{Z}_{st}(\boldsymbol{\theta}_0)[\mathbf{e}_{lt}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)\otimes\mathbf{I}_N]\mathbf{Z}_{lt}'(\boldsymbol{\theta}_0)\\
&-\delta[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0]\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\mathbf{Z}_{lt}'(\boldsymbol{\theta}_0)-2\mathbf{Z}_{st}(\boldsymbol{\theta}_0)\mathbf{Z}_{st}'(\boldsymbol{\theta}_0)-\frac{2\partial\delta\left[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0\right]}{\partial\varsigma}\{\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)\mathbf{Z}_{lt}'(\boldsymbol{\theta}_0)\\
&+\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)vec'[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)]\mathbf{Z}_{st}'(\boldsymbol{\theta}_0)+\mathbf{Z}_{st}(\boldsymbol{\theta}_0)vec[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)]\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\mathbf{Z}_{lt}'(\boldsymbol{\theta}_0)\\
&+\mathbf{Z}_{st}(\boldsymbol{\theta}_0)vec[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)]vec'[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)]\mathbf{Z}_{st}'(\boldsymbol{\theta}_0)\big\}\,.
\end{aligned}
$$

Clearly, the first four lines have zero conditional expectation, and the same is true of the sixth line by virtue of (C1). As for the remaining terms, we can write them as

$$
\begin{aligned}
&-\delta(\varsigma_t,\boldsymbol{\eta}_0)\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\mathbf{Z}_{lt}'(\boldsymbol{\theta}_0)-2\partial\delta(\varsigma_t,\boldsymbol{\eta}_0)/\partial\varsigma\cdot\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\varsigma_t\mathbf{u}_t\mathbf{u}_t'\mathbf{Z}_{lt}'(\boldsymbol{\theta}_0)\\
&-2\mathbf{Z}_{st}(\boldsymbol{\theta}_0)\mathbf{Z}_{st}'(\boldsymbol{\theta}_0)-2\partial\delta(\varsigma_t,\boldsymbol{\eta}_0)/\partial\varsigma\cdot\varsigma_t^2\mathbf{Z}_{st}(\boldsymbol{\theta}_0)vec(\mathbf{u}_t\mathbf{u}_t')vec'(\mathbf{u}_t\mathbf{u}_t')\mathbf{Z}_{st}'(\boldsymbol{\theta}_0),
\end{aligned}
$$

whose conditional expectation will be

$$
\begin{aligned}
&-\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\mathbf{Z}_{lt}'(\boldsymbol{\theta}_0)E[\delta(\varsigma_t;\boldsymbol{\eta}_0)+2(\varsigma_t/N)\cdot\partial\delta(\varsigma_t,\boldsymbol{\eta}_0)/\partial\varsigma|\boldsymbol{\eta}_0]-2\mathbf{Z}_{st}(\boldsymbol{\theta}_0)\mathbf{Z}_{st}'(\boldsymbol{\theta}_0)\\
&-\mathbf{Z}_{st}(\boldsymbol{\theta}_0)\frac{2E[\varsigma_t^2\cdot\partial\delta(\varsigma_t,\boldsymbol{\eta}_0)/\partial\varsigma|\boldsymbol{\eta}_0]}{N(N+2)}[(\mathbf{I}_{N^2}\otimes\mathbf{K}_{NN})+vec(\mathbf{I}_N)vec'(\mathbf{I}_N)]\mathbf{Z}_{st}'(\boldsymbol{\theta}_0).
\end{aligned}
$$

As for $\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\eta}t}(\boldsymbol{\phi}_0)$, it follows from (C5) and (C12) that we can write it as

$$
\begin{aligned}
&\{\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)+\mathbf{Z}_{st}(\boldsymbol{\theta}_0)vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)\right]\}\cdot\partial\delta\left[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0\right]/\partial\boldsymbol{\eta}'\\
=\;&[\mathbf{Z}_{lt}(\boldsymbol{\theta})\mathbf{u}_t\sqrt{\varsigma_t}+\mathbf{Z}_{st}(\boldsymbol{\theta})vec(\mathbf{u}_t\mathbf{u}_t')\varsigma_t]\cdot\partial\delta(\varsigma_t,\boldsymbol{\eta})/\partial\boldsymbol{\eta}',
\end{aligned}
$$

whose conditional expected value will be $\mathbf{Z}_{st}(\boldsymbol{\theta}_0)vec(\mathbf{I}_N)E[(\varsigma_t/N)\cdot\partial\delta(\varsigma_t,\boldsymbol{\eta}_0)/\partial\boldsymbol{\eta}'|\boldsymbol{\eta}]$. □

Fiorentini, Sentana and Calzolari (2003) provide the relevant expressions for the multivariate standardised Student t, while the expressions for the Kotz distribution and the DSMN are given in Amengual and Sentana (2010) (The expression for $\text{M}_{ss}(\kappa)$ for the Kotz distribution in Amengual and Sentana (2010) contains a typo. The correct value is $(N\kappa+2)/[(N+2)\kappa+2]$).

As for $\mathcal{I}(\boldsymbol{\phi}_0)$, while it is relatively straightforward to obtain closed-form expressions in conditionally homoskedastic, dynamic linear models such as multivariate regressions or VARs (see e.g. Amengual and Sentana (2010)), it is virtually impossible to do so in dynamic conditionally heteroskedastic models, as one has to resort to numerical or Monte Carlo integration methods to compute the required expected values (see e.g. Engle and Gonzalez-Rivera (1991) and Gonzalez-Rivera and Drost (1999)). Nevertheless, see Fiorentini and Sentana (2015, 2018) for closed-form expressions in the context of tests for univariate or multivariate conditional homoskedasticity, respectively.

C.4 Gaussian pseudo maximum likelihood estimators

An important special case of restricted ML estimator arises when $\bar{\boldsymbol{\eta}} = \mathbf{0}$, in which case $\tilde{\boldsymbol{\theta}}_T(\mathbf{0})$ coincides with the Gaussian PML estimator $\tilde{\boldsymbol{\theta}}_T$. Unlike what happens with other values of $\bar{\boldsymbol{\eta}}$, $\tilde{\boldsymbol{\theta}}_T$ remains root-T consistent for $\boldsymbol{\theta}_0$ under correct specification of $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ even though the true conditional distribution of $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}_0$ is neither Gaussian nor spherical, provided that it has bounded fourth moments. The proof is based on the fact that in those circumstances, the pseudo log-likelihood score, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})$, is also a vector martingale difference sequence when evaluated at $\boldsymbol{\theta}_0$, a property that inherits from

$$\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0}) = \begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0}) \\ \mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0}) \end{bmatrix} = \left\{ \begin{array}{c} \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \\ vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) - \mathbf{I}_N\right] \end{array} \right\}.$$

Importantly, this property is preserved even when the standardised innovations, $\boldsymbol{\varepsilon}_t^*$, are not stochastically independent of I_{t-1}.

The asymptotic distribution of the PML estimator of $\boldsymbol{\theta}$ is stated in the following result, which specialises Proposition 1 in Bollerslev and Wooldridge (1992) to models with *i.i.d.* innovations with shape parameters $\boldsymbol{\varrho}$:

Proposition C2 *Assume that the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied.*

1. *If* $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}$ *is i.i.d.* $D(\mathbf{0},\mathbf{I}_N,\boldsymbol{\varrho})$ *with* $tr[\mathcal{K}(\boldsymbol{\varrho})]<\infty$, *where* $\boldsymbol{\phi} = (\boldsymbol{\theta}',\boldsymbol{\varrho}')'$, *then* $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \to N\left[\mathbf{0}, \mathcal{C}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_0,\mathbf{0};\boldsymbol{\phi}_0)\right]$ *with*

$$\begin{aligned}
\mathcal{C}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi}) &= \mathcal{A}^{-1}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi})\mathcal{B}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi})\mathcal{A}^{-1}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi}),\\
\mathcal{A}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi}) &= -E\left[\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathcal{A}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi})|\boldsymbol{\phi}\right],\\
\mathcal{A}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi}) &= -E[\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\theta};\mathbf{0})|\, I_{t-1};\boldsymbol{\phi}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\mathbf{0})\mathbf{Z}'_{dt}(\boldsymbol{\theta}),\\
\mathcal{B}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi}) &= V\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathcal{B}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi})|\boldsymbol{\phi}\right],\\
\mathcal{B}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi}) &= V[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta};\mathbf{0})|\, I_{t-1};\boldsymbol{\phi}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\boldsymbol{\varrho})\mathbf{Z}'_{dt}(\boldsymbol{\theta}),
\end{aligned}$$

   and

$$\mathcal{K}(\boldsymbol{\varrho}) = V[\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})|\, I_{t-1};\boldsymbol{\phi}] = \begin{bmatrix} \mathbf{I}_N & \boldsymbol{\Phi}(\boldsymbol{\varrho}) \\ \boldsymbol{\Phi}'(\boldsymbol{\varrho}) & \boldsymbol{\Upsilon}(\boldsymbol{\varrho}) \end{bmatrix}, \tag{C22}$$

   where

$$\begin{aligned}
\boldsymbol{\Phi}(\boldsymbol{\varrho}) &= E[\boldsymbol{\varepsilon}_t^* vec'(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})|\, \boldsymbol{\phi}]\\
\boldsymbol{\Upsilon}(\boldsymbol{\varrho}) &= E[vec(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime} - \mathbf{I}_N)vec'(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime} - \mathbf{I}_N)|\, \boldsymbol{\phi}]
\end{aligned}$$

   depend on the multivariate third and fourth order cumulants of $\boldsymbol{\varepsilon}_t^*$, *so that* $\boldsymbol{\Phi}(\mathbf{0}) = \mathbf{0}$ *and* $\boldsymbol{\Upsilon}(\mathbf{0}) = (\mathbf{I}_{N^2} + \mathbf{K}_{NN})$ *if we use* $\boldsymbol{\varrho} = \mathbf{0}$ *to denote normality.*

2. *If* $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}_0$ *is i.i.d.* $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta}_0)$ *with* $\kappa_0 < \infty$, *then (C22) reduces to*

$$\mathcal{K}(\kappa) = \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & (\kappa+1)(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + \kappa vec(\mathbf{I}_N)vec'(\mathbf{I}_N) \end{bmatrix}, \tag{C23}$$

   which only depends on the true distribution through the population coefficient of multivariate excess kurtosis

$$\kappa = E(\varsigma_t^2|\boldsymbol{\eta})/[N(N+2)] - 1. \tag{C24}$$

Proof. The proof of the first part is based on a straightforward application of Proposition 1 in Bollerslev and Wooldridge (1992) to the *i.i.d.* case. Since $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0}) = \mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$, and $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$ is a vector martingale difference sequence, then to obtain $\mathcal{B}_t(\boldsymbol{\phi}_0)$ we only need to compute $V[\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})|I_{t-1};\boldsymbol{\phi}_0]$, which justifies (C22). Further, we will have that

$$\begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\theta}_0,\mathbf{0}) \\ \mathbf{e}_{st}(\boldsymbol{\theta}_0,\mathbf{0}) \end{bmatrix} = \begin{pmatrix} \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0) \\ vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0) - \mathbf{I}_N\right] \end{pmatrix} = \begin{bmatrix} \sqrt{\varsigma_t}\mathbf{u}_t \\ vec(\varsigma_t\mathbf{u}_t\mathbf{u}_t' - \mathbf{I}_N) \end{bmatrix}$$

for any spherical distribution, with ς_t and $\mathbf{u}_t$ both mutually and serially independent. Then (C23) follows from (C1) and (C2). As for $\mathcal{A}_t(\boldsymbol{\phi}_0)$, we know that its formula, which is valid regardless of the exact nature of the true conditional distribution, coincides with the expression for $\mathcal{B}_t(\boldsymbol{\phi}_0)$ under multivariate normality by the (conditional) information matrix equality. □

C.5 Spherically symmetric semiparametric estimators

As is well known, a single scoring iteration without line searches that started from $\tilde{\boldsymbol{\theta}}_T$ and some root-T consistent estimator of $\boldsymbol{\eta}$, say $\tilde{\boldsymbol{\eta}}_T$, would suffice to yield an estimator of $\boldsymbol{\phi}$ that would be asymptotically equivalent to the full-information ML estimator $\hat{\boldsymbol{\phi}}_T$, at least up to terms of order $O_p(T^{-1/2})$. Specifically,

$$\begin{pmatrix} \check{\boldsymbol{\theta}}_T - \tilde{\boldsymbol{\theta}}_T \\ \check{\boldsymbol{\eta}}_T - \tilde{\boldsymbol{\eta}}_T \end{pmatrix} = \begin{bmatrix} \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0) \\ \mathcal{I}_{\boldsymbol{\theta\eta}}'(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0) \end{bmatrix}^{-1} \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} \mathbf{s}_{\boldsymbol{\theta}t}(\tilde{\boldsymbol{\theta}}_T,\tilde{\boldsymbol{\eta}}_T) \\ \mathbf{s}_{\boldsymbol{\eta}t}(\tilde{\boldsymbol{\theta}}_T,\tilde{\boldsymbol{\eta}}_T) \end{bmatrix}.$$

If we use the partitioned inverse formula, then it is easy to see that

$$\check{\boldsymbol{\theta}}_T - \tilde{\boldsymbol{\theta}}_T = \left[\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}'(\boldsymbol{\phi}_0)\right]^{-1}$$
$$\times\frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{s}_{\boldsymbol{\theta}t}(\tilde{\boldsymbol{\theta}}_T,\tilde{\boldsymbol{\eta}}_T) - \mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\eta}t}(\tilde{\boldsymbol{\theta}}_T,\tilde{\boldsymbol{\eta}}_T)\right] = \mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\frac{1}{T}\sum_{t=1}^{T}\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\tilde{\boldsymbol{\theta}}_T,\tilde{\boldsymbol{\eta}}_T),$$

where

$$\mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) = [\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}'(\boldsymbol{\phi}_0)]^{-1}$$

and

$$\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0) = \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0) - \mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0) \quad \text{(C25)}$$

is the residual from the unconditional theoretical regression of the score corresponding to $\boldsymbol{\theta}$, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$, on the score corresponding to $\boldsymbol{\eta}$, $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}_0)$. This residual score is sometimes called the unrestricted parametric efficient score of $\boldsymbol{\theta}$, and its covariance matrix, $\mathcal{P}(\boldsymbol{\phi}_0) = [\mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)]^{-1}$, the marginal information matrix of $\boldsymbol{\theta}$, or the unrestricted parametric efficiency bound.

In the spherically symmetric case, we can easily prove that (C25) and its covariance matrix reduce to

$$\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\phi}_0) = \mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\cdot\left[\mathrm{M}_{sr}(\boldsymbol{\eta}_0)\mathcal{M}_{rr}^{-1}(\boldsymbol{\eta}_0)\mathbf{e}_{rt}(\boldsymbol{\phi}_0)\right] \quad \text{(C26)}$$

and

$$\mathcal{P}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}_s'(\boldsymbol{\phi}_0)\cdot\left[\mathrm{M}_{sr}(\boldsymbol{\eta}_0)\mathcal{M}_{rr}^{-1}(\boldsymbol{\eta}_0)\mathrm{M}_{sr}'(\boldsymbol{\eta}_0)\right], \quad \text{(C27)}$$

respectively, where

$$\mathbf{W}_s(\boldsymbol{\phi}_0) = \mathbf{Z}_d(\boldsymbol{\theta}_0)[\mathbf{0}', vec'(\mathbf{I}_N)]' = E[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0][\mathbf{0}', vec'(\mathbf{I}_N)]'$$
$$= E\left\{ \frac{1}{2}\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial \boldsymbol{\theta}} vec[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)] \right| \boldsymbol{\phi}_0 \Bigg\} = E[\mathbf{W}_{st}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0] = -E\left[\frac{\partial d_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right| \boldsymbol{\phi}_0 \Bigg], \tag{C28}$$

It is worth noting that the last summand of (C25) coincides with $\mathbf{Z}_d(\boldsymbol{\phi}_0)$ times the theoretical least squares projection of $\mathbf{e}_{dt}(\boldsymbol{\phi}_0)$ on (the linear span of) $\mathbf{e}_{rt}(\boldsymbol{\phi}_0)$, which is conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})$ from Proposition 3 of Fiorentini and Sentana (2007). Such an interpretation immediately suggests alternative estimators of $\boldsymbol{\theta}$ that replace a parametric assumption on the shape of the distribution of the standardised innovations $\boldsymbol{\varepsilon}_t^*$ by a more flexible alternative. Specifically, Hodgson and Vorkink (2003), Hafner and Rombouts (2007) and other authors have suggested spherically symmetric semiparametric estimators which allow for any member of the class of spherically symmetric distribution. To derive such estimators, these authors replace the linear span of $\mathbf{e}_{rt}(\boldsymbol{\phi}_0)$ by the so-called spherically symmetric tangent set, which is the Hilbert space generated by all time-invariant functions of $\varsigma_t(\boldsymbol{\theta}_0)$ with bounded second moments that have zero conditional means and are conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})$. The next proposition, which originally appeared as Proposition 7 in Fiorentini and Sentana (2007), provides the resulting spherically symmetric semiparametric efficient score and the corresponding efficiency bound:

Proposition C3 *When $\boldsymbol{\varepsilon}_t^*|I_{t-1}, \boldsymbol{\phi}$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta})$ with $-2/(N+2) < \kappa_0 < \infty$, the spherically symmetric semiparametric efficient score is given by:*

$$\mathring{\mathbf{s}}_{\theta t}(\boldsymbol{\phi}_0) = \mathbf{s}_{\theta t}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\left\{\left[\delta[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0]\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N} - 1\right] - \frac{2}{(N+2)\kappa_0 + 2}\left[\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N} - 1\right]\right\}, \tag{C29}$$

while the spherically symmetric semiparametric efficiency bound is

$$\mathring{\mathcal{S}}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}_s'(\boldsymbol{\phi}_0) \cdot \left\{ \left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1\right] - \frac{4}{N\left[(N+2)\kappa_0 + 2\right]} \right\}. \tag{C30}$$

Proof. First of all, it is easy to show that for any spherical distribution

$$\begin{aligned} \mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta}_0, \mathbf{0}) &= E\left[\begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\theta}_0, \mathbf{0}) \\ \mathbf{e}_{st}(\boldsymbol{\theta}_0, \mathbf{0}) \end{array} \right| \varsigma_t; \boldsymbol{\phi}_0 \Bigg] = E\left\{ \begin{array}{c} \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0) \\ vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0) - \mathbf{I}_N\right] \end{array} \right| \varsigma_t; \boldsymbol{\phi}_0 \Bigg\} \\ &= E\left[\begin{array}{c} \sqrt{\varsigma_t}\mathbf{u}_t \\ vec(\varsigma_t \mathbf{u}_t\mathbf{u}_t' - \mathbf{I}_N) \end{array} \right| \varsigma_t \Bigg] = \left(\frac{\varsigma_t}{N} - 1\right)\left[\begin{array}{c} \mathbf{0} \\ vec(\mathbf{I}_N) \end{array} \right], \end{aligned} \tag{C31}$$

and

$$\begin{aligned} \mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}_0) &= E\left[\begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\phi}_0) \\ \mathbf{e}_{st}(\boldsymbol{\phi}_0) \end{array} \right| \varsigma_t; \boldsymbol{\phi}_0 \Bigg] \\ &= E\left\{ \begin{array}{c} \delta[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0] \cdot \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0) \\ vec\left[\delta[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0] \cdot \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0) - \mathbf{I}_N\right] \end{array} \right| \varsigma_t; \boldsymbol{\phi}_0 \Bigg\} \\ &= E\left\{ \begin{array}{c} \delta(\varsigma_t, \boldsymbol{\eta}_0)\sqrt{\varsigma_t}\mathbf{u}_t \\ vec[\delta(\varsigma_t, \boldsymbol{\eta}_0)\varsigma_t \mathbf{u}_t\mathbf{u}_t' - \mathbf{I}_N] \end{array} \right| \varsigma_t \Bigg\} = \left[\delta(\varsigma_t, \boldsymbol{\eta}_0)\frac{\varsigma_t}{N}\text{-}1\right]\left[\begin{array}{c} \mathbf{0} \\ vec(\mathbf{I}_N) \end{array} \right], \end{aligned} \tag{C32}$$

where we have used again the fact that $E(\mathbf{u}_t) = \mathbf{0}$, $E(\mathbf{u}_t\mathbf{u}_t') = N^{-1}\mathbf{I}_N$, and ς_t and $\mathbf{u}_t$ are stochastically independent.

In addition, we can use the law of iterated expectations to show that

$$E[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})\mathbf{e}_{dt}'(\boldsymbol{\phi})|\boldsymbol{\phi}] = E\{E[[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})\mathbf{e}_{dt}'(\boldsymbol{\phi})|\varsigma_t,\boldsymbol{\phi}]|\boldsymbol{\phi}\} = E[\mathbf{e}_{dt}(\boldsymbol{\phi})\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\phi})|\boldsymbol{\phi}] = E[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\phi})|\boldsymbol{\phi}],$$
$$E[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})\mathbf{e}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}] = E\{E[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})\mathbf{e}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\varsigma_t,\boldsymbol{\phi}]|\boldsymbol{\phi}\} = E[\mathbf{e}_{dt}(\boldsymbol{\phi})\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}] = E[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}]$$

and

$$E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\mathbf{e}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\theta},0)|\boldsymbol{\phi}\right] = E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right].$$

Hence, to compute these matrices we simply need three scalar moments.

In this respect, we can use (C24) to show that

$$E\left[\left(\frac{\varsigma_t}{N}-1\right)^2\middle|\boldsymbol{\eta}\right] = \frac{(N+2)\kappa+2}{N}, \tag{C33}$$

so that

$$E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\mathbf{e}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right] = \frac{(N+2)\kappa+2}{N}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & vec(\mathbf{I}_N)vec'(\mathbf{I}_N)\end{pmatrix} = \mathring{\mathcal{K}}(\kappa).$$

We can also use Lemma 2 in Supplemental Appendix B to show that $E(\varsigma_t^2) = N(N+2)(\kappa+1) < \infty$ implies

$$E\left[\varsigma_t^2 \cdot \partial \ln h(\varsigma_t;\boldsymbol{\eta})/\partial\varsigma\middle|\, \boldsymbol{\eta}\right] = -E\left[2\varsigma_t|\boldsymbol{\eta}\right] = -2N.$$

If we then combine this result with (C20) and (C21), we will have that for any spherically symmetric distribution

$$E\left\{\left(\frac{\varsigma_t}{N}-1\right)\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}-1\right]\middle|\,\boldsymbol{\eta}\right\} = \frac{2}{N}, \tag{C34}$$

so that

$$E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})\mathbf{e}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right] = \mathring{\mathcal{K}}(0),$$

which coincides with the value of $E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\mathbf{e}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right]$ under normality.

Finally, Proposition C1 immediately implies that

$$E\left\{\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}-1\right]^2\middle|\,\boldsymbol{\eta}\right\} = \frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta})-1. \tag{C35}$$

Therefore, it trivially follows from the expressions for $\mathring{\mathcal{K}}(0)$ and $\mathring{\mathcal{K}}(\kappa_0)$ above that

$$E\left\{\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}) - \mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\mathbf{e}_{dt}'(\boldsymbol{\theta},\mathbf{0})\middle|\, I_{t-1};\boldsymbol{\phi}\right\}$$
$$= E\left\{\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}) - \mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\theta},\mathbf{0})\middle|\, I_{t-1};\boldsymbol{\phi}\right\} = \mathbf{0}$$

for any spherically symmetric distribution. In addition, we also know that

$$E\left\{\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}) - \mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\middle|\, I_{t-1};\boldsymbol{\phi}\right\} = \mathbf{0}.$$

Thus, even though $\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}_0) - \mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa_0)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})\right]$ is the residual from the theoretical regression of $\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})$ on a constant and $\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})$, it turns out that the second summand of (C29) belongs to the restricted tangent set, which is the Hilbert space spanned by all the time-invariant functions of $\boldsymbol{\varsigma}_t(\boldsymbol{\theta}_0)$ with bounded second moments that have zero conditional means and are conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$.

Now, if write (C29) as

$$\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\phi})\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}) + \mathbf{Z}_d(\boldsymbol{\phi})\mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0}),$$

then we can use the law of iterated expectations to show that the spherically symmetric semiparametric efficient score is indeed unconditionally orthogonal to the restricted tangent set.

Finally, the expression for the semiparametric efficiency bound will be

$$\begin{aligned}
E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathring{\mathbf{s}}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}] &= E\left[\begin{array}{c} \left\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\phi})\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}) - \mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\right\} \\ \times\left\{\mathbf{e}_{dt}(\boldsymbol{\phi})'\mathbf{Z}'_{dt}(\boldsymbol{\theta}) - \left[\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi}) - \mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathcal{K}}(0)\right]\mathbf{Z}'_d(\boldsymbol{\phi})\right\} \end{array}\middle|\,\boldsymbol{\phi}\right] \\
&= E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\phi})\mathbf{Z}_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}\right] \\
&-E\left\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\left[\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi}) - \mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathcal{K}}(0)\right]\mathbf{Z}'_d(\boldsymbol{\phi})|\boldsymbol{\phi}\right\} \\
&-E\left\{\mathbf{Z}_d(\boldsymbol{\phi})\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}) - \mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\mathbf{e}'_{dt}(\boldsymbol{\phi})\mathbf{Z}'_d(\boldsymbol{\phi})|\boldsymbol{\phi}\right\} \\
&+E\left\{\mathbf{Z}_d(\boldsymbol{\phi})\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}) - \mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\left[\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi}) - \mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathcal{K}}(0)\right]\mathbf{Z}'_d(\boldsymbol{\phi})|\boldsymbol{\phi}\right\} \\
&= \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}'_s(\boldsymbol{\phi}_0)\cdot\left\{\left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1\right] - \frac{4}{N\left[(N+2)\kappa+2\right]}\right\}
\end{aligned}$$

by virtue of the law of iterated expectations. □

In the case of the univariate GARCH-M model (19), we estimate the model parameters using reparametrisation 1 in section 4. Specifically,

$$\mathbf{Z}_{lt}(\boldsymbol{\vartheta}) = \frac{\partial\mu_t(\boldsymbol{\vartheta})/\partial\boldsymbol{\vartheta}}{\vartheta_i^{1/2}\sigma_t^{\circ}(\boldsymbol{\vartheta}_c)} = \frac{1}{\vartheta_i^{1/2}\sigma_t^{\circ}(\boldsymbol{\vartheta}_c)}\left[\begin{array}{c} \sigma_t^{\circ}(\boldsymbol{\vartheta}_c)\frac{\partial\delta}{\partial\boldsymbol{\vartheta}_c} + \frac{\delta}{2\sigma_t^{\circ}(\boldsymbol{\vartheta}_c)}\frac{\partial\sigma_t^{\circ 2}(\boldsymbol{\vartheta}_c)}{\partial\boldsymbol{\vartheta}_c} \\ 0 \end{array}\right] = \frac{1}{\vartheta_i^{1/2}}\left[\begin{array}{c} \frac{\partial\delta}{\partial\boldsymbol{\vartheta}_c} + \delta W_{st}(\boldsymbol{\vartheta}_c) \\ 0 \end{array}\right],$$

$$\mathbf{Z}_{st}(\boldsymbol{\vartheta}) = \frac{\partial\sigma_t^2(\boldsymbol{\vartheta})/\partial\boldsymbol{\vartheta}}{2\vartheta_{ic}\sigma_t^{\circ 2}(\boldsymbol{\vartheta}_c)} = \frac{1}{2\vartheta_i\sigma_t^{\circ 2}(\boldsymbol{\vartheta}_c)}\left[\begin{array}{c} \vartheta_i\frac{\partial\sigma_t^{\circ 2}(\boldsymbol{\vartheta}_c)}{\partial\boldsymbol{\vartheta}_c} \\ \sigma_t^{\circ 2}(\boldsymbol{\vartheta}_c) \end{array}\right] = \left[\begin{array}{c} W_{st}(\boldsymbol{\vartheta}_c) \\ \frac{1}{2}\vartheta_i^{-1} \end{array}\right],$$

$$W_{st}(\boldsymbol{\vartheta}_c) = \frac{1}{2\sigma_t^{\circ 2}(\boldsymbol{\vartheta}_c)}\frac{\partial\sigma_t^{\circ 2}(\boldsymbol{\vartheta}_c)}{\partial\boldsymbol{\vartheta}_c}$$

and

$$\varsigma_t(\boldsymbol{\vartheta}) = \varepsilon_t^{*2}(\boldsymbol{\vartheta}) = \vartheta_i^{-1}\sigma_t^{\circ -2}(\boldsymbol{\vartheta}_c)x_t^2.$$

On the other hand, we use the natural parametrisation of the multivariate market model in (20), so that $\boldsymbol{\theta}' = (\mathbf{a}', \mathbf{b}', \boldsymbol{\omega}')$, where $\boldsymbol{\omega} = vech(\boldsymbol{\Omega})$. Given the Jacobian matrices:

$$\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial(\mathbf{a}', \mathbf{b}', \boldsymbol{\omega}')} = \left(\begin{array}{ccc} \mathbf{I}_N & \mathbf{I}_N r_{Mt} & \mathbf{0} \end{array}\right), \tag{C36}$$

$$\frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial(\mathbf{a}', \mathbf{b}', \boldsymbol{\omega}')} = \left(\begin{array}{ccc} \mathbf{0} & \mathbf{0} & \mathbf{D}_N \end{array}\right), \tag{C37}$$

because $\partial vec(\mathbf{\Omega})/\partial vech'(\mathbf{\Omega})$ is the duplication matrix of order N (see Magnus and Neudecker, 1988), a direct application of (C4) immediately implies that

$$\begin{aligned}
\mathbf{s}_{\mathbf{a}t}(\boldsymbol{\theta}) &= \mathbf{\Omega}^{-1}\delta_t\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}),\\
\mathbf{s}_{\mathbf{b}t}(\boldsymbol{\theta}) &= \mathbf{\Omega}^{-1}r_{mt}\delta_t\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}),\\
\mathbf{s}_{\boldsymbol{\omega}t}(\boldsymbol{\theta}) &= \frac{1}{2}\mathbf{D}_N'(\mathbf{\Omega}^{-1}\otimes\mathbf{\Omega}^{-1})vec[\delta_t\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta})-\mathbf{\Omega}],
\end{aligned}$$

where $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \mathbf{r}_t - \mathbf{a} - \mathbf{b}r_{mt}$.

The last ingredient we need is

$$\mathbf{W}_s(\boldsymbol{\phi}_0) = [\mathbf{0}, \mathbf{0}, \frac{1}{2}vec'(\mathbf{\Omega}^{-1})\mathbf{D}_N]'$$

because

$$\mathbf{D}_N'(\mathbf{\Omega}^{-\frac{1}{2}\prime}\otimes\mathbf{\Omega}^{-\frac{1}{2}\prime})vec(\mathbf{I}_N) = \mathbf{D}_N'vec(\mathbf{\Omega}^{-1}).$$

In practice, $\mathbf{e}_{dt}(\boldsymbol{\phi})$ has to be replaced by a semiparametric estimate obtained from the joint density of $\boldsymbol{\varepsilon}_t^*$. However, the spherical symmetry assumption allows us to obtain such an estimate from a nonparametric estimate of the univariate density of ς_t, $h\left(\varsigma_t;\boldsymbol{\eta}\right)$, avoiding in this way the curse of dimensionality. Specifically, if we use expression (C19), then we can estimate $\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ non-parametrically by exploiting that

$$-\frac{2\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\varsigma} = -\frac{2\partial\ln h[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\varsigma} + \frac{N-2}{2}\frac{1}{\varsigma_t(\boldsymbol{\theta})}.$$

We can compute $h[\varsigma_t(\boldsymbol{\theta});\boldsymbol{\eta}]$ either directly by using a kernel for positive random variables (see Chen (2000)), or indirectly by using a faster standard Gaussian kernel after exploiting the Box-Cox-type transformation $v=\varsigma^k$ (see Hodgson, Linton and Vorkink (2002)). In the second case, the usual change of variable formula yields

$$p(v;\boldsymbol{\eta}) = \frac{\pi^{N/2}}{k\Gamma(N/2)}v^{-1+N/2k}\exp[c(\boldsymbol{\eta})+g(v^{1/k};\boldsymbol{\eta})],$$

whence

$$g(v^{1/k};\boldsymbol{\eta}) = \ln p(v;\boldsymbol{\eta}) + \left(1-\frac{N}{2k}\right)\ln v - \frac{N}{2}\ln 2\pi + \ln k - \ln\Gamma(N/2) - c(\boldsymbol{\eta})$$

and

$$\frac{\partial g(v^{1/k};\boldsymbol{\eta})}{\partial v^{1/k}} = k\frac{\partial\ln f(v;\boldsymbol{\eta})}{\partial v}v^{1-1/k} + \frac{k-N/2}{v^{1/k}}.$$

We use the second procedure in our Monte Carlo simulations because the distribution of $\varsigma_t(\boldsymbol{\theta})$ becomes more normal-like as N increases, which reduces the advantages of using kernels for positive variables. Specifically, we use a cubic root transformation to improve the approximation, with a common bandwidth parameter for both the density and its first derivative. Given that a proper cross-validation procedure is extremely costly to implement in a Monte Carlo exercise, we have done some experimentation to choose the optimal bandwidth by scaling up and down the automatic choices given in Silverman (1986).

In the univariate case, there is a conceptually simpler alternative that does not require working with $\varsigma_t = \varepsilon_t^{*2}$. In particular, we can exploit the fact that the density of ε_t^* is the same as the density of $-\varepsilon_t^*$ by assigning to $\pm\varepsilon_t^*$ the equally weighted average of the non-parametric density estimates at ε_t^* and $-\varepsilon_t^*$. Likewise, we can compute the equally weighted average of the absolute value of its derivatives and assign its $\pm$ value to ε_t^* and $-\varepsilon_t^*$, respectively.

D The general case of non-spherical distributions

D.1 Likelihood, score and Hessian for non-spherical distributions

In this section, we assume that, conditional on I_{t-1}, $\boldsymbol{\varepsilon}_t^*$ is independent and identically distributed, or $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0 \sim i.i.d.\ D(\mathbf{0},\mathbf{I}_N,\boldsymbol{\varrho}_0)$ for short, where $\boldsymbol{\varrho}$ are some q additional parameters that determine the shape of the distribution. Importantly, this distribution could substantially depart from a multivariate normal both in terms of skewness and kurtosis. Let $f(\boldsymbol{\varepsilon}^*;\boldsymbol{\varrho})$ denote the assumed conditional density of $\boldsymbol{\varepsilon}_t^*$ given I_{t-1} and those shape parameters $\boldsymbol{\varrho}$, which we assume is well defined. Let also $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\varrho})'$ denote the $p+q$ parameters of interest, which once again we assume variation free. Ignoring initial conditions, the log-likelihood function of a sample of size T for those values of $\boldsymbol{\theta}$ for which $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ has full rank will take the form $L_T(\boldsymbol{\phi}) = \sum_{t=1}^T l_t(\boldsymbol{\phi})$, where $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + \ln f\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}),\boldsymbol{\varrho}\right]$, $d_t(\boldsymbol{\theta}) = \ln|\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})|$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$, and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta})$.

The most common choices of square root matrices are the Cholesky decomposition, which leads to a lower triangular matrix for a given ordering of $\mathbf{y}_t$, or the spectral decomposition, which yields a symmetric matrix. The choice of square root matrix is non-trivial because $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ affects the value of the log-likelihood function and its score in multivariate non-spherical contexts. In what follows, we rely mostly on the Cholesky decomposition because it is much faster to compute than the spectral one, especially when $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ is time-varying. Nevertheless, we also discuss some modifications required for the spectral decomposition later on.

Let $\mathbf{s}_t(\boldsymbol{\phi})$ denote the score function $\partial l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$, and partition it into two blocks, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ and $\mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi})$, whose dimensions conform to those of $\boldsymbol{\theta}$ and $\boldsymbol{\varrho}$, respectively. Assuming that $\boldsymbol{\mu}_t(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ and $\ln f(\boldsymbol{\varepsilon}^*,\boldsymbol{\varrho})$ are differentiable, it trivially follows that

$$\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\varrho}) = \frac{\partial d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} + \frac{\partial\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}.$$

But since

$$\partial d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = -\frac{\partial vec'[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}} vec[\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})] = -\mathbf{Z}_{st}(\boldsymbol{\theta})vec(\mathbf{I}_N)$$

and

$$\begin{aligned}\frac{\partial\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} &= -\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} - [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})]\frac{\partial vec[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'} \\ &= -\{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\},\end{aligned} \quad \text{(D1)}$$

where

$$\left.\begin{array}{l}\mathbf{Z}_{lt}(\boldsymbol{\theta})=\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\cdot\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})\\ \mathbf{Z}_{st}(\boldsymbol{\theta})=\partial vec'[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}\cdot[\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})]\end{array}\right\}, \tag{D2}$$

it follows that

$$\begin{aligned}\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) &= [\mathbf{Z}_{lt}(\boldsymbol{\theta}),\mathbf{Z}_{st}(\boldsymbol{\theta})]\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right]=\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}),\\ \mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi}) &= \partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varrho}=\mathbf{e}_{rt}(\boldsymbol{\phi}),\end{aligned} \tag{D3}$$

with

$$\mathbf{e}_{dt}(\boldsymbol{\phi})=\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right]=\left[\begin{array}{c}-\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*,\\ -vec\left\{\mathbf{I}_N+\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\cdot\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\}\end{array}\right]. \tag{D4}$$

Similarly, let $\mathbf{h}_t(\boldsymbol{\phi})$ denote the Hessian function $\partial\mathbf{s}_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}'=\partial^2 l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'$. Assuming twice differentiability of the different functions involved, expression (D1) implies that

$$\frac{\partial\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial\boldsymbol{\theta}'}=-\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}\frac{\partial\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}=\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}\{\mathbf{Z}_{lt}'(\boldsymbol{\theta})+[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\} \tag{D5}$$

because

$$d\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})=-d\{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\}. \tag{D6}$$

In turn,

$$\begin{aligned}d\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})&=-dvec\left[\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\cdot\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right]\\ &=-[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]d\left\{\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\right\}-\left\{\mathbf{I}_N\otimes\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\right\}d\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\end{aligned} \tag{D7}$$

implies that

$$\begin{aligned}\frac{\partial\mathbf{e}_{st}(\boldsymbol{\phi})}{\partial\boldsymbol{\theta}'}&=\frac{\partial\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial\boldsymbol{\theta}'}=-[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}\frac{\partial\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}-\left\{\mathbf{I}_N\otimes\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\right\}\frac{\partial\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\\ &=\left\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}+\left[\mathbf{I}_N\otimes\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\right]\right\}\{\mathbf{Z}_{lt}'(\boldsymbol{\theta})+[\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\}.\end{aligned} \tag{D8}$$

Finally, (D6) and (D7) trivially imply that

$$\begin{aligned}\frac{\partial^2\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\varrho}'}&=-\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'},\\ \frac{\partial^2\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\varrho}'}&=-[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'}.\end{aligned}$$

Using these results, we can easily obtained the required expressions for

$$\begin{aligned}\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi})&=\mathbf{Z}_{lt}(\boldsymbol{\theta})\frac{\partial\mathbf{e}_{lt}(\boldsymbol{\phi})}{\partial\boldsymbol{\theta}'}+\mathbf{Z}_{st}(\boldsymbol{\theta})\frac{\partial\mathbf{e}_{st}(\boldsymbol{\phi})}{\partial\boldsymbol{\theta}'}\\ &\quad+\left[\mathbf{e}_{lt}'(\boldsymbol{\phi})\otimes\mathbf{I}_p\right]\frac{\partial vec[\mathbf{Z}_{lt}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'}+\left[\mathbf{e}_{st}'(\boldsymbol{\phi})\otimes\mathbf{I}_p\right]\frac{\partial vec[\mathbf{Z}_{st}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'}, \qquad \text{(D9)}\\ \mathbf{h}_{\boldsymbol{\theta\varrho}t}(\boldsymbol{\phi})&=\mathbf{Z}_{lt}(\boldsymbol{\theta})\partial\mathbf{e}_{lt}(\boldsymbol{\phi})/\partial\boldsymbol{\varrho}'+\mathbf{Z}_{st}(\boldsymbol{\theta})\partial\mathbf{e}_{st}(\boldsymbol{\phi})/\partial\boldsymbol{\varrho}', \qquad \text{(D10)}\\ \mathbf{h}_{\boldsymbol{\varrho\varrho}t}(\boldsymbol{\phi})&=\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'.\end{aligned}$$

In this regard, note that since (D6) and (D7) also imply that

$$\begin{aligned}\partial \mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}' &= -\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}', & \text{(D11)}\\ \partial \mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}' &= -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes \mathbf{I}_N]\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}', & \text{(D12)}\end{aligned}$$

respectively, it is clear that

$$\begin{aligned}\mathbf{Z}_{lt}(\boldsymbol{\theta})\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial\boldsymbol{\varrho}'}+\mathbf{Z}_{st}(\boldsymbol{\theta})\frac{\partial \mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial\boldsymbol{\varrho}'} &= -\{\mathbf{Z}_{lt}(\boldsymbol{\theta})+\mathbf{Z}_{st}(\boldsymbol{\theta})[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes \mathbf{I}_N]\}\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'}\\ &= \frac{\partial \boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho})}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'}\end{aligned}$$

so both ways of computing $\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\varrho}t}(\boldsymbol{\phi})$ indeed coincide.

Importantly, while $\mathbf{Z}_{lt}(\boldsymbol{\theta})$, $\mathbf{Z}_{st}(\boldsymbol{\theta})$, $\partial vec[\mathbf{Z}_{lt}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}'$ and $\partial vec[\mathbf{Z}_{st}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}'$ depend on the dynamic model specification, the first and second derivatives of $\ln f(\boldsymbol{\varepsilon}^*;\boldsymbol{\varrho})$ depend on the specific distribution assumed for estimation purposes.

For the standard (i.e. lower triangular) Cholesky decomposition of $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, we will have that

$$dvec(\boldsymbol{\Sigma}_t)=[(\boldsymbol{\Sigma}_t^{1/2}\otimes \mathbf{I}_N)+(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{1/2})\mathbf{K}_{NN}]dvec(\boldsymbol{\Sigma}_t^{1/2}).$$

Unfortunately, this transformation is singular, which means that we must find an analogous transformation between the corresponding *dvech's*. In this sense, we can write the previous expression as

$$dvech(\boldsymbol{\Sigma}_t)=[\mathbf{L}_N(\boldsymbol{\Sigma}_t^{1/2}\otimes \mathbf{I}_N)\mathbf{L}_N'+\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{1/2})\mathbf{K}_{NN}\mathbf{L}_N']dvech(\boldsymbol{\Sigma}_t^{1/2}), \tag{D13}$$

where $\mathbf{L}_N$ is the elimination matrix (see Magnus, 1988). We can then use the results in chapter 5 of Magnus (1988) to show that the above mapping will be lower triangular of full rank as long as $\boldsymbol{\Sigma}_t^{1/2}$ has full rank, which means that we can readily obtain the Jacobian matrix of $vech(\boldsymbol{\Sigma}_t^{1/2})$ from the Jacobian matrix of $vech(\boldsymbol{\Sigma}_t)$.

In the case of the symmetric square root matrix, the analogous transformation would be

$$dvech(\boldsymbol{\Sigma}_t)=[\mathbf{D}_N^+(\boldsymbol{\Sigma}_t^{1/2}\otimes \mathbf{I}_N)\mathbf{D}_N+\mathbf{D}_N^+(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{1/2})\mathbf{D}_N]dvech(\boldsymbol{\Sigma}_t^{1/2}),$$

where $\mathbf{D}_N^+=(\mathbf{D}_N'\mathbf{D}_N)^{-1}\mathbf{D}_N'$ is the Moore-Penrose inverse of the duplication matrix (see Magnus and Neudecker, 1988).

From a numerical point of view, the calculation of both $\mathbf{L}_N(\boldsymbol{\Sigma}_t^{1/2}\otimes \mathbf{I}_N)\mathbf{L}_N'$ and $\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{1/2})\mathbf{K}_{NN}\mathbf{L}_N'$ is straightforward. Specifically, given that $\mathbf{L}_N vec(\mathbf{A})=vech(\mathbf{A})$ for any square matrix $\mathbf{A}$, the effect of premultiplying by the $\frac{1}{2}N(N+1)\times N^2$ matrix $\mathbf{L}_N$ is to eliminate rows N+1, 2N+1 and 2N+2, 3N+1, 3N+2 and 3N+3, etc. Similarly, given that $\mathbf{L}_N\mathbf{K}_{NN}vec(\mathbf{A})=vech(\mathbf{A}')$, the effect of postmultiplying by $\mathbf{K}_{NN}\mathbf{L}_N'$ is to delete all columns but those in positions 1, N+1, 2N+1,...,N+2, 2N+2,..., N+3, 2N+3,..., N^2.

Let $\mathbf{F}_t$ denote the transpose of the inverse of $\mathbf{L}_N(\boldsymbol{\Sigma}_t^{1/2}\otimes \mathbf{I}_N)\mathbf{L}_N'+\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{1/2})\mathbf{K}_{NN}\mathbf{L}_N'$, which will be upper triangular. The fastest way to compute

$$\frac{\partial vec'[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}}[\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})]=\frac{1}{2}\frac{\partial vech'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}}\mathbf{F}_t\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{-1/2}) \tag{D14}$$

is as follows:

1. From the expression for $\partial vec'\left[\mathbf{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}$ we can readily obtain $\partial vech'\left[\mathbf{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}$ by simply avoiding the computation of the duplicated columns

2. Then we postmultiply the resulting matrix by $\mathbf{F}_t$

3. Next, we construct the matrix

$$\mathbf{L}_N(\mathbf{I}_N\otimes\mathbf{\Sigma}_t^{1/2})=\mathbf{L}_N\begin{pmatrix}\mathbf{\Sigma}_t^{-1/2} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma}_t^{-1/2} & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Sigma}_t^{-1/2}\end{pmatrix}$$

   by eliminating the first row from the second block, the first two rows from the third block, ..., and all the rows but the last one from the last block

4. Finally, we premultiply the resulting matrix by $\partial vech'\left[\mathbf{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}\cdot\mathbf{F}_t$.

D.2 Asymptotic distribution

Propositions 10.1, 13, C2.1 and D3 already deal explicitly with the general case, so there is no need to generalise them. In turn, Propositions 6, 7, 8, 9 and their proofs continue to be valid if we change $\boldsymbol{\eta}$ by $\boldsymbol{\varrho}$. The same happens to Proposition 5, provided we erase the row and columns corresponding to $\mathring{\boldsymbol{\theta}}_T$ and its influence function $\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$. On the other hand, Propositions 10.2, 11, 12, C2.2 and C3 are specific to the spherically symmetric case. Therefore, the only proposition that really requires a proper generalisation is Proposition C1.

Proposition D1 *If $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}$ is i.i.d. $D(\mathbf{0},\mathbf{I}_N,\boldsymbol{\varrho})$ with density $f(\boldsymbol{\varepsilon}^*,\boldsymbol{\varrho})$, then*

$$\begin{aligned}\mathcal{I}_t(\boldsymbol{\phi}) &= \mathbf{Z}_t(\boldsymbol{\theta})\mathcal{M}(\boldsymbol{\varrho})\mathbf{Z}_t'(\boldsymbol{\theta}),\\ \mathbf{Z}_t(\boldsymbol{\theta}) &= \begin{pmatrix}\mathbf{Z}_{dt}(\boldsymbol{\theta}) & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_q\end{pmatrix}=\begin{pmatrix}\mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q\end{pmatrix},\end{aligned}$$

and

$$\mathcal{M}(\boldsymbol{\varrho})=\begin{bmatrix}\mathcal{M}_{dd}(\boldsymbol{\varrho}) & \mathcal{M}_{dr}(\boldsymbol{\varrho})\\ \mathcal{M}_{dr}'(\boldsymbol{\varrho}) & \mathcal{M}_{rr}(\boldsymbol{\varrho})\end{bmatrix}=\begin{bmatrix}\mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{lr}(\boldsymbol{\varrho})\\ \mathcal{M}_{ls}'(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho}) & \mathcal{M}_{sr}(\boldsymbol{\varrho})\\ \mathcal{M}_{lr}'(\boldsymbol{\varrho}) & \mathcal{M}_{sr}'(\boldsymbol{\varrho}) & \mathcal{M}_{rr}(\boldsymbol{\varrho})\end{bmatrix},$$

with

$$\begin{aligned}&\mathcal{M}_{ll}(\boldsymbol{\varrho})=V[\mathbf{e}_{lt}(\boldsymbol{\phi})|\boldsymbol{\phi}]=E\left[\partial^2\ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\right|\boldsymbol{\varrho}\right],\\ &\mathcal{M}_{ls}(\boldsymbol{\varrho})=E[\mathbf{e}_{lt}(\boldsymbol{\phi})\mathbf{e}_{st}(\boldsymbol{\phi})'|\boldsymbol{\phi}]=E\left[\partial^2\ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\cdot(\boldsymbol{\varepsilon}_t^{\prime*}\otimes\mathbf{I}_N)\right|\boldsymbol{\varrho}\right],\\ &\mathcal{M}_{ss}(\boldsymbol{\varrho})=V[\mathbf{e}_{st}(\boldsymbol{\phi})|\boldsymbol{\phi}]=E\left[(\boldsymbol{\varepsilon}_t^*\otimes\mathbf{I}_N)\cdot\partial^2\ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\cdot(\boldsymbol{\varepsilon}_t^{*\prime}\otimes\mathbf{I}_N)|\boldsymbol{\varrho}\right]-\mathbf{K}_{NN},\\ &\mathcal{M}_{lr}(\boldsymbol{\varrho})=E[\mathbf{e}_{lt}(\boldsymbol{\phi})\mathbf{e}_{rt}'(\boldsymbol{\phi})|\boldsymbol{\phi}]=-E\left[\partial^2\ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\varrho}\right],\\ &\mathcal{M}_{sr}(\boldsymbol{\varrho})=E[\mathbf{e}_{st}(\boldsymbol{\phi})\mathbf{e}_{rt}'(\boldsymbol{\phi})|\boldsymbol{\phi}]=-E\left[(\boldsymbol{\varepsilon}_t^*\otimes\mathbf{I}_N)\partial^2\ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\varrho}\right],\end{aligned}$$

and

$$\mathcal{M}_{rr}(\boldsymbol{\varrho})=V[\mathbf{e}_{rt}(\boldsymbol{\phi})|\boldsymbol{\phi}]=-E\left[\partial^2\ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'|\boldsymbol{\phi}\right].$$

Proof. Since the distribution of $\boldsymbol{\varepsilon}_t^*$ given I_{t-1} is assumed to be *i.i.d.*, then it is easy to see from (D3) that $\mathbf{e}_t(\boldsymbol{\phi}) = [\mathbf{e}'_{dt}(\boldsymbol{\phi}), \mathbf{e}'_{rt}(\boldsymbol{\phi})]'$ will inherit the martingale difference property of the score $\mathbf{s}_t(\boldsymbol{\phi}_0)$. As a result, the conditional information matrix will be given by

$$\begin{bmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{bmatrix} \begin{bmatrix} \mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{lr}(\boldsymbol{\varrho}) \\ \mathcal{M}'_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho}) & \mathcal{M}_{sr}(\boldsymbol{\varrho}) \\ \mathcal{M}'_{lr}(\boldsymbol{\varrho}) & \mathcal{M}'_{sr}(\boldsymbol{\varrho}) & \mathcal{M}_{rr}(\boldsymbol{\varrho}) \end{bmatrix} \begin{bmatrix} \mathbf{Z}'_{lt}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{Z}'_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta})\mathcal{M}_{ll}(\boldsymbol{\varrho})\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + \mathbf{Z}_{st}(\boldsymbol{\theta})\mathcal{M}'_{ls}(\boldsymbol{\varrho})\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + \mathbf{Z}_{lt}(\boldsymbol{\theta})\mathcal{M}_{ls}(\boldsymbol{\varrho})\mathbf{Z}'_{st}(\boldsymbol{\theta}) + \mathbf{Z}_{st}(\boldsymbol{\theta})\mathcal{M}_{ss}(\boldsymbol{\varrho})\mathbf{Z}'_{st}(\boldsymbol{\theta}) & \mathbf{Z}_{lt}(\boldsymbol{\theta})\mathcal{M}_{lr}(\boldsymbol{\varrho}) + \mathbf{Z}_{st}(\boldsymbol{\theta})\mathcal{M}_{sr}(\boldsymbol{\varrho}) \\ \mathcal{M}'_{lr}(\boldsymbol{\varrho})\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + \mathcal{M}'_{sr}(\boldsymbol{\varrho})\mathbf{Z}'_{st}(\boldsymbol{\theta}) & \mathcal{M}_{rr}(\boldsymbol{\varrho}) \end{bmatrix},$$

where

$$\begin{bmatrix} \mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{lr}(\boldsymbol{\varrho}) \\ \mathcal{M}'_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho}) & \mathcal{M}_{sr}(\boldsymbol{\varrho}) \\ \mathcal{M}'_{lr}(\boldsymbol{\varrho}) & \mathcal{M}'_{sr}(\boldsymbol{\varrho}) & \mathcal{M}_{rr}(\boldsymbol{\varrho}) \end{bmatrix} = V \left[\left. \begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho}) \\ \mathbf{e}_{st}(\boldsymbol{\theta}, \boldsymbol{\varrho}) \\ \mathbf{e}_{rt}(\boldsymbol{\theta}, \boldsymbol{\varrho}) \end{array} \right| \boldsymbol{\theta}, \boldsymbol{\varrho} \right],$$

which confirms the variance of the score part of the proposition.

As for the expected value of the Hessian expressions, it is easy to see that

$$E[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi})|z_t, I_{t-1}; \boldsymbol{\phi}] = \mathbf{Z}_{lt}(\boldsymbol{\theta}) E\left[\left. \frac{\partial \mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} \right| z_t, I_{t-1}; \boldsymbol{\phi} \right] + \mathbf{Z}_{st}(\boldsymbol{\theta}) E\left[\left. \frac{\partial \mathbf{e}_{st}(\boldsymbol{\theta}, \boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} \right| z_t, I_{t-1}; \boldsymbol{\phi} \right]$$

because

$$E\left[\mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho})|z_t, I_{t-1}; \boldsymbol{\phi}\right] = -E\left[\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial \boldsymbol{\varepsilon}^* | z_t, I_{t-1}; \boldsymbol{\phi}\right] = \mathbf{0} \tag{D15}$$

and

$$E\left[\mathbf{e}_{st}(\boldsymbol{\theta}, \boldsymbol{\varrho})|z_t, I_{t-1}; \boldsymbol{\phi}\right] = -E\left[vec\{\mathbf{I}_N + \partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial \boldsymbol{\varepsilon}^* \cdot \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\} | z_t, I_{t-1}; \boldsymbol{\phi}\right] = \mathbf{0}. \tag{D16}$$

Expression (D5) then leads to

$$E\left[\left. \frac{\partial \mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} \right| z_t, I_{t-1}; \boldsymbol{\phi} \right] = E\left[\left. \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \{\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}'_{st}(\boldsymbol{\theta})\} \right| z_t, I_{t-1}; \boldsymbol{\phi} \right]$$

$$= E\left[\left. \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \right| \boldsymbol{\phi} \right] \mathbf{Z}'_{lt}(\boldsymbol{\theta}) + E\left[\left. \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} [\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \right| \boldsymbol{\phi} \right] \mathbf{Z}'_{st}(\boldsymbol{\theta}).$$

Likewise, equation (D8) leads to

$$E\left[\left. \frac{\partial \mathbf{e}_{st}(\boldsymbol{\theta}, \boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} \right| z_t, I_{t-1}; \boldsymbol{\phi} \right] = E\left[\left\{ [\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} + \left[\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \right] \right\} \right.$$

$$\left. \times \{\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}'_{st}(\boldsymbol{\theta})\} \right| z_t, I_{t-1}; \boldsymbol{\phi} \Big] = E\left[[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \left. \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \right| \boldsymbol{\phi} \right] \mathbf{Z}'_{lt}(\boldsymbol{\theta})$$

$$+ E\left[[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} [\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \right| z_t, I_{t-1}; \boldsymbol{\phi} \Big] \mathbf{Z}'_{st}(\boldsymbol{\theta}) - \mathbf{K}_{NN}\mathbf{Z}'_{st}(\boldsymbol{\theta})$$

because of (D15) and (D16), which in turn implies

$$\begin{aligned}
& E\left\{\left[\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right][\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\right| z_t, I_{t-1};\boldsymbol{\phi}\right\} \\
= \; & \mathbf{K}_{NN} E\left\{\mathbf{K}_{NN}\left[\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right][\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\right| z_t, I_{t-1};\boldsymbol{\phi}\right\} \\
= \; & \mathbf{K}_{NN} E\left\{\left[\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \otimes \mathbf{I}_N\right][\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\right| z_t, I_{t-1};\boldsymbol{\phi}\right\} \\
= \; & \mathbf{K}_{NN} E\left\{\left[\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N\right]\right| z_t, I_{t-1};\boldsymbol{\phi}\right\} = -\mathbf{K}_{NN}
\end{aligned}$$

in view of Theorem 3.1 in Magnus (1988).

As a result, the information matrix equality implies that

$$\begin{aligned}
\mathcal{M}_{ll}(\boldsymbol{\varrho}) &= E\left\{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\right|\boldsymbol{\phi}\right\} \\
\mathcal{M}_{ls}(\boldsymbol{\varrho}) &= E\left\{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\cdot[\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\right|\boldsymbol{\phi}\right\} \\
\mathcal{M}_{ss}(\boldsymbol{\varrho}) &= E\left\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\cdot[\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\right|\boldsymbol{\phi}\right\} - \mathbf{K}_{NN}
\end{aligned}$$

Similarly, equation (D10) implies that

$$E[\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\varrho}t}(\boldsymbol{\phi})|z_t, I_{t-1};\boldsymbol{\phi}] = E[\mathbf{Z}_{lt}(\boldsymbol{\theta})\partial\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}' + \mathbf{Z}_{st}(\boldsymbol{\theta})\partial\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}'|z_t, I_{t-1};\boldsymbol{\phi}].$$

But then the information matrix equality together with equations (D11) and (D12) imply that

$$\begin{aligned}
E[\partial\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}'|z_t, I_{t-1};\boldsymbol{\phi}] &= -E\{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\phi}\} = \mathcal{M}_{lr}(\boldsymbol{\varrho}), \\
E[\partial\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}'|z_t, I_{t-1};\boldsymbol{\phi}] &= -E\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\phi}\} = \mathcal{M}_{sr}(\boldsymbol{\varrho}).
\end{aligned}$$

Finally, the information matrix equality also implies that

$$\mathcal{M}_{rr}(\boldsymbol{\varrho}) = -E\{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'|\boldsymbol{\phi}\},$$

as required. □

D.3 Cross-sectionally independent disturbances

Let us now specialise the results in the previous two subsections for the case in which the disturbances are cross-sectionally independent. Specifically, we assume that the conditional density of $\boldsymbol{\varepsilon}_t^*$ given I_{t-1} and the shape parameters $\boldsymbol{\varrho}$ can be factorised as

$$\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}),\boldsymbol{\varrho}] = \sum_{i=1}^{N} \ln f[\varepsilon_{it}^*(\boldsymbol{\theta}),\boldsymbol{\varrho}_i],$$

where $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = [\varepsilon_{1t}^*(\boldsymbol{\theta}),\ldots,\varepsilon_{Nt}^*(\boldsymbol{\theta})]'$ and $\boldsymbol{\varrho} = (\boldsymbol{\varrho}_1,\ldots,\boldsymbol{\varrho}_N)$, with $\dim(\boldsymbol{\varrho}_i) = q_i$ and $\sum_{i=1}^{N} q_i = q$.

The main simplification in the expressions for the scores result from the fact that

$$\mathbf{e}_{lt}(\boldsymbol{\phi}) = \left\{\begin{array}{c} -\frac{\partial f[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial \varepsilon_1^*} \\ \vdots \\ -\frac{\partial f[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial \varepsilon_N^*} \end{array}\right\},$$

$$\mathbf{e}_{st}(\boldsymbol{\phi}) = -vec\left\{\begin{array}{ccc} 1+\frac{\partial \ln f[\varepsilon^*_{1t}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial \varepsilon^*_1}\varepsilon^*_{1t}(\boldsymbol{\theta}) & \dots & \frac{\partial \ln f[\varepsilon^*_{1t}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial \varepsilon^*_1}\varepsilon^*_{Nt}(\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \frac{\partial \ln f[\varepsilon^*_{Nt}(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial \varepsilon^*_N}\varepsilon^*_{1t}(\boldsymbol{\theta}) & \dots & 1+\frac{\partial \ln f[\varepsilon^*_{Nt}(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial \varepsilon^*_N}\varepsilon^*_{Nt}(\boldsymbol{\theta}) \end{array}\right\}$$

and

$$\mathbf{e}_{rt}(\boldsymbol{\phi}) = \left\{\begin{array}{c} \frac{\partial \ln f[\varepsilon^*_{1t}(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial \boldsymbol{\varrho}_1} \\ \vdots \\ \frac{\partial \ln f[\varepsilon^*_{Nt}(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial \boldsymbol{\varrho}_N} \end{array}\right\},$$

so that the derivatives involved correspond to the underlying univariate densities.

When any of the N distributions is symmetric, then these expressions simplify further as

$$-\frac{\partial f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i} = \delta(\varepsilon^{*2}_{it};\boldsymbol{\varrho}_i)\varepsilon^*_{it}.$$

Additional simplifications in the expressions for the Hessian arise because $\partial^2 \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}$, $\partial^2 \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'$ and $\partial^2 \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'$ are (block) diagonal matrices with representative elements $\partial^2 \ln f[\varepsilon^*_{it}(\boldsymbol{\theta});\boldsymbol{\varrho}_i]/\partial\varepsilon^*_i\partial\varepsilon^*_i$, $\partial^2 \ln f[\varepsilon^*_{it}(\boldsymbol{\theta});\boldsymbol{\varrho}_i]/\partial\varepsilon^*_i\partial\boldsymbol{\varrho}'_i$ and $\partial^2 \ln f[\varepsilon^*_{it}(\boldsymbol{\theta});\boldsymbol{\varrho}_i]/\partial\boldsymbol{\varrho}_i\partial\boldsymbol{\varrho}'_i$, respectively.

As for the information matrix, Proposition D1 simplifies to

Proposition D2 *If $\boldsymbol{\varepsilon}^*_t|I_{t-1};\boldsymbol{\phi}$ is i.i.d. $D(\mathbf{0},\mathbf{I}_N,\boldsymbol{\varrho})$ with density $f(\boldsymbol{\varepsilon}^*,\boldsymbol{\varrho}) = \prod_{i=1}^N f(\varepsilon^*_{it},\boldsymbol{\varrho}_i)$, then the information matrix will be given by a special case of Proposition D1 in which $\mathcal{M}_{ll}$ will be a diagonal matrix of order N with typical element*

$$\mathrm{M}_{ll}(\boldsymbol{\varrho}_i) = V\left[\left.\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\right|\boldsymbol{\varrho}\right],$$

$\mathcal{M}_{ls} = \mathrm{M}_{ls}\mathbf{E}'_N$, where M_{ls} also a diagonal matrix of order N with typical element

$$\mathrm{M}_{ls}(\boldsymbol{\varrho}_i) = cov\left[\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}, \left.\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\varepsilon^*_{it}\right|\boldsymbol{\varrho}\right],$$

$\mathcal{M}_{ss}$ is the sum of the commutation matrix $\mathbf{K}_{NN}$ and a block diagonal matrix $\boldsymbol{\Upsilon}$ of order N^2 in which each of the N diagonal blocks is a diagonal matrix of size N with the following structure:

$$\boldsymbol{\Upsilon}_i = \begin{bmatrix} \mathrm{M}_{ll}(\boldsymbol{\varrho}_1) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathrm{M}_{ll}(\boldsymbol{\varrho}_{i-1}) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathrm{M}_{ss}(\boldsymbol{\varrho}_i)-1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathrm{M}_{ll}(\boldsymbol{\varrho}_{i+1}) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathrm{M}_{ll}(\boldsymbol{\varrho}_N) \end{bmatrix},$$

where

$$\mathrm{M}_{ss}(\boldsymbol{\varrho}_i) = V\left[\left.\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\varepsilon^*_{it}\right|\boldsymbol{\varrho}\right],$$

$\mathcal{M}_{lr}$ is an $N \times q$ block diagonal matrix with typical diagonal block of size $1 \times q_i$

$$\mathrm{M}_{lr}(\boldsymbol{\varrho}_i) = -cov\left[\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}, \left.\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \boldsymbol{\varrho}_i}\right|\boldsymbol{\varrho}\right],$$

$\mathcal{M}_{sr} = \mathbf{E}_N \mathrm{M}_{sr}$, where$\mathrm{M}_{sr}$ another block diagonal matrix of order $N \times q$ with typical block of size $1 \times q_i$

$$\mathrm{M}_{sr}(\boldsymbol{\varrho}_i) = cov\left[\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\varepsilon^*_{it}, \left.\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \boldsymbol{\varrho}_i}\right|\boldsymbol{\varrho}\right],$$

and $\mathcal{M}_{rr}$ is an $q \times q$ block diagonal matrix with typical block of size $q_i \times q_i$

$$\mathrm{M}_{rr}(\boldsymbol{\varrho}_i) = V\left[\left.\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \boldsymbol{\varrho}_i}\right|\boldsymbol{\varrho}_i\right].$$

Proof. The expression for $\mathcal{M}_{ll}$ follows trivially from the fact that

$$cov\left[\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}, \left.\frac{\partial \ln f(\varepsilon^*_{jt};\boldsymbol{\varrho}_j)}{\partial \varepsilon^*_j}\right|\boldsymbol{\varrho}\right] = 0$$

for $i \neq j$ because of the cross-sectional independence of the shocks.

The same property also implies that $\mathcal{M}_{ls} = \mathrm{M}_{ls}\mathbf{E}'_N$ because for $i \neq j \neq k$

$$\begin{aligned}
E\left[\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\frac{\partial \ln f(\varepsilon^*_{jt};\boldsymbol{\varrho}_j)}{\partial \varepsilon^*_j}\varepsilon^*_{it}\right|\left.\boldsymbol{\varrho}\right] &= 0 \text{ since } E\left[\left.\frac{\partial \ln f(\varepsilon^*_{jt};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_j}\right|\boldsymbol{\varrho}\right] = 0,\\
E\left[\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\varepsilon^*_{jt}\right|\left.\boldsymbol{\varrho}\right] &= 0 \text{ since } E\left(\varepsilon^*_{jt}\middle|\boldsymbol{\varrho}\right) = 0,\\
E\left[\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\left(\frac{\partial \ln f(\varepsilon^*_{jt};\boldsymbol{\varrho}_j)}{\partial \varepsilon^*_j}\varepsilon^*_{jt}+1\right)\right|\left.\boldsymbol{\varrho}\right] &= 0 \text{ since } E\left[\left.\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\right|\boldsymbol{\varrho}\right] = 0
\end{aligned}$$

and

$$E\left[\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\frac{\partial \ln f(\varepsilon^*_{jt};\boldsymbol{\varrho}_j)}{\partial \varepsilon^*_j}\varepsilon^*_{kt}\right|\left.\boldsymbol{\varrho}\right] = 0 \text{ since } E\left(\varepsilon^*_{kt}\middle|\boldsymbol{\varrho}\right) = 0.$$

The expression for $\mathcal{M}_{ss}$ is slightly more involved. First, most but not all the off-diagonal terms will be 0. Specifically, when $i \neq j$

$$E\left[\left(\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\varepsilon^*_{it}+1\right)\frac{\partial \ln f(\varepsilon^*_{jt};\boldsymbol{\varrho}_j)}{\partial \varepsilon^*_j}\varepsilon^*_{it}\right|\left.\boldsymbol{\varrho}\right] = 0 \text{ since } E\left[\left.\frac{\partial \ln f(\varepsilon^*_{jt};\boldsymbol{\varrho}_j)}{\partial \varepsilon^*_j}\right|\boldsymbol{\varrho}\right] = 0,$$

$$E\left[\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\varepsilon^*_{it}\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\varepsilon^*_{jt}\right|\left.\boldsymbol{\varrho}\right] = 0 \text{ since } E\left(\varepsilon^*_{jt}\middle|\boldsymbol{\varrho}\right) = 0$$

and

$$E\left[\left(\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\varepsilon^*_{it}+1\right)\left(\frac{\partial \ln f(\varepsilon^*_{jt};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_j}\varepsilon^*_{jt}+1\right)\right|\left.\boldsymbol{\varrho}\right] = 0 \text{ since } E\left[\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\varepsilon^*_{it}+1\middle|\boldsymbol{\varrho}\right] = 0$$

However,

$$E\left[\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\varepsilon^*_{jt}\frac{\partial \ln f(\varepsilon^*_{jt};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_j}\varepsilon^*_{it}\right|\left.\boldsymbol{\varrho}\right] = 1 \text{ since } E\left[\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\varepsilon^*_{it}+1\middle|\boldsymbol{\varrho}\right] = 0.$$

In contrast, the diagonal terms, which can only take two forms, are different from 0. Specifically, they will be either

$$E\left[\left(\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\varepsilon^*_{it}+1\right)^2\middle|\boldsymbol{\varrho}\right] = \mathrm{M}_{ss}(\boldsymbol{\varrho}_i) \text{ since } E\left[\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\varepsilon^*_{it}+1\middle|\boldsymbol{\varrho}\right] = 0$$

or

$$E\left[\left(\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\varepsilon_{jt}^*\right)^2\middle|\boldsymbol{\varrho}\right] = E\left[\left(\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\right)^2\middle|\boldsymbol{\varrho}\right] = \mathrm{M}_{ll}(\boldsymbol{\varrho}_i) \text{ since } E\left(\varepsilon_{jt}^{*2}\middle|\boldsymbol{\varrho}\right) = 1.$$

As a result, we can write $\mathcal{M}_{ss} = \mathbf{K}_{NN} + \boldsymbol{\Upsilon}$.

The cross-sectional independence of the shocks also implies the block diagonal structure of $\mathcal{M}_{lr}$ and $\mathcal{M}_{rr}$, as well as the fact that $\mathcal{M}_{sr} = \mathbf{E}_N \mathrm{M}_{sr}$. As expected, the same expressions are obtained by taking the expected value of the (minus) Hessian. □

When one of the univariate distributions is symmetric, then $\mathrm{M}_{ls}(\boldsymbol{\varrho}_i) = \mathrm{M}_{lr}(\boldsymbol{\varrho}_i) = 0$. One popular example will be the univariate standardised Student t distribution with $\nu = \eta^{-1}$ degrees of freedom, which is such that

$$\ln f[\varepsilon_{it}^*(\boldsymbol{\theta});\eta_i] = c(\eta_i) - \left(\frac{\eta_i+1}{2\eta_i}\right)\log\left[1+\frac{\eta_i}{1-2\eta_i}\varepsilon_{it}^{*2}(\boldsymbol{\theta})\right],$$

with

$$c(\eta_i) = \log\left(\frac{\eta_i+1}{2\eta_i}\right) - \log\left[\Gamma\left(\frac{1}{2\eta_i}\right)\right] - \frac{1}{2}\log\left(\frac{1-2\eta_i}{\eta_i}\right) - \frac{1}{2}\log\pi.$$

Here,

$$\delta(\varepsilon_t^{*2};\eta) = \frac{\eta+1}{1-2\eta+\eta\varepsilon_t^{*2}}$$

and

$$\begin{aligned}\frac{\partial \ln f(\varepsilon_{it}^*;\eta)}{\partial \eta} &= \frac{1}{2\eta(1-2\eta)} - \frac{1}{2\eta^2}\left[\psi\left(\frac{\eta+1}{2\eta}\right) - \psi\left(\frac{1}{2\eta}\right)\right] \\ &\quad - \frac{\eta+1}{1-2\eta+\eta\varepsilon_{it}^{*2}}\frac{\varepsilon_{it}^{*2}}{2\eta(1-2\eta)} + \frac{1}{2\eta^2}\ln\left(1+\frac{\eta}{1-2\eta}\varepsilon_{it}^{*2}\right).\end{aligned}$$

In addition

$$\begin{aligned}\mathrm{M}_{ll}(\boldsymbol{\varrho}_i) &= \frac{\nu_i(\nu_i+1)}{(\nu_i-2)(\nu_i+3)}, \\ \mathrm{M}_{ss}(\boldsymbol{\varrho}_i) &= \frac{2\nu_i}{\nu_i+3}, \\ \mathrm{M}_{sr}(\boldsymbol{\varrho}_i) &= -\frac{6\nu_i^2}{(\nu_i-2)(\nu_i+1)(\nu_i+3)}\end{aligned}$$

and

$$\mathrm{M}_{rr}(\boldsymbol{\varrho}_i) = \frac{\nu_i^4}{4}\left[\psi'\left(\frac{\nu_i}{2}\right) - \psi'\left(\frac{\nu_i+1}{2}\right)\right] - \frac{\nu_i^4(\nu_i-3)(\nu_i+4)}{2(\nu_i-2)^2(\nu_i+1)(\nu_i+3)},$$

where $\psi'(x) = \partial^2 \ln\Gamma(x)/\partial x^2$ is the so-called tri-gamma function (Abramowitz and Stegun 1964), which reduce to 1, 1, 0 and 3/2 respectively, under normality (see Fiorentini, Sentana and Calzolari (2003)). As a result, when all shocks are in fact Gaussian, $\mathcal{M}_{ss} = \mathbf{K}_{NN} + \mathbf{I}_{N^2}$, which confirms that not all elements of $\mathbf{C}$ can be identified with a Gaussian log-likelihood function because rank$(\mathbf{K}_{NN} + \mathbf{I}_{N^2}) = N(N+1)/2$ (see section 4 in Magnus and Sentana (2020) for a general expression for the eigenvalues of $(\mathbf{K}_{NN} + \boldsymbol{\Upsilon})$.

D.4 Semiparametric estimators

In Supplemental Appendix C.5 we interpreted the last summand of (C25) as $\mathbf{Z}_d(\boldsymbol{\phi}_0)$ times the theoretical least squares projection of $\mathbf{e}_{dt}(\boldsymbol{\phi}_0)$ on (the linear span of) $\mathbf{e}_{rt}(\boldsymbol{\phi}_0)$, which is conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$ from Proposition 3 in Fiorentini and Sentana (2007). Such an interpretation allowed Gonzalez-Rivera and Drost (1999) to replace a parametric assumption on the shape of the distribution of the standardised innovations $\boldsymbol{\varepsilon}_t^*$ by a fully non-parametric alternative. Specifically, in a univariate context they replaced the linear span of $\mathbf{e}_{rt}(\boldsymbol{\phi}_0)$ by the so-called unrestricted tangent set, which is the Hilbert space generated by all the time-invariant functions of $\boldsymbol{\varepsilon}_t^*$ with bounded second moments that have zero conditional means and are conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$. The next proposition, which originally appeared as Proposition 6 in Fiorentini and Sentana (2007), describes the resulting semiparametric efficient score and the corresponding efficiency bound for multivariate conditionally heteroskedastic models whose conditionally mean is not identically zero:

Proposition D3 *If $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\theta},\boldsymbol{\rho}$ is i.i.d. $D(\mathbf{0},\mathbf{I}_N,\boldsymbol{\rho})$ with density function $f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\rho})$, where $\boldsymbol{\rho}$ denotes the possibly infinite dimensional vector of shape parameters and $\boldsymbol{\rho}=\mathbf{0}$ normality, and both its Fisher information matrix for location and scale,*

$$\mathcal{M}_{dd}(\boldsymbol{\theta},\boldsymbol{\rho}) = V\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})|I_{t-1};\boldsymbol{\theta},\boldsymbol{\rho}\right]$$
$$= V\left\{\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\rho})\\ \mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\rho})\end{array}\right]\middle|\boldsymbol{\theta},\boldsymbol{\rho}\right\} = V\left\{\left[\begin{array}{c}-\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\rho}]/\partial\boldsymbol{\varepsilon}^*\\ -vec\left\{\mathbf{I}_N+\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\rho}]/\partial\boldsymbol{\varepsilon}^*\cdot\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\}\end{array}\right]\middle|\boldsymbol{\theta},\boldsymbol{\rho}\right\}$$

and the matrix of third and fourth order central moments $\mathcal{K}(\boldsymbol{\rho})$ in (C22) are bounded, then the semiparametric efficient score will be given by:

$$\ddot{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) = \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\rho})\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\rho}) - \mathcal{K}(0)\,\mathcal{K}^+(\boldsymbol{\rho})\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\right], \tag{D17}$$

while the semiparametric efficiency bound is

$$\ddot{\mathcal{S}}(\boldsymbol{\phi}) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta},\boldsymbol{\rho}) - \mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\rho})\left[\mathcal{M}_{dd}(\boldsymbol{\theta},\boldsymbol{\rho}) - \mathcal{K}(0)\,\mathcal{K}^+(\boldsymbol{\rho})\mathcal{K}(0)\right]\mathbf{Z}_d'(\boldsymbol{\theta},\boldsymbol{\rho}), \tag{D18}$$

where $+$ denotes Moore-Penrose inverses and $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta},\boldsymbol{\rho}) = E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{M}_{dd}(\boldsymbol{\theta},\boldsymbol{\rho})\mathbf{Z}_{dt}'(\boldsymbol{\theta})|\boldsymbol{\theta},\boldsymbol{\rho}\right]$.

Proof. It trivially follows from expressions (B3) and (C22) in appendices B and C, respectively, that

$$E\left\{\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathcal{K}(0)\,\mathcal{K}^+(\boldsymbol{\varrho})\,\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\mathbf{e}_{dt}'(\boldsymbol{\theta},\mathbf{0})\,|I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}\right\} = \mathbf{0}$$

for any distribution. In addition, we also know that

$$E\left\{\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathcal{K}(0)\,\mathcal{K}^+(\boldsymbol{\varrho})\,\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]|I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}\right\} = \mathbf{0}.$$

Hence, the second summand of (D17), which can be interpreted as $\mathbf{Z}_d(\boldsymbol{\phi}_0)$ times the residual from the theoretical regression of $\mathbf{e}_{dt}(\boldsymbol{\phi}_0)$ on a constant and $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$, belongs to the unrestricted tangent set, which is the Hilbert space spanned by all the time-invariant functions of $\boldsymbol{\varepsilon}_t^*$ with zero conditional means and bounded second moments that are conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$.

Now, if we write (D17) as

$$[\mathbf{Z}_{dt}(\boldsymbol{\theta})-\mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})]\,\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})+\mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})\mathcal{K}\left(0\right)\mathcal{K}^{+}\left(\boldsymbol{\varrho}\right)\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0}),$$

then we can use the law of iterated expectations to show that the semiparametric efficient score (D17) evaluated at the true parameter values will be unconditionally orthogonal to the unrestricted tangent set because so is $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$, and $E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})-\mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})|\boldsymbol{\theta},\boldsymbol{\varrho}\right]=\mathbf{0}$.

Finally, the expression for the semiparametric efficiency bound will be

$$E\left[\begin{array}{c}\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})-\mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})-\mathcal{K}\left(0\right)\mathcal{K}^{+}\left(\boldsymbol{\varrho}\right)\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\}\\ \times\{\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})'\mathbf{Z}_{dt}'(\boldsymbol{\theta})-\left[\mathbf{e}_{dt}'(\boldsymbol{\theta},\boldsymbol{\varrho})-\mathbf{e}_{dt}'(\boldsymbol{\theta},0)\mathcal{K}^{+}\left(\boldsymbol{\varrho}\right)\mathcal{K}\left(0\right)\right]\mathbf{Z}_d'(\boldsymbol{\theta},\boldsymbol{\varrho})\}\end{array}\middle|\,\boldsymbol{\theta},\boldsymbol{\varrho}\right]$$
$$=E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})\mathbf{e}_{dt}'(\boldsymbol{\theta},\boldsymbol{\varrho})\mathbf{Z}_{dt}(\boldsymbol{\theta})|\boldsymbol{\theta},\boldsymbol{\varrho}\right]$$
$$-E\left\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})\left[\mathbf{e}_{dt}'(\boldsymbol{\theta},\boldsymbol{\varrho})-\mathbf{e}_{dt}'(\boldsymbol{\theta},\mathbf{0})\mathcal{K}^{+}\left(\boldsymbol{\varrho}\right)\mathcal{K}\left(0\right)\right]\mathbf{Z}_d'(\boldsymbol{\theta},\boldsymbol{\varrho})|\boldsymbol{\theta},\boldsymbol{\varrho}\right\}$$
$$-E\left\{\mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})-\mathcal{K}\left(0\right)\mathcal{K}^{+}\left(\boldsymbol{\varrho}\right)\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\mathbf{e}_{dt}(\boldsymbol{\phi})'\mathbf{Z}_{dt}'(\boldsymbol{\theta})|\boldsymbol{\theta},\boldsymbol{\varrho}\right\}$$
$$+E\left\{\mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})-\mathcal{K}\left(0\right)\mathcal{K}^{+}\left(\boldsymbol{\varrho}\right)\mathbf{e}_{dt}(\boldsymbol{\theta},0)\right]\left[\mathbf{e}_{dt}'(\boldsymbol{\theta},\boldsymbol{\varrho})-\mathbf{e}_{dt}'(\boldsymbol{\theta},0)\mathcal{K}^{+}\left(\boldsymbol{\varrho}\right)\mathcal{K}\left(0\right)\right]\mathbf{Z}_d'(\boldsymbol{\theta},\boldsymbol{\varrho})|\boldsymbol{\theta},\boldsymbol{\varrho}\right\}$$
$$=\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta},\boldsymbol{\varrho})-\mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})\left[\mathcal{M}_{dd}\left(\boldsymbol{\varrho}\right)-\mathcal{K}\left(0\right)\mathcal{K}^{+}\left(\boldsymbol{\varrho}\right)\mathcal{K}\left(0\right)\right]\mathbf{Z}_d'(\boldsymbol{\theta},\boldsymbol{\varrho})$$

by virtue of (C22), (B3) and the law of iterated expectations. □

In the case of the univariate GARCH-M model (19), we estimate the model parameters using reparametrisation 2 in section 4. Specifically, expressions (D2) and (D4) become

$$\mathbf{Z}_{lt}(\boldsymbol{\varphi})=\frac{\partial\mu_t(\boldsymbol{\varphi})/\partial\boldsymbol{\varphi}}{\varphi_{ic}^{1/2}\sigma_t^{\diamond}(\boldsymbol{\varphi}_c)}=\frac{1}{\varphi_{ic}^{1/2}\sigma_t^{\diamond}(\boldsymbol{\varphi}_c)}\left[\begin{array}{c}\frac{1}{2}\varphi_{im}\sigma_t^{\diamond-1}(\boldsymbol{\varphi}_c)\partial\sigma_t^{\diamond 2}(\boldsymbol{\varphi}_c)/\partial\boldsymbol{\varphi}_c\\ \sigma_t^{\diamond}(\boldsymbol{\varphi}_c)\\ 0\end{array}\right]=\left[\begin{array}{c}\varphi_{im}\varphi_{ic}^{-1/2}\mathbf{W}_{\boldsymbol{\varphi}_c t}(\boldsymbol{\varphi}_c)\\ \varphi_{ic}^{-1/2}\\ 0\end{array}\right],$$
$$\mathbf{Z}_{st}(\boldsymbol{\varphi})=\frac{\partial\sigma_t^2(\boldsymbol{\varphi})/\partial\boldsymbol{\varphi}}{2\varphi_{ic}\sigma_t^{\diamond 2}(\boldsymbol{\varphi}_c)}=\frac{1}{2\varphi_{ic}\sigma_t^{\diamond 2}(\boldsymbol{\varphi}_c)}\left[\begin{array}{c}\varphi_{ic}\partial\sigma_t^{\diamond 2}(\boldsymbol{\varphi}_c)/\partial\boldsymbol{\varphi}_c\\ 0\\ \sigma_t^{\diamond 2}(\boldsymbol{\varphi}_c)\end{array}\right]=\left[\begin{array}{c}\mathbf{W}_{\boldsymbol{\varphi}_c t}(\boldsymbol{\varphi}_c)\\ 0\\ \frac{1}{2}\varphi_{ic}^{-1}\end{array}\right]$$

and

$$e_{lt}(\boldsymbol{\varphi},\boldsymbol{\varrho})=-\frac{\partial\ln f[\epsilon_t(\boldsymbol{\varphi});\boldsymbol{\rho}]}{\partial\varepsilon},$$
$$e_{st}(\boldsymbol{\varphi},\boldsymbol{\varrho})=-\left\{1+\epsilon_t(\boldsymbol{\varphi})\frac{\partial\ln f[\epsilon_t(\boldsymbol{\varphi});\boldsymbol{\rho}]}{\partial\varepsilon}\right\},$$

respectively, where

$$\epsilon_t(\boldsymbol{\varphi})=\frac{\epsilon_t^{\diamond}(\boldsymbol{\varphi}_c)-\varphi_{im}}{\varphi_{ic}^{1/2}}=\frac{x_t}{\varphi_{ic}^{1/2}\sigma_t^{\diamond}(\boldsymbol{\varphi}_c)}-\frac{\varphi_{im}}{\varphi_{ic}^{1/2}}=\frac{x_t-\varphi_{im}\sigma_t^{\diamond}(\boldsymbol{\varphi}_c)}{\varphi_{ic}^{1/2}\sigma_t^{\diamond}(\boldsymbol{\varphi}_c)}$$

and

$$\mathbf{W}_{\boldsymbol{\varphi}_c t}(\boldsymbol{\varphi}_c)=\frac{1}{2\sigma_t^{\diamond 2}(\boldsymbol{\varphi}_c)}\frac{\partial\sigma_t^{\diamond 2}(\boldsymbol{\varphi}_c)}{\partial\boldsymbol{\varphi}_c}.$$

Then, a direct application of (D3) yields

$$\mathbf{s}_{\boldsymbol{\varphi}t}(\boldsymbol{\phi})=\left[\begin{array}{cc}\mathbf{Z}_{lt}(\boldsymbol{\varphi}) & \mathbf{Z}_{st}(\boldsymbol{\varphi})\end{array}\right]\left[\begin{array}{c}e_{lt}(\boldsymbol{\varphi},\boldsymbol{\varrho})\\ e_{st}(\boldsymbol{\varphi},\boldsymbol{\varrho})\end{array}\right]=\left[\begin{array}{c}\mathbf{W}_t(\boldsymbol{\varphi}_c)\mathbf{r}'(\boldsymbol{\varphi}_i)\\ \boldsymbol{\Delta}(\varphi_{ic})\end{array}\right]\left[\begin{array}{c}e_{lt}(\boldsymbol{\varphi},\boldsymbol{\varrho})\\ e_{st}(\boldsymbol{\varphi},\boldsymbol{\varrho})\end{array}\right],$$

where

$$\mathbf{r}(\boldsymbol{\varphi}_i) = \left(\begin{array}{cc} \varphi_{im}\varphi_{ic}^{-1/2} & 1 \end{array} \right)'$$

and

$$\boldsymbol{\Delta}(\varphi_{ic}) = \left(\begin{array}{cc} \varphi_{ic}^{-1/2} & 0 \\ 0 & \frac{1}{2}\varphi_{ic}^{-1} \end{array} \right).$$

On the other hand, we use again the natural parametrisation of the multivariate market model in (20). As a result, the Jacobian matrix (C36) in Supplemental Appendix C remains relevant, so that

$$\begin{aligned} \mathbf{s}_{\mathbf{a}t}(\boldsymbol{\theta}) &= -\boldsymbol{\Omega}^{-1/2}\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\rho}]/\partial\boldsymbol{\varepsilon}^*, \\ \mathbf{s}_{\mathbf{b}t}(\boldsymbol{\theta}) &= -\boldsymbol{\Omega}^{-1/2} r_{mt}\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\rho}]/\partial\boldsymbol{\varepsilon}^*, \end{aligned}$$

where $\boldsymbol{\Omega}^{1/2}$ is a matrix square root of $\boldsymbol{\Omega}$.

If we choose the Cholesky decomposition, we can use expression (D14) to obtain

$$\mathbf{s}_{\boldsymbol{\omega}t}(\boldsymbol{\theta}) = -\frac{1}{2}\mathbf{D}_N'\mathbf{F}\mathbf{L}_N(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-\frac{1}{2}})vec\left\{\mathbf{I}_N + \partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\rho}]/\partial\boldsymbol{\varepsilon}^* \cdot \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\},$$

where $\mathbf{F}$ denotes the transpose of the inverse of $\mathbf{L}_N(\boldsymbol{\Omega}^{1/2} \otimes \mathbf{I}_N)\mathbf{L}_N' + \mathbf{L}_N(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{1/2})\mathbf{K}_{NN}\mathbf{L}_N'$.

Finally, it is worth noting that it is possible to avoid the use of explicit Moore-Penrose generalised inverses in the computation of the correction by exploiting the fact that

$$\mathcal{K}(\boldsymbol{\rho}) = \left(\begin{array}{cc} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_N \end{array} \right) \left[\begin{array}{cc} \mathbf{I}_N & E[\boldsymbol{\varepsilon}_t^* vech'(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})|\,\boldsymbol{\theta},\boldsymbol{\varrho}] \\ E[vech(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})\boldsymbol{\varepsilon}_t^{\prime *}|\,\boldsymbol{\theta},\boldsymbol{\varrho}] & E[vech(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})vech'(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime}) - \mathbf{I}_N|\,\boldsymbol{\theta},\boldsymbol{\varrho}] \end{array} \right] \left(\begin{array}{cc} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_N' \end{array} \right)$$

and

$$\mathcal{K}(\mathbf{0}) = \left(\begin{array}{cc} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N^2} + \mathbf{K}_{NN} \end{array} \right)$$

imply that

$$\begin{aligned} &\mathcal{K}(0)\,\mathcal{K}^+(\boldsymbol{\rho})\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0}) = \left(\begin{array}{cc} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & 2\mathbf{D}^{+\prime} \end{array} \right) \\ &\times \left[\begin{array}{cc} \mathbf{I}_N & E[\boldsymbol{\varepsilon}_t^* vech'(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})|\,\boldsymbol{\theta},\boldsymbol{\varrho}] \\ E[vech(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})\boldsymbol{\varepsilon}_t^{\prime *}|\,\boldsymbol{\theta},\boldsymbol{\varrho}] & E[vech(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})vech'(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime}) - \mathbf{I}_N|\,\boldsymbol{\theta},\boldsymbol{\varrho}] \end{array} \right]^{-1} \left[\begin{array}{c} \boldsymbol{\varepsilon}_t^* \\ vech(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime} - \mathbf{I}) \end{array} \right]. \end{aligned}$$

Nevertheless, $f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\rho})$ has to be replaced by a nonparametric estimator, which increasingly suffers from the curse of dimensionality as the cross-sectional dimension N increases. In line with the usual practice, we employ a standard multivariate Gaussian kernel. Once again, we have done some experimentation to choose optimal bandwidths by scaling up and down the automatic choices given in Silverman (1986) because a proper cross-validation procedure is extremely costly to implement in a Monte Carlo exercise when $N = 3$.

E Other results

E.1 Standardised two component mixtures of multivariate normals

Consider the following mixture of two multivariate normals

$$\varepsilon_t \sim \begin{cases} N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) & \text{with probability} \quad \lambda, \\ N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) & \text{with probability} \quad 1-\lambda. \end{cases}$$

Let d_t denote a Bernoulli variable which takes the value 1 with probability λ and 0 with probability $1-\lambda$. As is well known, the unconditional mean vector and covariance matrix of the observed variables are:

$$\begin{aligned} E(\varepsilon_t) &= E[E(\varepsilon_t|d_t)] = \lambda\boldsymbol{\mu}_1 + (1-\lambda)\boldsymbol{\mu}_2, \\ V(\varepsilon_t) &= V[E(\varepsilon_t|d_t)] + E[V(\varepsilon_t|d_t)] = \lambda(1-\lambda)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)' + \lambda\boldsymbol{\Sigma}_1 + (1-\lambda)\boldsymbol{\Sigma}_2. \end{aligned}$$

Therefore, this random vector will be standardised if and only if

$$\begin{aligned} \lambda\boldsymbol{\mu}_1 + (1-\lambda)\boldsymbol{\mu}_2 &= \mathbf{0}, \\ \lambda(1-\lambda)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)' + \lambda\boldsymbol{\Sigma}_1 + (1-\lambda)\boldsymbol{\Sigma}_2 &= \mathbf{I}. \end{aligned}$$

Let us initially assume that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}$ but that the mixture is not degenerate, so that $\lambda \neq 0, 1$. Let $\boldsymbol{\Sigma}_{1L}\boldsymbol{\Sigma}'_{1L}$ and $\boldsymbol{\Sigma}_{2L}\boldsymbol{\Sigma}'_{2L}$ denote the Cholesky decompositions of the covariance matrices of the two components. Then, we can write

$$\lambda\boldsymbol{\Sigma}_1 + (1-\lambda)\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_{1L}[\lambda\mathbf{I}_N + (1-\lambda)\boldsymbol{\Sigma}_{1L}^{-1}\boldsymbol{\Sigma}_{2L}\boldsymbol{\Sigma}'_{2L}\boldsymbol{\Sigma}_{1L}^{-1\prime}]\boldsymbol{\Sigma}'_{1L} = \boldsymbol{\Sigma}_{1L}(\lambda\mathbf{I}_N + \mathbf{K}_L\mathbf{K}'_L)\boldsymbol{\Sigma}'_{1L},$$

where $\mathbf{K}_L = \sqrt{1-\lambda}\boldsymbol{\Sigma}_{1L}^{-1}\boldsymbol{\Sigma}_{2L}$ remains a lower triangular matrix. Given that $\mathbf{I}_N = \mathbf{e}_1\mathbf{e}_1 + \ldots + \mathbf{e}_N\mathbf{e}_N$, where $\mathbf{e}_i$ is the i^{th} vector of the canonical basis, the Cholesky decomposition of $\lambda\mathbf{I}_N + \mathbf{K}_L\mathbf{K}'_L$, say $\mathbf{J}_L\mathbf{J}'_L$, can be computed by means of N rank-one updates that sequentially add $\sqrt{\lambda}\mathbf{e}_i\sqrt{\lambda}\mathbf{e}'_i$ for $i = 1, \ldots, N$. The special form of those vectors can be efficiently combined with the usual rank-one update algorithms to speed up this process (see e.g. Sentana (1999) and the references therein). In any case, the elements of $\mathbf{J}_L$ will be functions of λ and the $N(N+1)/2$ elements in $\mathbf{K}_L$. If we then choose $\boldsymbol{\Sigma}_{1L} = \mathbf{J}_L^{-1}$, we will guarantee that $\lambda\boldsymbol{\Sigma}_1 + (1-\lambda)\boldsymbol{\Sigma}_2 = \mathbf{I}_N$. Therefore, we can achieve a standardised two-component mixture of two multivariate normals with 0 means by drawing with probability λ one random variable from a distribution with covariance matrix $\mathbf{J}_L^{-1\prime}\mathbf{J}_L^{-1}$, and with probability $1-\lambda$ from another distribution with covariance matrix $(1-\lambda)^{-1}\mathbf{K}_L\mathbf{K}'_L$.

Let us now turn to the case in which the means of the components are no longer 0. The zero unconditional mean condition is equivalent to $\boldsymbol{\mu}_1 = (1-\lambda)\boldsymbol{\delta}$ and $\boldsymbol{\mu}_2 = -\lambda\boldsymbol{\delta}$, so that $\boldsymbol{\delta}$ measures the difference between the two means. Thus, the unconditional covariance matrix will be $\lambda(1-\lambda)\boldsymbol{\delta}\boldsymbol{\delta}' + \mathbf{I}_N$ after imposing the restrictions on $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ in the previous paragraph. Once again, the Cholesky decomposition of this matrix is very easy to obtain because it can be regarded as a positive rank-one update of the identity matrix, whose decomposition is trivial.

Thus, we can parametrise a standardised mixture of two multivariate normals, which usually involves $2N$ mean parameters, $2N(N+1)/2$ covariance parameters and one mixing parameter, in terms of the N mean difference parameters in $\boldsymbol{\delta}$, the $N(N+1)/2$ relative variance parameters in $\mathbf{K}_L$ and the mixing parameter λ, the remaining N mean parameters and $N(N+1)/2$ covariance ones freed up to target any unconditional mean vector and covariance matrix.

Mencía and Sentana (2009) explain how to standardise Bernoulli location-scale mixtures of normals, which are a special case of the two component mixtures we have just discussed in which $\boldsymbol{\Sigma}_2 = \varkappa\boldsymbol{\Sigma}_1$. Straightforward algebra confirms that the standardisation procedure described above simplifies to the one they provide in their Proposition 1.

E.2 Non-causal ARMA models

Consider the following AR(2) process:

$$(1-\alpha_1 L)(1-\alpha_2 L)x_t = \mu + \xi_t, \tag{E1}$$

where ξ_t is a possibly non-Gaussian *i.i.d.* sequence, $\alpha_1, \alpha_2 \in \mathbb{R}$, $|\alpha_1| < 1$, $|\alpha_2| > 1$ but $\alpha_2 \neq \alpha_1^{-1}$. Higher order process with possibly complex roots can be handled analogously, but the algebra gets messier. Brockwell and Davis (1987) showed that x_t can be written as the following doubly infinite MA process

$$x_t = \frac{-\alpha_2^{-1}\mu}{(1-\alpha_1)(1-\alpha_2^{-1})} - (\ldots + \alpha_2^{-2}L^{-3} + \alpha_2^{-1}L^{-2} + L^{-1} + \alpha_1 + \alpha_1^2 L + \alpha_1^3 L^2 + \alpha_1^4 L^3 + \ldots)\frac{\xi_t}{\alpha_2 - \alpha_1},$$

which they called mixed causal/non-causal because x_t effectively depends on past, present and future values of the underlying innovations. Nevertheless, by looking at the spectral density of x_t they also showed that this process has the following purely causal AR(2) representation:

$$(1-\alpha_1 L)(1-\alpha_2^{-1} L)x_t = \nu + u_t, \tag{E2}$$

where u_t is a white noise but not necessarily serially independent sequence, with variance $\sigma_u^2 = \alpha_2^{-2}\sigma_\xi^2$ and $\nu = -\alpha_2^{-1}\mu$. Thus, the situation is entirely analogous to the well known multiple invertible and non-invertible representations of MA processes.

Breidt et al (1991) showed that a non-Gaussian log-likelihood function based on the assumption that the distribution of ξ_t is *i.i.d.* with 0 mean and finite variance σ_ξ^2 will be able to consistently estimate the values of the two autoregressive roots that appear in (E1) as well as the true drift and variance of the innovations. In contrast, a Gaussian log-likelihood function, which effectively exploits the information in the spectral density of x_t, can only consistently estimate the parameters in (E2).

At first sight, it might appear that one cannot apply the procedures we have developed in the paper to assess the adequacy of the non-Gaussian distribution chosen for the purposes of estimating the "structural" parameters because the Gaussian pseudo log-likelihood cannot consistently estimate them. However, under correct specification, the non-Gaussian log-likelihood function will also estimate α_1, α_2^{-1}, $-\alpha_2^{-1}\mu$ and $\alpha_2^{-2}\sigma_\xi^2$ consistently. Therefore, one can easily develop a DWH specification test to check the validity of the distributional assumption for ξ_t by comparing the non-Gaussian coefficient estimators of those "reduced form" parameters with the Gaussian ones. The score versions of those tests that we discussed in section 2.1 are also straightforward. As we have argued in section 3.7, power gains may be obtained by focusing on ν and σ_u^2.

E.3 Additional Monte Carlo results

In this section, we look at the sampling distribution of the estimators we used in section 4 to compute the DWH tests of the univariate Garch-M model and the multivariate market model.

Univariate GARCH-M Table 1S displays the Monte Carlo medians and interquartile ranges of the estimators. The results broadly confirm the theoretical predictions in terms of bias and relative efficiency. It is worth noticing that the bias of the restricted (unrestricted) Student t maximum likelihood estimators of the scale parameter is negative (positive) when the log-likelihood is misspecified, which suggests that our tests will have good power for pairwise comparisons involving this parameter, at least for the distributions considered in the exercise. In turn, the location parameter estimators are biased only when the true distribution is asymmetric.

Multivariate market model Table 2S displays the Monte Carlo medians and interquartile ranges of the estimators for several representative parameters in addition to the global scale parameter $\vartheta_i = |\mathbf{\Omega}|^{1/N}$. Specifically, we exploit the exchangeability of our design to pool the results of all the elements of the vectors of intercepts $\mathbf{a}$ and slopes $\mathbf{b}$, and the "vectors" of residual covariance parameters $vecd(\mathbf{\Omega}^\circ)$, $vecl(\mathbf{\Omega}^\circ)$, $vecd(\mathbf{\Omega})$ and $vecl(\mathbf{\Omega})$. Once again, the results are in line with the theoretical predictions. Moreover, the biases of the restricted and unrestricted Student t maximum likelihood estimators of the global scale parameter have opposite signs, as in the univariate case. Finally, the location parameters are only biased in the asymmetric distribution simulations. Therefore, we expect tests that involve the intercepts to increase power in that case, but to result in a waste of degrees of freedom otherwise.

Additional references

Abramowitz, M., and Stegun, I.A. (1964): *Handbook of mathematical functions*, AMS 55, National Bureau of Standards.

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

Breidt, J., Davis, R.A., Lh, K.S. and Rosenblatt, M. (1991): "Maximum likelihood estimation for noncausal autoregressive processes", *Journal of Multivariate Analysis* 36, 175-198.

Brockwell, P.J. and Davis, R.A. (1987): *Time series: theory and methods* (2nd ed.), Springer.

Chen, S.X. (2000): "Probability density function estimation using gamma kernels", *Annals of the Institute of Statistical Mathematics* 52, 471-480.

Fang, K.-T., Kotz, S., and Ng, K.-W. (1990), *Symmetric multivariate and related distributions*, Chapman and Hall.

Fiorentini, G. and Sentana, E. (2015): "Tests for serial dependence in static, non-Gaussian factor models", in S.J. Koopman and N. Shephard (eds.) *Unobserved Components and Time Series Econometrics*, pp. 118-189, Oxford University Press.

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

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

Sentana, E. (1999): "Econometric applications of positive rank-one modifications of the symmetric factorization of a positive semi-definite matrix", *Spanish Economic Review* 1, 79-90.

Silverman B.W. (1986): *Density estimation*, Chapman and Hall.

TABLE 1S: Univariate GARCH-M: Parameter estimators.

Parameter		β	γ	δ, φ_{im}	$\vartheta_i, \varphi_{ic}$	$\eta = 1/\nu$
True value		0.85	0.1	0.05	1.0	
	RML	0.8467	0.0960	0.0506	1.0404	0.0833
		(0.0375)	(0.0348)	(0.0314)	(0.4132)	
	UML	0.8467	0.0959	0.0507	1.0397	0.0815
Student t_{12}		(0.0376)	(0.0350)	(0.0315)	(0.4125)	(0.0276)
	PML	0.8464	0.0956	0.0508	1.0420	
		(0.0392)	(0.0363)	(0.0324)	(0.4331)	
	RML	0.8467	0.0956	0.0505	1.0137	0.0833
		(0.0383)	(0.0344)	(0.0315)	(0.3986)	
	UML	0.8468	0.0959	0.0504	1.0392	0.1232
Student t_8		(0.0381)	(0.0343)	(0.0314)	(0.4077)	(0.0276)
	PML	0.8460	0.0955	0.0504	1.0439	
		(0.0423)	(0.0384)	(0.0333)	(0.4539)	
	RML	0.8461	0.0955	0.0506	0.8706	0.0833
		(0.0437)	(0.0383)	(0.0278)	(0.3817)	
	UML	0.8470	0.0967	0.0502	1.3990	0.3604
GC(0,3.2)		(0.0371)	(0.0338)	(0.0254)	(0.5748)	(0.0264)
	PML	0.8460	0.0956	0.0506	1.0425	
		(0.0429)	(0.0377)	(0.0327)	(0.4476)	
	RML	0.8460	0.0956	0.1117	0.8601	0.0833
		(0.0436)	(0.0386)	(0.0358)	(0.3848)	
	UML	0.8475	0.0970	0.1723	1.5853	0.3865
GC(-.9,3.2)		(0.0356)	(0.0321)	(0.0380)	(0.6728)	(0.0265)
	PML	0.8459	0.0956	0.0511	1.0453	
		(0.0431)	(0.0381)	(0.0326)	(0.4626)	

Monte Carlo medians and (interquartile ranges) of RML (Student t-based maximum likelihood with 12 degrees of freedom), UML (unrestricted Student t-based maximum likelihood), and PML (Gaussian pseudo maximum likelihood) estimators. GC (Gram-Charlier expansion). Sample length=2,000. Replications=20,000.

TABLE 2S: Multivariate market model: Parameter estimators.

Parameter		**a**	**b**	ϑ_i	ω°_{ii}	ω°_{ij}	ω_{ii}	ω_{ij}	$\eta = 1/\nu$
True value		0.112	1	2.8917	1.0845	0.3253	3.136	0.9408	
	RML	0.1124	0.9989	2.8702	1.0872	0.3262	3.1215	0.9355	0.0833
		(0.1040)	(0.1173)	(0.1696)	(0.0808)	(0.0705)	(0.2955)	(0.2115)	
Student t_{12}	UML	0.1123	0.9989	2.8674	1.0872	0.3262	3.1176	0.9347	0.0810
		(0.1041)	(0.1174)	(0.1815)	(0.0808)	(0.0706)	(0.3043)	(0.2117)	(0.0280)
	PML	0.1124	0.9998	2.8646	1.0873	0.3262	3.1147	0.9341	
		(0.1066)	(0.1213)	(0.1807)	(0.0849)	(0.0738)	(0.3125)	(0.2200)	
	RML	0.1127	0.9989	2.7652	1.0874	0.3259	3.0077	0.9008	0.0833
		(0.1015)	(0.1148)	(0.1763)	(0.0832)	(0.0723)	(0.2980)	(0.2078)	
Student t_8	UML	0.1126	0.9989	2.8683	1.0875	0.3259	3.1211	0.9352	0.1233
		(0.1013)	(0.1144)	(0.2088)	(0.0831)	(0.0718)	(0.3301)	(0.2171)	(0.0304)
	PML	0.1126	0.9988	2.8618	1.0877	0.3259	3.1129	0.9318	
		(0.1075)	(0.1219)	(0.2085)	(0.0927)	(0.0803)	(0.3649)	(0.2391)	
	RML	0.1123	0.9995	2.0600	1.0882	0.3264	2.2402	0.6705	0.0833
		(0.0803)	(0.0912)	(0.1989)	(0.0975)	(0.0853)	(0.2945)	(0.1886)	
DSMN(0.2,0.1)	UML	0.1125	0.9997	3.5341	1.0878	0.3262	3.8521	1.1545	0.3474
		(0.0775)	(0.0874)	(0.8393)	(0.0877)	(0.0765)	(0.9692)	(0.3848)	(0.0372)
	PML	0.1119	1.0000	2.8483	1.0907	0.3266	3.1071	0.9280	
		(0.1071)	(0.1202)	(0.3197)	(0.1241)	(0.1077)	(0.4966)	(0.3275)	
	RML	-0.0003	1.0004	2.0275	1.0829	0.3140	2.1962	0.6351	0.0833
		(0.0830)	(0.0891)	(0.1984)	(0.0991)	(0.0868)	(0.2980)	(0.1900)	
DLSMN(0.2,0.1,0.5)	UML	-0.0576	1.0006	3.7270	1.0753	0.2986	4.0177	1.1141	0.3616
		(0.0831)	(0.0854)	(0.9916)	(0.0880)	(0.0763)	(1.1239)	(0.4204)	(0.0373)
	PML	0.1119	1.0010	2.8485	1.0908	0.3271	3.1067	0.9295	
		(0.1065)	(0.1204)	(0.3152)	(0.1252)	(0.1097)	(0.4948)	(0.3306)	

Monte Carlo medians and (interquartile ranges) of RML (Student t-based maximum likelihood with 12 degrees of freedom), UML (unrestricted Student t-based maximum likelihood), and PML (Gaussian pseudo maximum likelihood) estimators. DSMN (discrete scale mixture of two normals), DLSMN (discrete location-scale mixture of two normals). Sample length=500. Replications=20,000.