May/2018
Working Paper 18-22
rcea.org/RePEc/pdf/wp18-22.pdf

SPECIFICATION TESTS FOR NON-GAUSSIAN MAXIMUM LIKELIHOOD ESTIMATORS

Gabriele Fiorentini
Università di Firenze, Italy
RCEA

Enrique Sentana
CEMFI, Spain

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: May 2018

Abstract

We propose generalised DWH specification tests which simultaneously compare three or more likelihood-based estimators of conditional mean and variance parameters 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. To design powerful and reliable tests, we determine the rank deficiencies of the differences between the estimators' asymptotic covariance matrices under the null of correct specification, and take into account that some parameters remain consistently estimated under the alternative of distributional misspecification. Finally, we provide finite sample results through Monte Carlo simulations.

Keywords: Durbin-Wu-Hausman Tests, Partial Adaptivity, Semiparametric Estimators, Singular Covariance Matrices.

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 for useful comments and suggestions, as well as audiences at CREST, Konstanz, Tokyo, 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) and the Bilgi CEFIS conference on "Advances in Econometrics Methods" (Instanbul, March 2018). Of course, the usual caveat applies. The second author gratefully acknowledges support from the Santander - CEMFI Research Chair on Finance.

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. As a result, empirical researchers often specify a parametric leptokurtic distribution, which then they use to estimate their models by maximum likelihood (ML). The dominant commercially available econometric packages have responded to this practice by offering ML procedures that estimate the parameters of the conditional means, variances and covariances of the observed series either jointly with the parameters characterising the shape of the assumed distribution or allowing the user to fix these to some sensible values. Specifically, EVIEWS and STATA support Student t and Generalised Error distributions (GED) in univariate models (see the ARCH sections of IHS Global Inc (2015) and StataCorp LP (2015)), while STATA additionally allows for Student t innovations in multivariate ones (see the MARCH section of StataCorp LP (2015)).

An additional non-trivial benefit of these procedures is that they deliver (weakly) more efficient estimators of the mean and variance parameters, especially if the shape parameters can be fixed to their true values. The problem with non-Gaussian ML estimators, 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 irrespective of the degree of asymmetry and kurtosis of the conditional distribution of the observed variables, so long as the first two moments are correctly specified and the fourth moments are bounded.

If researchers were only interested in the first two conditional moments of the data, the semiparametric (SP) estimators of Engle and Gonzalez-Rivera (1991) and Gonzalez-Rivera and Drost (1999) would provide an attractive solution because they are not only consistent but also attain full efficiency for a subset of the parameters, as shown by Linton (1993), Drost and Klaassen (1997), Drost, Klaassen and Werker (1997) and Sun and Stengos (2006) in several univariate time series examples. Unfortunately, SP estimators suffer from the curse of dimensionality when the number of series involved, N, is moderately large, which severely limits their use in multivariate models. 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 distribution will again contaminate these estimators.

In a companion paper (Fiorentini and Sentana (2018)), we characterise the mean and variance parameters that distributionally misspecified ML estimators can consistently estimate, and provide simple closed-form consistent estimators for the rest. Francq, Lepage and Zakoïan (2011) and Fan, Qi and Xiu (2014) have also proposed alternative consistent estimators for univariate GARCH models without mean, which are asymptotically equivalent to ours in that context. Nevertheless, many empirical researchers will continue to rely on the estimators that the econometric software packages provide. For that reason, it would be desirable that they would routinely complement their empirical results with some formal indication of the validity of the parametric assumptions they make for estimation purposes.

There are several ways to do so. 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 nesting distribution for the multivariate Student t. An alternative procedure would be an information matrix test that compares some or all of the elements of the expected Hessian and the variance of the score. But when an empirical researcher relies on standard software for calculating some estimators of $\boldsymbol{\theta}$ and their asymptotic standard errors, a more natural approach to testing the distributional specification would be to compare those estimators on a pairwise basis using simple Durbin-Wu-Hausman (DWH) tests. 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. 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. We also explore several important issues related to the practical implementation of these tests, including its two score versions, their numerical invariance to reparametrisations and their application to subsets of parameters.

To design powerful and reliable tests, though, we first need to study the consistency and efficiency properties of the different estimators involved. In particular, we need to figure out the rank of the difference between the corresponding asymptotic covariance matrices under the null of correct specification to select 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. Otherwise our tests will use up degrees of freedom without providing any power gains.

Importantly, we find 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 Var models with non-Gaussian innovations). Obviously, our approach also applies in cross-sectional models with exogenous regressors, as well as in static ones. Another important feature of our analysis is that we explicitly look at the unrestricted ML procedure that jointly estimates the shape parameters, as well as the Gaussian PML, SP, SSP and restricted ML estimators considered in the existing literature.

The rest of the paper is organised 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 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 those tests can be found in section 4. Finally, we present our conclusions in section 5. Proofs and auxiliary results are gathered in appendices.

2 Durbin-Wu-Hausman tests

2.1 Wald and two score versions

As we mentioned in the introduction, DWH tests exploit the diverging behaviour under the alternative of two estimators for the same parameters to test the correct specification of the model under the null. The standard calculation of DWH tests, though, requires the prior computation of those two estimators. 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 spells out those equivalences:

Proposition 1 *Let $\hat{\boldsymbol{\theta}}_T$ and $\tilde{\boldsymbol{\theta}}_T$ denote two root-T consistent, asymptotically Gaussian, 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 the weighting matrices $\tilde{\mathcal{S}}_{mT}$ and $\tilde{\mathcal{S}}_{nT}$, respectively. Then, under standard regularity conditions*

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

and

$$T(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T)'\boldsymbol{\Delta}^{-}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) - T\bar{\mathbf{n}}_T'(\hat{\boldsymbol{\theta}}_T)\mathcal{S}_n\mathcal{J}_n(\boldsymbol{\theta}_0)\boldsymbol{\Psi}_n^{-}\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n\bar{\mathbf{n}}_T(\hat{\boldsymbol{\theta}}_T) = o_p(1),$$

where $^{-}$ denotes a generalised inverse, $\boldsymbol{\Delta}$, $\boldsymbol{\Psi}_m$ and $\boldsymbol{\Psi}_n$ are the limiting variances of $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T)$, $\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T)$ 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

$$\begin{aligned}\boldsymbol{\Delta} &= \left[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right]^{-1}\boldsymbol{\Psi}_m\left[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right]^{-1} \\ &= \left[\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n\mathcal{J}_n(\boldsymbol{\theta}_0)\right]^{-1}\boldsymbol{\Psi}_n\left[\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n\mathcal{J}_n(\boldsymbol{\theta}_0)\right]^{-1},\end{aligned}$$

with

$$\mathcal{J}_m(\boldsymbol{\theta}) = \operatorname*{plim}_{T\to\infty}\partial\bar{\mathbf{m}}_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}', \quad \mathcal{J}_n(\boldsymbol{\theta}) = \operatorname*{plim}_{T\to\infty}\partial\bar{\mathbf{n}}_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}',$$
$$\mathcal{S}_m = \operatorname*{plim}_{T\to\infty}\tilde{\mathcal{S}}_{mT}, \quad \mathcal{S}_n = \operatorname*{plim}_{T\to\infty}\tilde{\mathcal{S}}_{nT},$$

and

$$\operatorname{rank}\left[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right] = \operatorname{rank}\left[\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n\mathcal{J}_n(\boldsymbol{\theta}_0)\right] = p = \dim(\boldsymbol{\theta}).$$

An intuitive way of re-interpreting the asymptotic equivalence between the original Wald-type version of the DWH test and the two alternative score versions is to think of the latter as Wald-type 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 $\boldsymbol{\pi}_m(\boldsymbol{\theta}) = \mathcal{J}_m'(\boldsymbol{\theta})\mathcal{S}_m E[\mathbf{m}_t(\boldsymbol{\theta})]$ and $\boldsymbol{\pi}_n(\boldsymbol{\theta}) = \mathcal{J}_n'(\boldsymbol{\theta})\mathcal{S}_n E[\mathbf{n}_t(\boldsymbol{\theta})]$. 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 $\mathcal{J}_m'(\boldsymbol{\theta}_m)\mathcal{S}_m E[\mathbf{m}_t(\boldsymbol{\theta}_m)] = \mathbf{0}$ and $\mathcal{J}_n'(\boldsymbol{\theta}_n)\mathcal{S}_n E[\mathbf{n}_t(\boldsymbol{\theta}_n)] = \mathbf{0}$, 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.[1]

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. We will revisit these issues in sections 2.2 and 4.

[1] 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).

Consider a sequence of local alternatives such that

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

As is well known, the asymptotic distribution any of the DWH statistics above is chi-square with $r = \text{rank}(\boldsymbol{\Delta})$ degrees of freedom and non-centrality parameter $(\boldsymbol{\theta}_m - \boldsymbol{\theta}_n)'\boldsymbol{\Delta}^-(\boldsymbol{\theta}_m - \boldsymbol{\theta}_n)$ (see e.g. Hausman (1978) or Holly (1987)), which reduces to a central χ^2_r under the null hypothesis that both sets of moments are correctly specified. As a result, 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.

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:

Proposition 2 *Consider a homeomorphic, continuously differentiable transformation $\boldsymbol{\pi}(.)$ from $\boldsymbol{\theta}$ to a new set of parameters $\boldsymbol{\pi}$, with* rank $[\partial\boldsymbol{\pi}'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}] = p = \dim(\boldsymbol{\theta})$ *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*

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

   where $\boldsymbol{\Psi}^{\sim}_{mT}$ and $\boldsymbol{\Psi}^{\sim}_{\mathfrak{m}T}$, 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] A trivial non-random example of discontinuities is the sequence $1/T$, which converges to 0 while $(1/T)^- = T$ 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.

2.3 Subsets of parameters

In some well-known examples, generalised inverses can be avoided by working with a subset of parameters. In particular, if the two estimators of the last p_2 elements of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_{2T}$ and $\tilde{\boldsymbol{\theta}}_{2T}$, share the same asymptotic distribution, 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 there may be other reasons for focusing on a subset. For example, 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, then 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*

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

as two sets of p_1 transformed sample moment conditions, where

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

Then,

$$T(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T)'\boldsymbol{\Delta}^{-}_{11}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) - T\bar{\mathbf{m}}^{\perp\prime}_{T}(\tilde{\boldsymbol{\theta}}_T)\boldsymbol{\Psi}^{-}_{\mathbf{m}^{\perp}_1}\bar{\mathbf{m}}^{\perp\prime}_{T}(\tilde{\boldsymbol{\theta}}_T) = o_p(1)$$

and

$$T(\tilde{\boldsymbol{\theta}}_{1T} - \hat{\boldsymbol{\theta}}_{1T})'\boldsymbol{\Delta}^{-}_{11}(\tilde{\boldsymbol{\theta}}_{1T} - \hat{\boldsymbol{\theta}}_{1T}) - T\bar{\mathbf{n}}^{\perp\prime}_{1T}(\hat{\boldsymbol{\theta}}_T)\boldsymbol{\Psi}^{-}_{\mathbf{n}^{\perp}_1}\bar{\mathbf{n}}^{\perp}_{1T}(\hat{\boldsymbol{\theta}}_T) = o_p(1),$$

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

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

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

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

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

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 these 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 whereby 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, though, 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.[3] 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:

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

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

Then the simultaneous DWH tests that compares all J estimators can be decomposed as the sum of $J-1$ consecutive pairwise DWH tests, which are asymptotically mutually independent under the null of correct specification and sequences of local alternatives.

Therefore, the asymptotic distribution of the simultaneous DWH test will be a non-central chi-square 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 by chance under the null will be $1-(1-\alpha)^{J-1}$, where α is the significance level of each pairwise test.

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

Positive semidefiniteness of the covariance structure in (1) implies that one can rank (in the usual positive semidefinite sense) the asymptotic variance of the J estimators as

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

so that the sequence of estimators follows a decreasing efficiency order. Nevertheless, (1) 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 (1) 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.

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:

$$\begin{array}{c} \mathbf{y}_t = \boldsymbol{\mu}_t(\boldsymbol{\theta}) + \mathbf{\Sigma}_t^{1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*, \\ \boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}(I_{t-1};\boldsymbol{\theta}), \\ \mathbf{\Sigma}_t(\boldsymbol{\theta}) = \mathbf{\Sigma}(I_{t-1};\boldsymbol{\theta}), \end{array}$$

where $\boldsymbol{\mu}()$ and $vech\,[\mathbf{\Sigma}()]$ are $N\times1$ and $N(N+1)/2\times1$ vector functions describing the conditional mean vector and covariance matrix known up to the $p\times1$ 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 $\mathbf{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ is some particular "square root" matrix such that $\mathbf{\Sigma}_t^{1/2}(\boldsymbol{\theta})\mathbf{\Sigma}_t^{1/2\prime}(\boldsymbol{\theta}) = \mathbf{\Sigma}_t(\boldsymbol{\theta})$. 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

$$\left.\begin{array}{c} E(\mathbf{y}_t|I_{t-1}) = \boldsymbol{\mu}_t(\boldsymbol{\theta}_0) \\ V(\mathbf{y}_t|I_{t-1}) = \mathbf{\Sigma}_t(\boldsymbol{\theta}_0) \end{array}\right\}.$$

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 appendix B 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 appendix A.1 for a brief introduction to spherically symmetric distributions).[4] 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. For that reason, we define η as $1/\nu$, which will always remain in the finite range $[0,1/2)$ under our assumptions. Obviously, in the univariate case, any symmetric distribution, including the GED (also known as the Generalised Gaussian distribution), is spherically symmetric too.[5]

For illustrative purposes, we consider the following two empirically relevant examples throughout the paper:

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

$$\left.\begin{array}{r} r_{Mt} = \mu_t(\boldsymbol{\theta}) + \sigma_t(\boldsymbol{\theta})\varepsilon_t^*, \\ \mu_t(\boldsymbol{\theta}) = \tau\sigma_t(\boldsymbol{\theta}), \\ \sigma_t^2(\boldsymbol{\theta}) = \omega + \alpha r_{Mt-1}^2 + \beta\sigma_{t-1}^2(\boldsymbol{\theta}). \end{array}\right\} \tag{2}$$

The conditional mean and variance parameters are $\boldsymbol{\theta}' = (\tau,\omega,\alpha,\beta)$. Importantly, this model nests the one considered by Francq, Lepage and Zakoïan (2011) and Fan, Qi and Xiu (2014) when $\tau = 0$.

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

$$\mathbf{r}_t = \mathbf{a} + \mathbf{b}r_{Mt} + \boldsymbol{\Omega}^{1/2}\boldsymbol{\varepsilon}_t^*. \tag{3}$$

The conditional mean and variance parameters are $\boldsymbol{\theta}' = (\mathbf{a}',\mathbf{b}',\boldsymbol{\omega}')$, where $\boldsymbol{\omega} = vech(\boldsymbol{\Omega})$ and $\boldsymbol{\Omega} = \boldsymbol{\Omega}^{1/2}\boldsymbol{\Omega}'^{1/2}$.

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 $\boldsymbol{\theta}$:

[4] Nevertheless, Propositions 10, 14, A2 and A3 already deal explicitly with the general case, while Propositions 6, 7, 8 and 9 continue to be valid without sphericity. As explained below, the same applies to Proposition 5.

[5] See Gillier (2005) for a spherically symmetric multivariate version of the GED.

1. **Restricted ML (RML)**: $\hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}})$

which is such that

$$\hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}}) = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L_T(\boldsymbol{\theta}, \bar{\boldsymbol{\eta}}).$$

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

2. **Joint or Unrestricted ML (UML)**: $\hat{\boldsymbol{\theta}}_T$

which is obtained as

$$(\hat{\boldsymbol{\theta}}_T, \hat{\boldsymbol{\eta}}_T) = \arg\max_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} L_T(\boldsymbol{\theta}, \boldsymbol{\eta}).$$

This is characterised by the feasible parametric efficiency bound

$$\mathcal{P}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0). \quad (4)$$

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.[6] 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 typically computed with one BHHH iteration starting from a consistent estimator, but this time based on the efficient score. 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.

In appendix A, we provide further details on these five estimators and their efficiency bounds.

[6] 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.

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 $\varepsilon_t^*|I_{t-1};\phi_0$ is i.i.d. $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta}_0)$ with bounded fourth moments, then*

$$\lim_{T\to\infty} V\left[\frac{\sqrt{T}}{T}\sum_{t=1}^{T}\begin{pmatrix}\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)\\ \mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\phi}_0)\\ \mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)\\ \ddot{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)\\ \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0})\end{pmatrix}\right] = \begin{bmatrix}\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{P}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}(\boldsymbol{\phi}_0) & \mathcal{A}(\boldsymbol{\phi}_0)\\ \mathcal{P}(\boldsymbol{\phi}_0) & \mathcal{P}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}(\boldsymbol{\phi}_0) & \mathcal{A}(\boldsymbol{\phi}_0)\\ \mathring{\mathcal{S}}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}(\boldsymbol{\phi}_0) & \mathcal{A}(\boldsymbol{\phi}_0)\\ \ddot{\mathcal{S}}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}(\boldsymbol{\phi}_0) & \mathcal{A}(\boldsymbol{\phi}_0)\\ \mathcal{A}(\boldsymbol{\phi}_0) & \mathcal{A}(\boldsymbol{\phi}_0) & \mathcal{A}(\boldsymbol{\phi}_0) & \mathcal{A}(\boldsymbol{\phi}_0) & \mathcal{B}(\boldsymbol{\phi}_0)\end{bmatrix}, \quad (5)$$

and

$$\lim_{T\to\infty} V\left[\sqrt{T}\begin{pmatrix}\hat{\boldsymbol{\theta}}_T(\boldsymbol{\eta}_0)-\boldsymbol{\theta}_0\\ \hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\\ \mathring{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\\ \ddot{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\\ \tilde{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\end{pmatrix}\right] = \begin{bmatrix}\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\\ \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{P}^{-1}(\boldsymbol{\phi}_0) & \mathcal{P}^{-1}(\boldsymbol{\phi}_0) & \mathcal{P}^{-1}(\boldsymbol{\phi}_0) & \mathcal{P}^{-1}(\boldsymbol{\phi}_0)\\ \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{P}^{-1}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0)\\ \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{P}^{-1}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0)\\ \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{P}^{-1}(\boldsymbol{\phi}_0) & \mathring{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0) & \ddot{\mathcal{S}}^{-1}(\boldsymbol{\phi}_0) & \mathcal{C}(\boldsymbol{\phi}_0)\end{bmatrix}. \quad (6)$$

Therefore, the five estimators have the Matryoshka doll covariance structure in (1), 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:

$$\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) \geq \mathcal{P}(\boldsymbol{\phi}_0) \geq \mathring{\mathcal{S}}(\boldsymbol{\phi}_0) \geq \ddot{\mathcal{S}}(\boldsymbol{\phi}_0) \geq \mathcal{C}^{-1}(\boldsymbol{\phi}_0). \quad (7)$$

We would like to emphasise that 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 appendix B 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 (7) suggests that they should focus on adjacent estimators. However, the number of degrees of freedom and the diverging behaviour of the estimators under the alternative hypothesis also play a very important role, as we discussed in section 2.1.

Nevertheless, we also saw in section 2.4 that there is no reason why researchers should choose just one such pair. 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.

Although in principle one could compare all five estimators, researchers may choose to disregard $\ddot{\boldsymbol{\theta}}_T - \tilde{\boldsymbol{\theta}}_T$ because both 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 $\mathring{\boldsymbol{\theta}}_T - \tilde{\boldsymbol{\theta}}_T$ 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 ones included in standard software packages even though increases in power might be obtained under the maintained assumption of *iid* innovations by comparing $\ddot{\boldsymbol{\theta}}_T$ instead of $\tilde{\boldsymbol{\theta}}_T$ with $\mathring{\boldsymbol{\theta}}_T$, $\hat{\boldsymbol{\theta}}_T$ and $\hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}})$. The next proposition provides detailed expressions for the different ingredients of the DWH test statistics described in Proposition 1 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* $\mathbf{y}_t$

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

and

$$\lim_{T\to\infty} V[\sqrt{T}\bar{\mathbf{s}}'_{\boldsymbol{\theta}T}(\hat{\boldsymbol{\theta}}_T, \mathbf{0})] = \mathcal{B}(\boldsymbol{\phi}_0) - \mathcal{A}(\boldsymbol{\phi}_0)\mathcal{P}^{-1}(\boldsymbol{\phi}_0)\mathcal{A}(\boldsymbol{\phi}_0),$$

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 different ingredients of the DWH test statistics in Proposition 1 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* $\mathbf{y}_t$

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

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

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, LM and Likelihood Ratio (LR) tests 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 A1 implies that in the spherically symmetric case

$$\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0) = \mathbf{W}_s(\boldsymbol{\phi}_0)\mathrm{M}_{sr}(\boldsymbol{\eta}_0),$$

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 $\mathrm{M}_{sr}(\boldsymbol{\eta}_0) \neq \mathbf{0}$, the Student t 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

$$\tilde{\boldsymbol{\eta}}'\mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)[\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)]^{-}\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\tilde{\boldsymbol{\eta}}, \tag{8}$$

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 will have zero power 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).

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

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

and

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

where

$$\bar{\mathbf{s}}_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2\boldsymbol{\eta}T}(\boldsymbol{\theta},\boldsymbol{\eta}) = \bar{\mathbf{s}}_{\boldsymbol{\theta}_1T}(\boldsymbol{\theta},\boldsymbol{\eta}) - [\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\,\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\eta}}(\boldsymbol{\phi}_0)]\begin{bmatrix}\mathcal{I}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\theta}_2\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\\ \mathcal{I}'_{\boldsymbol{\theta}_2\boldsymbol{\eta}}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\end{bmatrix}^{-1}\begin{bmatrix}\bar{\mathbf{s}}_{\boldsymbol{\theta}_2T}(\boldsymbol{\theta},\boldsymbol{\eta})\\ \bar{\mathbf{s}}_{\boldsymbol{\eta}T}(\boldsymbol{\theta},\boldsymbol{\eta})\end{bmatrix}, \quad (9)$$

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

while

$$\begin{aligned}\bar{\mathbf{s}}_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2T}(\boldsymbol{\theta},\mathbf{0}) &= \bar{\mathbf{s}}_{\boldsymbol{\theta}_1T}(\boldsymbol{\theta},\mathbf{0}) - \mathcal{A}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\mathcal{A}^{-1}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\bar{\mathbf{s}}_{\boldsymbol{\theta}_2T}(\boldsymbol{\theta},\mathbf{0}), \\ \mathcal{A}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0) &= [\mathcal{A}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0) - \mathcal{A}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\mathcal{A}^{-1}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\mathcal{A}'_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)]^{-1}.\end{aligned}$$

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* $\mathbf{y}_t$

$$\begin{aligned}\lim_{T\to\infty} V\{\sqrt{T}[\hat{\boldsymbol{\theta}}_{1T} - \hat{\boldsymbol{\theta}}_{1T}(\bar{\boldsymbol{\eta}})]\} &= \mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0) - \mathcal{I}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0), \\ \lim_{T\to\infty} V[\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2T}(\hat{\boldsymbol{\theta}}_T, \bar{\boldsymbol{\eta}})] &= [\mathcal{I}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)]^{-1}\mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)[\mathcal{I}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)]^{-1} - [\mathcal{I}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)]^{-1}\end{aligned}$$

and

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

where $\bar{\mathbf{s}}_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2\boldsymbol{\eta}T}(\boldsymbol{\theta},\boldsymbol{\eta})$ *is defined in (9),*

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

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 A1 and A2 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.[7] 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.

In what follows, we will simplify the presentation by concentrating on Wald versions of DWH tests, but all our results can be readily applied to their two asymptotically equivalent score versions by virtue of Propositions 1 and 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. In particular, there is one instance in which all the degrees of freedom are equal to 0, namely when the true conditional distribution is Gaussian. In that case, the PML estimator is obviously fully efficient, which implies that the other estimators of $\boldsymbol{\theta}$ must also be efficient in view of (7).[8] More formally,

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

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

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 $\kappa_0<\infty$, and $\mathbf{Z}_l(\boldsymbol{\phi}_0)\neq\mathbf{0}$, 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.

[7] Unfortunately, DWH tests that involve the Gaussian PMLE will not work properly with unbounded fourth moments, which violates one of the assumptions of Proposition A2.

[8] 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.

For practical purposes, this result indicates that one should not use any DWH test, including those in Propositions 6 and 7, to test for normality or when one suspects that the true distribution is Gaussian because the test statistics might become numerically unstable in that case. Unfortunately, one cannot simply compare the relevant covariance matrices because the estimated versions of $\mathcal{A}(\boldsymbol{\phi}_0)$, $\mathcal{B}(\boldsymbol{\phi}_0)$ and $\mathcal{P}(\boldsymbol{\phi}_0)$ will not coincide unless we evaluate them at $\boldsymbol{\eta} = \mathbf{0}$ using their theoretical expressions. In this regard, it is worth remembering that under normality, the unrestricted ML estimator $\hat{\eta}_T$ of the reciprocal of degrees of freedom of a multivariate Student t will be 0 approximately half the time only (see Fiorentini, Sentana and Calzolari (2003)).

There are other non-Gaussian distributions for which some but not all of the differences will be 0. In particular,

Proposition 11 *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 $N[(N+2)\kappa_0+2]$.*

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,[9] the density of a leptokurtic Kotz distribution has a pole at 0, which is a potential drawback from an empirical point of view.

For practical purposes, Proposition 11 implies that DWH tests based on $\sqrt{T}[\hat{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}})]$, $\sqrt{T}(\mathring{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T)$ and $\sqrt{T}[\mathring{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}})]$ might also become numerically unstable when the true distribution is a non-Gaussian Kotz. Although the problem would be easy to detect in comparing the restricted and unrestricted ML estimators if one uses the theoretical expressions for computing $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})$ and $\mathcal{P}(\boldsymbol{\phi})$, the same is not necessarily true in comparisons involving the spherically symmetric estimator, unless one exploits the properties of the Kotz distribution in estimating $\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)$ on the basis of (A27).

Proposition 11 provides a 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. However, it is not necessary:

Proposition 12 *If $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta}_0)$ $\mathbf{0}$, and $\mathbf{W}_s(\boldsymbol{\phi}_0) \neq \mathbf{0}$, then $\mathcal{P}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)$ only if $\mathrm{M}_{sr}(\boldsymbol{\eta}_0) = \mathbf{0}$.*

[9] 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 $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) = \mathcal{C}(\boldsymbol{\phi})$ in those models under exactly the same assumptions.

In this case, the DWH test considered in Proposition 7 will have no power to detect misspecification because it will converge to 0 asymptotically. This is confirmed by the fact that its non-centrality parameter in (8) will be identically 0 in those circumstances. Once again, though, numerical problems might arise if the estimated value of $\mathcal{I}_{\boldsymbol{\theta\eta}}$ is not identically $\mathbf{0}$.

Finally, Proposition 5 in Fiorentini and Sentana (2018) illustrates a very different reason for the DWH test considered in Proposition 6 to be degenerate. Specifically, they show that if one uses a Student t log-likelihood function for estimating $\boldsymbol{\theta}$ but the true distribution is such that $\kappa < 0$, then $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) = o_p(1)$.[10]

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 and 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*

$$\left.\begin{array}{c} \boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}_t(\boldsymbol{\vartheta}_c), \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \vartheta_i \boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_c) \end{array}\right\} \quad \forall t. \tag{10}$$

Expression (10) 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.

Such a reparametrisation is not unique, since we can always multiply the overall scale parameter ϑ_i by some scalar positive smooth function of $\boldsymbol{\vartheta}_c$, $k(\boldsymbol{\vartheta}_c)$ say, and divide $\boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_c)$ by the same function without violating (10) or redefining $\boldsymbol{\vartheta}_c$. Although it is by no means essential, a particularly convenient normalisation would be such that

$$E[\ln |\boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_c)| | \boldsymbol{\phi}_0] = k \quad \forall \boldsymbol{\vartheta}_c. \tag{11}$$

For the examples in section 3.1, reparametrisation 1 is as follows:

[10] As Fiorentini and Sentana (2018) explain, this result is valid not only for the Student t, but also for other log-likelihood functions in which the shape parameters are inequality restricted.

Univariate GARCH-M: We can write model (2) as

$$\left.\begin{array}{c} r_{Mt} = \mu_t(\boldsymbol{\vartheta}_c) + \vartheta_i^{1/2}\sigma_t^{\circ}(\boldsymbol{\vartheta}_c)\varepsilon_t^*, \\ \mu_t(\boldsymbol{\vartheta}_c) = \delta\sigma_t^{\circ}(\boldsymbol{\vartheta}_c), \\ \sigma_t^{\circ 2}(\boldsymbol{\vartheta}) = 1 + \gamma r_{Mt-1}^2 + \beta\sigma_{t-1}^{\circ 2}(\boldsymbol{\vartheta}_c). \end{array}\right\}$$

The transformed conditional mean and variance parameters are $\boldsymbol{\vartheta}_c' = (\delta, \gamma, \beta)$ and ϑ_i, whose relationship to the original ones is $\tau = \vartheta_i^{-1/2}\delta$, $\alpha = \vartheta_i\gamma$ and $\omega = \vartheta_i$.

Imposing (11) in this model is tricky because we need to obtain

$$E\left\{\ln\left[(1-\beta)^{-1} + \gamma\sum\nolimits_{j=0}^{\infty} r_{Mt-1-j}^2\right]\right\}$$

as a function of $\boldsymbol{\vartheta}_c$, which is probably best computed by numerically quadrature.

Multivariate market model: We can write model (3) as

$$\mathbf{r}_t = \mathbf{a} + \mathbf{b}r_{Mt} + \vartheta_i\boldsymbol{\Omega}^{\circ 1/2}(\boldsymbol{\varpi})\boldsymbol{\varepsilon}_t^*.$$

The transformed conditional mean and variance parameters are $\boldsymbol{\vartheta}_c' = (\mathbf{a}', \mathbf{b}', \boldsymbol{\varpi}')$ and ϑ_i, where $\boldsymbol{\varpi}$ contains $N(N+1)/2 - 1$ elements. Following Amengual and Sentana (2010), we can impose (11) with $\vartheta_i = |\boldsymbol{\Omega}|^{1/N}$ and $\boldsymbol{\Omega}^{\circ}(\boldsymbol{\varpi}) = \boldsymbol{\Omega}/|\boldsymbol{\Omega}|^{1/N}$ so as to ensure that $|\boldsymbol{\Omega}^{\circ}(\boldsymbol{\varpi})| = 1$ (see Appendix A.5 in Fiorentini and Sentana (2018) for explicit parametrisations that achieve this goal).

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

Proposition 13 *1. If $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta})$ and (10) 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

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

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

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

2. If in addition condition (11) holds at $\boldsymbol{\vartheta}_{c0}$, 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 spherically symmetric semiparametric 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 13 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 unrestricted semiparametric 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 and variance parameters* $\boldsymbol{\theta}$ *into an alternative set* $\boldsymbol{\psi} = (\boldsymbol{\psi}'_c, \boldsymbol{\psi}'_{im}, \boldsymbol{\psi}'_{ic},)'$*, where* $\boldsymbol{\psi}_{im}$ *is* $N \times 1$*,* $\boldsymbol{\psi}_{ic} = vech(\boldsymbol{\Psi}_{ic})$*,* $\boldsymbol{\Psi}_{ic}$ *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*

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

This parametrisations simply requires the pseudo-standardised residuals

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

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

Again, (14) is not unique, since it continues to hold with the same $\boldsymbol{\psi}_c$ if we replace $\boldsymbol{\Psi}_{ic}$ by $\mathbf{K}^{-1/2}(\boldsymbol{\psi}_c)\boldsymbol{\Psi}_{ic}\mathbf{K}^{-1/2\prime}(\boldsymbol{\psi}_c)$ and $\boldsymbol{\psi}_{im}$ by $\mathbf{K}^{-1/2}(\boldsymbol{\psi}_c)\boldsymbol{\psi}_{im} - \mathbf{l}(\boldsymbol{\psi}_c)$, and adjust $\boldsymbol{\mu}_t^{\diamond}(\boldsymbol{\psi}_c)$ and $\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)$ accordingly, where $\mathbf{l}(\boldsymbol{\psi}_c)$ and $\mathbf{K}(\boldsymbol{\psi}_c)$ are a $N \times 1$ vector and a $N \times N$ positive definite matrix of smooth functions of $\boldsymbol{\psi}_c$, respectively. A particularly convenient normalisation would be:

$$\left.\begin{array}{r} E\left[\partial\boldsymbol{\mu}_t^{\diamond\prime}(\boldsymbol{\psi}_c)/\partial\boldsymbol{\psi}_c \cdot \boldsymbol{\Sigma}_t^{\diamond -1/2}(\boldsymbol{\psi}_c)\middle|\, \boldsymbol{\phi}_0\right] = \mathbf{0} \\ E\left\{\partial vec[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)]/\partial\boldsymbol{\psi}_c \cdot \left[\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\boldsymbol{\psi}_c)\right]\middle|\, \boldsymbol{\phi}_0\right\} = \mathbf{0} \end{array}\right\}. \tag{16}$$

For the examples in section 3.1, reparametrisation 2 is as follows:

Univariate GARCH-M: We can write model (2) as

$$\left.\begin{array}{c} r_{Mt} = \psi_{im}\mu_t^{\diamond}(\boldsymbol{\psi}_c) + \psi_{ic}^{1/2}\sigma_t^{\diamond}(\boldsymbol{\psi}_c)\varepsilon_t^*, \\ \mu_t^{\diamond}(\boldsymbol{\psi}_c) = \sigma_t^{\diamond}(\boldsymbol{\psi}_c), \\ \sigma_t^{\diamond}(\boldsymbol{\psi}_c) = 1 + \gamma r_{Mt-1}^2 + \beta\sigma_{t-1}^{\diamond 2}(\boldsymbol{\vartheta}_c). \end{array}\right\}$$

The new conditional mean and variance parameters are $\boldsymbol{\psi}_c' = (\gamma, \beta)$, ψ_{im} and ψ_{ic}, whose relationship to the original ones is $\tau = \psi_{ic}^{1/2}\psi_{im}$, $\alpha = \psi_{ic}\gamma$ and $\omega = \psi_{ic}$.

Multivariate market model: We can write model (3) as

$$\mathbf{r}_t = \boldsymbol{\psi}_{im} + \mathbf{b}r_{Mt} + \boldsymbol{\Psi}_{ic}^{1/2}\boldsymbol{\varepsilon}_t^*.$$

The new conditional mean and variance parameters are $\boldsymbol{\psi}_c = \mathbf{b}$, $\boldsymbol{\psi}_{im}$ and $\boldsymbol{\psi}_{ic} = vech(\boldsymbol{\Psi}_{ic})$.

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

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

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

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

$$\boldsymbol{\psi}_{imT}(\boldsymbol{\psi}_c) = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\psi}_c), \tag{17}$$

$$\boldsymbol{\psi}_{icT}(\boldsymbol{\psi}_c) = vech\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\psi}_c) - \boldsymbol{\psi}_{imT}(\boldsymbol{\psi}_c)\right]\left[\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\psi}_c) - \boldsymbol{\psi}_{imT}(\boldsymbol{\psi}_c)\right]'\right\}. \tag{18}$$

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

2. If in addition condition (16) holds at $\boldsymbol{\psi}_{c0}$, then

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

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

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 14 for our purposes is that while the DWH test based on $\sqrt{T}(\tilde{\boldsymbol{\psi}}_T - \ddot{\boldsymbol{\psi}}_T)$ will have a maximum of $p - N(N+3)/2$ degrees of freedom, those based on $\sqrt{T}[\hat{\boldsymbol{\psi}}_T - \hat{\boldsymbol{\psi}}_T(\bar{\boldsymbol{\eta}})]$, $\sqrt{T}(\ddot{\boldsymbol{\psi}}_T - \hat{\boldsymbol{\psi}}_T)$ and $\sqrt{T}[\ddot{\boldsymbol{\psi}}_T - \hat{\boldsymbol{\psi}}_T(\bar{\boldsymbol{\eta}})]$ 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 $\sqrt{T}(\tilde{\boldsymbol{\psi}}_{cT} - \ddot{\boldsymbol{\psi}}_{cT})$, $\sqrt{T}[\hat{\boldsymbol{\psi}}_{iT} - \hat{\boldsymbol{\psi}}_{iT}(\bar{\boldsymbol{\eta}})]$, $\sqrt{T}(\ddot{\boldsymbol{\psi}}_{iT} - \hat{\boldsymbol{\psi}}_{iT})$ and $\sqrt{T}[\ddot{\boldsymbol{\psi}}_{iT} - \hat{\boldsymbol{\psi}}_{iT}(\bar{\boldsymbol{\eta}})]$ are identical to the analogous tests based on the entire vector $\boldsymbol{\psi}$, 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 in Fiorentini and Sentana (2018) we show that the parameters that are efficiently estimated by the semiparametric procedures continue to be consistently estimated by the parametric ML 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.

Specifically, Proposition 1 in Fiorentini and Sentana (2018) states that in the situation discussed in Proposition 13, $\boldsymbol{\vartheta}_c$ 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.[11] Therefore, we will maximise power in those circumstances if we base our DWH tests on the overall scale parameter ϑ_i exclusively.

Similarly, Proposition 3 in Fiorentini and Sentana (2018) states that in the context of Proposition 14, $\boldsymbol{\psi}_c$ 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 $\boldsymbol{\psi}_{im}$ and $\boldsymbol{\psi}_{ic}$ will be inconsistently estimated.[12] 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 $\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\psi}_c)$ in (15).

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 both cases, we evaluate the three asymptotically equivalent versions of the tests 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: We consider first the univariate GARCH-M model (2) analysed in Drost and Klaassen (1997). As we have already seen, 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

[11] See Figure 1 in Fiorentini and Sentana (2018) for the inconsistency in estimating ϑ_i in model (3) under distributional misspecification.

[12] See Figures 2A-B in Fiorentini and Sentana (2018) for the inconsistencies in estimating ψ_{im} and ψ_{ic} in model (2) under distributional misspecification.

$\boldsymbol{\psi}_c = (\beta, \gamma)'$, ψ_{im} and ψ_{ic}, where $\gamma = \alpha/\omega$, $\psi_{im} = \tau\omega^{1/2}$ and $\psi_{ic} = \omega$ (reparametrisation 2).

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 $\delta = \psi_{im} = 0.05$, $\gamma = 0.1$ and $\vartheta_i = \psi_{ic} = 1$. 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 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.

Table 2 contains the empirical rejections rates of the three pairwise tests in Propositions 8 and 9, together with the corresponding threesome 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,[13] 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 2 shows, with DWH3 having the best raw (i.e. non-size adjusted) power, and the LR ranking second. In turn, the threesome 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: In our second exercise, we consider the multivariate market model (3). In this case, we can write it in terms of $\boldsymbol{\vartheta}_c' = (\mathbf{a}', \mathbf{b}', \boldsymbol{\varpi}')$ and ϑ_i, with $\vartheta_i = |\boldsymbol{\Omega}|^{1/N}$ and $\boldsymbol{\Omega}^\circ(\boldsymbol{\varpi}) = \boldsymbol{\Omega}/|\boldsymbol{\Omega}|^{1/N}$ (reparametrisation 1) or $\boldsymbol{\psi}_c = \mathbf{b}$, $\boldsymbol{\psi}_{im} = \mathbf{a}$ and $\boldsymbol{\psi}_{ic} = vech(\boldsymbol{\Psi}_{ic}) = vech(\boldsymbol{\Omega})$ (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

[13] 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.

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 A.2, 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 $\boldsymbol{\Omega} = \mathbf{D}^{1/2}\mathbf{R}\mathbf{D}^{1/2}$, with $\mathbf{D} = 3.136\ \mathbf{I}_3$ and the off diagonal terms of the correlation matrix $\mathbf{R}$ equal to 0.3. Finally, 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 3 displays the Monte Carlo medians and interquartile ranges of the estimators for several representative parameters in addition to the global scale parameter $\vartheta_i = |\boldsymbol{\Omega}|^{1/N}$. 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 $vecd(\boldsymbol{\Omega}^{\circ})$, $vecl(\boldsymbol{\Omega}^{\circ})$, $vecd(\boldsymbol{\Omega})$ and $vecl(\boldsymbol{\Omega})$. 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.

Table 4 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 $vech(\boldsymbol{\Omega})$, 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.

5 Conclusions

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. We also explore several important issues related to the practical implementation of these 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 beyond univariate ARCH models. In particular, our results apply to conditionally homoskedastic, dynamic linear models such as VARs or multivariate regressions, which remain the workhorse in empirical macroeconomics and asset pricing contexts.

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. Nevertheless, we also consider two semiparametric estimators, one of which imposes the assumption that the standardised innovations follow a spherical distribution. For that reason, in the main text we particularise our results to log likelihood functions for spherical distributions, postponing the general case to Appendix B.

But to design powerful and reliable tests, we first need to study the consistency and efficiency properties of the different estimators involved. In particular, we need to figure out the rank of the difference between the corresponding asymptotic covariance matrices under the null of correct specification to select the right number of degrees of freedom. Consequently, we discuss several situations in which some of the estimators are equally efficient for some of the parameters. More specifically, we show that in the spherical case the SSP estimator is adaptive for all but one global scale parameter in an appropriate reparametrisation of the model. We also show that when the conditional distribution is not only leptokurtic or platykurtic but also potentially asymmetric, the general SP estimator is adaptive for a more restricted set of parameters in an alternative reparametrisation, which covers the slope coefficients of many conditionally homoskedastic multivariate regression models, including VARs. Importantly, we prove that both semiparametric estimators share a saddle point efficiency property: they are as inefficient as the Gaussian PMLE for the parameters that they cannot estimate adaptively.

We also take into account that some parameters remain consistently estimated under the alternative of incorrect distributional specification. Otherwise our tests will use up degrees of freedom without providing any power gains. In this regard, the results of Fiorentini and Sentana

(2018) 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.

Although our proposed tests apply to any multivariate conditionally heteroskedastic dynamic regression model, in our Monte Carlo experiments we study in detail two empirically relevant examples: a univariate Garch-M and a multivariate market model. We find that while many of our proposed tests work quite well, both in terms of size and power, 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. Nevertheless, we find very accurate sizes for one of the score versions of the test that compares the restricted and unrestricted ML estimators, which we show is not only invariant to reparametrisations but also numerically identical to the LM test of the null hypothesis that the shape parameter chosen by the researcher is correct. In addition, we find an almost null correlation between the statistics that compare the restricted and unrestricted ML estimators and the ones that compare the Gaussian PMLE with the unrestricted MLE, which confirms that the distribution of our proposed simultaneous tests can be approximated by the distribution of the sum of the two pairwise DWH tests.

An interesting extension of our Monte Carlo analysis would look at the power of our tests in models with time-varying shape parameters (see Fiorentini and Sentana (2010) for some limited results) or misspecified first and second moment dynamics.

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

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

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

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

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

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.

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

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.

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.

Fan, J., Qi, L. and Xiu, D. (2014): "Quasi maximum likelihood estimation of GARCH models with heavy-tailed likelihoods", *Journal of Business and Economic Statistics* 32, 193-198.

Fang, K.-T., Kotz, S. and Ng, K.-W. (1990), *Symmetric multivariate and related distributions*, Chapman and Hall.

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. (2010): "New testing approaches for mean-variance predictability", mimeo, CEMFI.

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.

Fiorentini, G. and Sentana, E. (2018): "Consistent non-Gaussian pseudo maximum likelihood estimators", CEMFI Working Paper 1802.

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., Lepage, G. and Zakoïan, J.-M. (2011): "Two-stage non Gaussian QML estimation of GARCH models and testing the efficiency of the Gaussian QMLE", *Journal of Econometrics* 165, 246-257.

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", available at http://dx.doi.org/10.2139/ssrn.2265027

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*, Cambridge University Press.

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

IHS Global Inc (2015): *Eviews 9: Command and Programming Reference.*

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.

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

Magnus, J.R. and Neudecker, H. (1988): *Matrix differential calculus with applications in Statistics and Econometrics*, 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.

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.

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.

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.

StataCorp LP (2015): *STATA Time Series Reference Manual Release 14.*

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

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

White, H. (1994): *Estimation, inference and specification analysis*, Econometric Society Monographs, Cambridge.

Appendix

A Proofs and auxiliary results

A.1 Some useful results on spherical distributions

A spherically symmetric random vector of dimension N, $\boldsymbol{\varepsilon}_t^\circ$, is fully characterised in Theorem 2.5 (iii) of Fang, Kotz and Ng (1990) as $\boldsymbol{\varepsilon}_t^\circ = e_t\mathbf{u}_t$, 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 $\boldsymbol{\varepsilon}_t^\circ$. 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 $\boldsymbol{\varepsilon}_t^\circ$ by setting $E(e_t^2) = N$, so that $E(\boldsymbol{\varepsilon}_t^\circ) = \mathbf{0}$, $V(\boldsymbol{\varepsilon}_t^\circ) = \mathbf{I}_N$. Specifically, if $\boldsymbol{\varepsilon}_t^\circ$ 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 $\boldsymbol{\varepsilon}_t^\circ$ 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 $V(\boldsymbol{\varepsilon}_t^\circ) = \mathbf{I}_N$ are given by

$$E(\boldsymbol{\varepsilon}_t^\circ\boldsymbol{\varepsilon}_t^{\circ\prime} \otimes \boldsymbol{\varepsilon}_t^\circ) = \mathbf{0}, \tag{A1}$$

$$E(\boldsymbol{\varepsilon}_t^\circ\boldsymbol{\varepsilon}_t^{\circ\prime} \otimes \boldsymbol{\varepsilon}_t^\circ\boldsymbol{\varepsilon}_t^{\circ\prime}) = E[vec(\boldsymbol{\varepsilon}_t^\circ\boldsymbol{\varepsilon}_t^{\circ\prime})vec'(\boldsymbol{\varepsilon}_t^\circ\boldsymbol{\varepsilon}_t^\circ)] = (\kappa_0+1)[(\mathbf{I}_{N^2}+\mathbf{K}_{NN})+vec\,(\mathbf{I}_N)\,vec'\,(\mathbf{I}_N)]. \tag{A2}$$

A.2 Standardised two component mixtures of multivariate normals

Consider the following mixture of two multivariate normals

$$\boldsymbol{\varepsilon}_t \sim \begin{cases} N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) & \text{with probability } \lambda, \\ N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) & \text{with probability } 1-\lambda. \end{cases}$$

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:

$$\begin{aligned} E(\boldsymbol{\varepsilon}_t) &= E[E(\boldsymbol{\varepsilon}_t|d_t)] = \lambda\boldsymbol{\mu}_1 + (1-\lambda)\boldsymbol{\mu}_2, \\ V(\boldsymbol{\varepsilon}_t) &= V[E(\boldsymbol{\varepsilon}_t|d_t)] + E[V(\boldsymbol{\varepsilon}_t|d_t)] = \lambda(1-\lambda)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' + \lambda\boldsymbol{\Sigma}_1 + (1-\lambda)\boldsymbol{\Sigma}_2. \end{aligned}$$

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.

A.3 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 $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + c(\boldsymbol{\eta}) + g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$, $d_t(\boldsymbol{\theta}) = \ln|\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})|$ 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 $g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$ are differentiable, then

$$\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) = \partial c(\boldsymbol{\eta})/\partial\boldsymbol{\eta} + \partial g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]/\partial\boldsymbol{\eta} = \mathbf{e}_{rt}(\boldsymbol{\phi}), \tag{A3}$$

while

$$\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) = \frac{\partial d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} + \frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial\varsigma}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = [\mathbf{Z}_{lt}(\boldsymbol{\theta}), \mathbf{Z}_{st}(\boldsymbol{\theta})]\begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{e}_{st}(\boldsymbol{\phi}) \end{bmatrix} = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}), \tag{A4}$$

where

$$\begin{aligned}
\partial d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} &= -\mathbf{Z}_{st}(\boldsymbol{\theta})vec(\mathbf{I}_N), \\
\partial\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} &= -2\{\mathbf{Z}_{lt}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) + \mathbf{Z}_{st}(\boldsymbol{\theta})vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right]\}, && \text{(A5)} \\
\mathbf{Z}_{lt}(\boldsymbol{\theta}) &= \partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\cdot\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}), \\
\mathbf{Z}_{st}(\boldsymbol{\theta}) &= \frac{1}{2}\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}\cdot[\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})], \\
\mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\eta}) &= \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}), && \text{(A6)} \\
\mathbf{e}_{st}(\boldsymbol{\theta}, \boldsymbol{\eta}) &= vec\left\{\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) - \mathbf{I}_N\right\}, && \text{(A7)}
\end{aligned}$$

and

$$\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] = -2\partial g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]/\partial\varsigma \tag{A8}$$

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 Appendix B, 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

$$
\begin{aligned}
\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\phi}) &= \frac{\partial^2 d_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} + \frac{\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}), \eta\right]}{(\partial \varsigma)^2}\frac{\partial \varsigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial \varsigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} + \frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}), \eta\right]}{\partial \varsigma}\frac{\partial^2 \varsigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}, && \text{(A9)} \\
\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\eta}t}(\boldsymbol{\phi}) &= \partial \varsigma_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \cdot \partial^2 g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]/\partial \varsigma \partial \boldsymbol{\eta}', && \text{(A10)} \\
\mathbf{h}_{\boldsymbol{\eta}\boldsymbol{\eta}t}(\boldsymbol{\phi}) &= \partial^2 c(\boldsymbol{\eta})/\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}' + \partial^2 g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]/\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}',
\end{aligned}
$$

where

$$
\partial^2 d_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}' = 2\mathbf{Z}_{st}(\boldsymbol{\theta})\mathbf{Z}_{st}'(\boldsymbol{\theta}) - \frac{1}{2}\left\{vec'\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right] \otimes \mathbf{I}_p\right\} \partial vec\left\{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial \boldsymbol{\theta}\right\}/\partial \boldsymbol{\theta}', \quad \text{(A11)}
$$

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

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 (see Kotz (1975)) and discrete scale mixture of normals, and Amengual, Fiorentini and Sentana (2013) for polynomial expansions).

A.4 Asymptotic distribution under correct specification

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 $\mathcal{I}(\boldsymbol{\phi}_0) = E[\mathcal{I}_t(\boldsymbol{\phi}_0)|\boldsymbol{\phi}_0]$,

$$
\begin{aligned}
\mathcal{I}_t(\boldsymbol{\phi}) &= V\left[\mathbf{s}_t(\boldsymbol{\phi})|I_{t-1}; \boldsymbol{\phi}\right] = \mathbf{Z}_t(\boldsymbol{\theta})\mathcal{M}(\boldsymbol{\phi})\mathbf{Z}_t'(\boldsymbol{\theta}) = -E\left[\mathbf{h}_t(\boldsymbol{\phi})|I_{t-1}; \boldsymbol{\phi}\right], \\
\mathbf{Z}_t(\boldsymbol{\theta}) &= \begin{pmatrix} \mathbf{Z}_{dt}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{pmatrix}, \qquad \text{(A12)}
\end{aligned}
$$

and $\mathcal{M}(\boldsymbol{\phi}) = V\left[\mathbf{e}_t(\boldsymbol{\phi})|\boldsymbol{\phi}\right]$. 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.

Proposition 1 in Fiorentini and Sentana (2007), which generalises Propositions 3 in Lange, Little and Taylor (1989), 1 in Fiorentini, Sentana and Calzolari (2003) and 5.2 in Hafner and Rombouts (2007), provides detailed expressions for $\mathcal{M}(\boldsymbol{\phi})$. We reproduce it here to facilitate its comparison to Proposition B5:

Proposition A1 *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*

$$\mathcal{M}(\boldsymbol{\eta}) = \begin{pmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{M}_{ss}(\boldsymbol{\eta}) & \mathcal{M}_{sr}(\boldsymbol{\eta}) \\ \mathbf{0} & \mathcal{M}'_{sr}(\boldsymbol{\eta}) & \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{pmatrix}, \tag{A13}$$

$$\mathcal{M}_{ll}(\boldsymbol{\eta}) = \mathrm{M}_{ll}(\boldsymbol{\eta})\mathbf{I}_N, \tag{A14}$$

$$\mathcal{M}_{ss}(\boldsymbol{\eta}) = \mathrm{M}_{ss}(\boldsymbol{\eta})\left(\mathbf{I}_{N^2} + \mathbf{K}_{NN}\right) + [\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1]vec(\mathbf{I}_N)vec'(\mathbf{I}_N), \tag{A15}$$

$$\mathcal{M}_{sr}(\boldsymbol{\eta}) = vec(\mathbf{I}_N)\mathrm{M}_{sr}(\boldsymbol{\eta}), \tag{A16}$$

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

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

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

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 $\mathrm{M}_{ss}(\kappa)$ 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 (2010, 2015) for closed-form expressions in the context of tests for univariate or multivariate conditional homoskedasticity, respectively.

A.5 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

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

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 $\boldsymbol{\rho}$:

Proposition A2 *Assume that the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied.*

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

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

   and

$$\mathcal{K}(\boldsymbol{\rho})=V[\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})|\,I_{t-1};\boldsymbol{\varphi}]=\begin{bmatrix}\mathbf{I}_N & \boldsymbol{\Phi}(\boldsymbol{\rho})\\ \boldsymbol{\Phi}(\boldsymbol{\rho}) & \boldsymbol{\Upsilon}(\boldsymbol{\rho})\end{bmatrix}, \tag{A17}$$

   where

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

   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 $\boldsymbol{\rho}=\mathbf{0}$ to denote normality.

2. *If $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}_0$ is i.i.d. $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\rho}_0)$ with $\kappa_0<\infty$, then (A17) reduces to*

$$\mathcal{K}(\kappa) = \begin{bmatrix}\mathbf{I}_N & \mathbf{0}\\ \mathbf{0} & (\kappa+1)(\mathbf{I}_{N^2}+\mathbf{K}_{NN})+\kappa vec(\mathbf{I}_N)vec'(\mathbf{I}_N)\end{bmatrix}, \tag{A18}$$

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

$$\kappa = E(\varsigma_t^2|\boldsymbol{\eta})/[N(N+2)]-1. \tag{A19}$$

A.6 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

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

where

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

and

$$\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) = \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0) - \mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0) \quad \text{(A20)}$$

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 $\boldsymbol{\eta}$, $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}_0)$. The residual score $\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)$ is sometimes called the unrestricted parametric efficient score of $\boldsymbol{\theta}$, and its variance, $\mathcal{P}(\boldsymbol{\phi}_0)$, the marginal information matrix of $\boldsymbol{\theta}$, or the unrestricted parametric efficiency bound. In this respect, note that $\mathcal{I}^{\theta\theta}(\boldsymbol{\phi}_0)$, which is the inverse of $\mathcal{P}(\boldsymbol{\phi}_0)$, coincides with the first block of $\mathcal{I}^{-1}(\boldsymbol{\phi}_0)$, and therefore it gives us the asymptotic variance of the feasible ML estimator, $\hat{\boldsymbol{\theta}}_T$. In contrast, $\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)$ would give us the asymptotic variance of the restricted ML estimator $\tilde{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}})$, provided of course that we could fix the shape parameters $\boldsymbol{\eta}$ to their true values.

In the spherically symmetric case, we can easily prove that (A20) and (4) reduce to

$$\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) = \mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\cdot\left[\mathrm{M}_{sr}(\boldsymbol{\eta}_0)\mathcal{M}^{-1}_{rr}(\boldsymbol{\eta}_0)\mathbf{e}_{rt}(\boldsymbol{\phi}_0)\right] \quad \text{(A21)}$$

and

$$\mathcal{P}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}'_s(\boldsymbol{\phi}_0)\cdot\left[\mathrm{M}_{sr}(\boldsymbol{\eta}_0)\mathcal{M}^{-1}_{rr}(\boldsymbol{\eta}_0)\mathrm{M}'_{sr}(\boldsymbol{\eta}_0)\right], \quad \text{(A22)}$$

respectively, where

$$\mathbf{W}_s(\boldsymbol{\phi}_0) = \mathbf{Z}_d(\boldsymbol{\phi}_0)[\mathbf{0}', vec'(\mathbf{I}_N)]' = E[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0][\mathbf{0}', vec'(\mathbf{I}_N)]'$$
$$= E\left\{\frac{1}{2}\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]/\partial\boldsymbol{\theta}\cdot vec[\boldsymbol{\Sigma}^{-1}_t(\boldsymbol{\theta}_0)]\middle|\boldsymbol{\phi}_0\right\} = E[\mathbf{W}_{st}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0] = -E\left\{\partial d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}|\,\boldsymbol{\phi}_0\right\}. \quad \text{(A23)}$$

It is worth noting that the last summand of (A20) 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 Lemma 3 in Appendix A.7. 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 nonparametric or semiparametric alternatives. In this section, we shall consider two such estimators.

The first one is fully nonparametric, and therefore replaces 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 generalises the univariate results of Gonzalez-Rivera and Drost (1999) and Propositions 3 and 4 in Hafner and Rombouts (2007) to multivariate models in which the conditional mean vector is not identically zero, describes the resulting semiparametric efficient score and the corresponding efficiency bound:

Proposition A3 *If $\boldsymbol{\varepsilon}_t^*|I_{t-1}; \boldsymbol{\theta}, \boldsymbol{\varrho}$ is i.i.d. $D(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\varrho})$ with density function $f(\boldsymbol{\varepsilon}_t^*; \boldsymbol{\varrho})$, where $\boldsymbol{\varrho}$ contains some shape parameters and $\boldsymbol{\varrho} = \mathbf{0}$ denotes normality, such that both its Fisher information matrix for location and scale,*

$$\mathcal{M}_{dd}(\boldsymbol{\theta}, \boldsymbol{\varrho}) = V\left[\mathbf{e}_{dt}(\boldsymbol{\theta}, \boldsymbol{\varrho})|I_{t-1}; \boldsymbol{\theta}, \boldsymbol{\varrho}\right]$$
$$= V\left\{\left[\begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho}) \\ \mathbf{e}_{st}(\boldsymbol{\theta}, \boldsymbol{\varrho}) \end{array}\right]\middle| \boldsymbol{\theta}, \boldsymbol{\varrho}\right\} = V\left\{\left[\begin{array}{c} -\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial \boldsymbol{\varepsilon}^* \\ -vec\left\{\mathbf{I}_N + \partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial \boldsymbol{\varepsilon}^* \cdot \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\} \end{array}\right]\middle| \boldsymbol{\theta}, \boldsymbol{\varrho}\right\}$$

and the matrix of third and fourth order central moments $\mathcal{K}(\boldsymbol{\varrho})$ are bounded, then the semiparametric efficient score will be given by:

$$\ddot{\mathbf{s}}_{\boldsymbol{\theta} t}(\boldsymbol{\phi}) = \mathbf{Z}_{dt}(\boldsymbol{\theta}, \boldsymbol{\varrho})\mathbf{e}_{dt}(\boldsymbol{\theta}, \boldsymbol{\varrho}_0) - \mathbf{Z}_d(\boldsymbol{\theta}, \boldsymbol{\varrho})\left[\mathbf{e}_{dt}(\boldsymbol{\theta}, \boldsymbol{\varrho}) - \mathcal{K}(0)\,\mathcal{K}^+(\boldsymbol{\varrho})\mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0})\right], \tag{A24}$$

while the semiparametric efficiency bound is

$$\ddot{\mathcal{S}}(\boldsymbol{\phi}) = \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}, \boldsymbol{\varrho}) - \mathbf{Z}_d(\boldsymbol{\theta}, \boldsymbol{\varrho})\left[\mathcal{M}_{dd}(\boldsymbol{\theta}, \boldsymbol{\varrho}) - \mathcal{K}(0)\,\mathcal{K}^+(\boldsymbol{\varrho})\mathcal{K}(0)\right]\mathbf{Z}_d'(\boldsymbol{\theta}, \boldsymbol{\varrho}), \tag{A25}$$

where $+$ denotes Moore-Penrose inverses and $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}, \boldsymbol{\varrho}) = E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{M}_{dd}(\boldsymbol{\theta}, \boldsymbol{\varrho})\mathbf{Z}_{dt}'(\boldsymbol{\theta})|\boldsymbol{\theta}, \boldsymbol{\varrho}\right]$.

In practice, however, $f(\boldsymbol{\varepsilon}_t^*; \boldsymbol{\varrho})$ has to be replaced by a nonparametric estimator, which suffers from the curse of dimensionality. For this reason, Hodgson and Vorkink (2003), Hafner and Rombouts (2007) and other authors have suggested to limit the admissible distributions to the class of spherically symmetric ones. As a consequence, the restricted tangent set in this case becomes 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 amends and extends Proposition 9 in Hafner and Rombouts (2007), provides the resulting spherically symmetric semiparametric efficient score and the corresponding efficiency bound:

Proposition A4 *When $\varepsilon_t^*|I_{t-1},\phi$ 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:*

$$\mathring{\mathbf{s}}_{\theta t}(\phi_0)=\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\phi_0)-\mathbf{W}_s(\phi_0)\left\{\left[\delta[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0]\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N}-1\right]-\frac{2}{(N+2)\kappa_0+2}\left[\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N}-1\right]\right\}, \quad (A26)$$

while the spherically symmetric semiparametric efficiency bound is

$$\mathring{\mathcal{S}}(\phi_0)=\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\phi_0)-\mathbf{W}_s(\phi_0)\mathbf{W}_s'(\phi_0)\cdot\left\{\left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta}_0)-1\right]-\frac{4}{N\left[(N+2)\kappa_0+2\right]}\right\}. \quad (A27)$$

Once again, $\mathbf{e}_{dt}(\phi)$ has to be replaced in practice by a semiparametric estimate obtained from the joint density of ε_t^*. However, the spherical symmetry assumption allows us to obtain such an estimate from a nonparametric estimate of the univariate density of ς_t, $h(\varsigma_t;\boldsymbol{\eta})$, avoiding in this way the curse of dimensionality.

A.7 Lemmata

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 $\mathrm{rank}\left[\partial\boldsymbol{\pi}'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\right]$ 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

$$\frac{\partial\bar{\mathbf{m}}_T'(\hat{\boldsymbol{\theta}}_T)}{\partial\boldsymbol{\theta}}\tilde{\mathcal{S}}_{\mathrm{m}T}\bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T)=\mathbf{0}, \quad (A28)$$

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

$$\frac{\partial\bar{\mathbb{m}}_T'(\hat{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\tilde{\mathcal{S}}_{\mathrm{m}T}\bar{\mathbb{m}}_T(\hat{\boldsymbol{\pi}}_T)=\frac{\partial\boldsymbol{\theta}'(\hat{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\frac{\partial\bar{\mathbf{m}}_T'[\boldsymbol{\theta}(\hat{\boldsymbol{\pi}}_T)]}{\partial\boldsymbol{\theta}}\tilde{\mathcal{S}}_{\mathrm{m}T}\bar{\mathbf{m}}_T[\boldsymbol{\theta}(\hat{\boldsymbol{\pi}}_T)]=\mathbf{0} \quad (A29)$$

by the chain rule for derivatives. Given that $\mathrm{rank}\left[\partial\boldsymbol{\theta}'(\boldsymbol{\pi})/\partial\boldsymbol{\pi}\right]$ 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 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 (A28) and (A29) by $\bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T) = \mathbf{0}$, and $\bar{\mathfrak{m}}_T(\hat{\boldsymbol{\pi}}_T) = \mathbf{0}$, 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*

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

as long as the required expectations are defined and bounded.

Proof. If we differentiate

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

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*

$$E\left\{\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\left[\mathbf{e}'_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}),\mathbf{e}'_{rt}(\boldsymbol{\theta},\boldsymbol{\varrho})\right]\Big|\, I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}\right\} = [\mathcal{K}(\mathbf{0})\,|\mathbf{0}]. \tag{A30}$$

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

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

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 $\mathbf{e}_t(\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0)$. □

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

$$\begin{aligned} \frac{\partial \bar{\mathbf{m}}_T'(\hat{\boldsymbol{\theta}}_T)}{\partial \boldsymbol{\theta}} \tilde{\mathcal{S}}_{mT} \bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T) &= \mathbf{0}, \quad \text{(A31)} \\ \frac{\partial \bar{\mathbf{n}}_T'(\tilde{\boldsymbol{\theta}}_T)}{\partial \boldsymbol{\theta}} \tilde{\mathcal{S}}_{nT} \bar{\mathbf{n}}_T(\tilde{\boldsymbol{\theta}}_T) &= \mathbf{0}. \quad \text{(A32)} \end{aligned}$$

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

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

and

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

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

The Wald version of the DWH test 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 (A31) and (A32) around $\boldsymbol{\theta}_0$ imply that

$$\begin{aligned} \sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) &= -\left[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right]^{-1}\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_0) + o_p(1), \\ \sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) &= -\left[\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n(\boldsymbol{\theta}_0)\mathcal{J}_n(\boldsymbol{\theta}_0)\right]^{-1}\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n\sqrt{T}\bar{\mathbf{n}}_T(\boldsymbol{\theta}_0) + o_p(1). \quad \text{(A33)} \end{aligned}$$

Therefore,

$$\begin{aligned} \sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) &= \left\{ \left[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right]^{-1}\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m \quad -\left[\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n\mathcal{J}_n(\boldsymbol{\theta}_0)\right]^{-1}\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n \right\} \\ &\quad \times \begin{bmatrix} \sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_0) \\ \sqrt{T}\bar{\mathbf{n}}_T(\boldsymbol{\theta}_0) \end{bmatrix} + o_p(1). \quad \text{(A34)} \end{aligned}$$

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

$$\mathcal{J}_m'(\boldsymbol{\theta}_n)\mathcal{S}_m E[\mathbf{m}_t(\boldsymbol{\theta}_n)] = \mathbf{0}. \quad \text{(A35)}$$

If we knew $\boldsymbol{\theta}_n$, it would be straightforward to test whether (A35) holds. But since we do not know it, we replace it by its consistent estimator $\tilde{\boldsymbol{\theta}}_T$, which satisfies (A32). 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 (A35) 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 write this expansion as

$$\mathcal{J}_m'(\tilde{\boldsymbol{\theta}}_T)\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T) = \mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_0) + \mathcal{J}_m(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) + o_p(1)$$
$$= \mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_0)$$
$$-[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)][\mathcal{J}_n(\boldsymbol{\theta}_0)\mathcal{S}_n(\boldsymbol{\theta}_0)\mathcal{J}_n'(\boldsymbol{\theta}_0)]^{-1}\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n\sqrt{T}\bar{\mathbf{n}}_T(\boldsymbol{\theta}_0) + o_p(1) \qquad \text{(A36)}$$

in view of (A33).

But a comparison between (A36) and (A34) makes clear that

$$\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) = \left[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right]^{-1}[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T)] + o_p(1), \qquad \text{(A37)}$$

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. □

Proposition 2

The Wald-type version of the Hausman test for the original parameters is computed in practice as

$$T(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T)'\boldsymbol{\Delta}_T^{\sim}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T), \qquad \text{(A38)}$$

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 can be computed as

$$T(\tilde{\boldsymbol{\pi}}_T - \hat{\boldsymbol{\pi}}_T)'\left[\frac{\partial\boldsymbol{\theta}(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}'}\right]^{-1}\boldsymbol{\Delta}_T^{\sim}\left[\frac{\partial\boldsymbol{\theta}'(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\right]^{-1}(\tilde{\boldsymbol{\pi}}_T - \hat{\boldsymbol{\pi}}_T). \qquad \text{(A39)}$$

Lemma 1 states the numerical invariance of GMM estimators and criterion functions to reparametrisations when the weighting matrix remains the same. In particular, it implies 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 (A38) and (A39) to differ. However, when the mapping from $\boldsymbol{\theta}$ to $\boldsymbol{\pi}$ is affine, the Jacobian of the inverse transformation is the constant matrix $\mathbf{A}^{-1}$, which yields

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

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 (A31). Therefore, we can compute the alternative DWH test in practice as

$$T\bar{\mathbf{m}}_T'(\tilde{\boldsymbol{\theta}}_T)\tilde{\mathcal{S}}_{mT}\frac{\partial\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T)}{\partial\boldsymbol{\theta}'}\boldsymbol{\Psi}_{mT}^{\sim}\frac{\partial\bar{\mathbf{m}}_T'(\tilde{\boldsymbol{\theta}}_T)}{\partial\boldsymbol{\theta}}\tilde{\mathcal{S}}_{mT}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T). \tag{A40}$$

Lemma 1 implies that $\bar{\mathfrak{m}}_T(\boldsymbol{\pi}) = \bar{\mathbf{m}}_T[\boldsymbol{\theta}(\boldsymbol{\pi})]$ 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 $\mathcal{J}_{\mathfrak{m}}'(\boldsymbol{\pi}_0)\mathcal{S}_m\sqrt{T}\bar{\mathfrak{m}}_T(\tilde{\boldsymbol{\pi}}_T)$ will be

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

As a consequence, it seems natural to use

$$\left[\frac{\partial\boldsymbol{\theta}(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}'}\right]^{-1}\boldsymbol{\Psi}_{mT}^{\sim}\left[\frac{\partial\boldsymbol{\theta}'(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\right]^{-1} \tag{A41}$$

as a consistent estimator of a generalised inverse of $\boldsymbol{\Psi}_{\mathfrak{m}}$, 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

$$T\bar{\mathfrak{m}}_T'(\tilde{\boldsymbol{\pi}}_T)\tilde{\mathcal{S}}_{mT}\frac{\partial\bar{\mathfrak{m}}_T(\tilde{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}'}\left[\frac{\partial\boldsymbol{\theta}(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}'}\right]^{-1}\boldsymbol{\Psi}_{mT}^{\sim}\left[\frac{\partial\boldsymbol{\theta}'(\dot{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\right]^{-1}\frac{\partial\bar{\mathfrak{m}}_T'(\tilde{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\tilde{\mathcal{S}}_{mT}\bar{\mathfrak{m}}_T(\tilde{\boldsymbol{\pi}}_T). \tag{A42}$$

But if we combine the chain rule for derivatives with the results in Lemma 1, we can immediately prove that

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

which in turn implies that

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

Therefore, (A40) and (A42) 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 (A41). □

Proposition 3

Again, we focus on the first result, as the second one is entirely analogous. Let us start from the asymptotic equivalence relationship (A37). Given that

$$\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)=\begin{bmatrix}\mathcal{J}_{1m}'(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{1m}(\boldsymbol{\theta}) & \mathcal{J}_{1m}'(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{2m}(\boldsymbol{\theta})\\ \mathcal{J}_{2m}'(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{1m}(\boldsymbol{\theta}) & \mathcal{J}_{2m}'(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{2m}(\boldsymbol{\theta})\end{bmatrix}$$

and

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

the application of the partitioned inverse formula yields

$$\sqrt{T}(\tilde{\boldsymbol{\theta}}_{1T}-\hat{\boldsymbol{\theta}}_{1T})=\left[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right]^{11}\bar{\mathbf{m}}_{1T}^{\perp}(\tilde{\boldsymbol{\theta}}_T,\mathcal{S}_m)$$

where

$$\left[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\right]^{11}=\begin{bmatrix}\mathcal{J}_{1m}'(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{1m}(\boldsymbol{\theta})\\ -\mathcal{J}_{1m}'(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{2m}(\boldsymbol{\theta})[\mathcal{J}_{2m}'(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{2m}(\boldsymbol{\theta})]^{-1}\mathcal{J}_{2m}'(\boldsymbol{\theta})\mathcal{S}_m\mathcal{J}_{1m}(\boldsymbol{\theta})\end{bmatrix}^{-1}.$$

Given that $[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)]^{11}$ 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 $\bar{\mathbf{m}}_{1T}^{\perp}(\tilde{\boldsymbol{\theta}}_T,\mathcal{S}_n)$.

Proposition 4

Given that

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

it follows immediately that

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

which in turn implies the asymptotic independence of non-overlapping test statistics. □

Proposition 5

Given that we have obtained all the information bounds in Propositions A1-A4, we simply need to compute the off-diagonal elements. Let us start with the first row. Straightforward manipulations imply that

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

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,

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

by virtue of the law of iterated expectations, together with (A59), (A60) and (A61). 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 $\boldsymbol{\varsigma}_t(\boldsymbol{\theta}_0)$ with bounded second moments that have zero conditional means and are conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$, 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 A3 that

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

by virtue of the law of iterated expectations, together with (A17) and (A30). 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

$$\begin{gathered}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\eta}}(\boldsymbol{\phi})\mathcal{I}^{-1}_{\boldsymbol{\eta\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\eta}}(\boldsymbol{\phi})\mathcal{I}^{-1}_{\boldsymbol{\eta\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\eta}}(\boldsymbol{\phi})\mathcal{I}^{-1}_{\boldsymbol{\eta\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\eta}}(\boldsymbol{\phi})\mathcal{I}^{-1}_{\boldsymbol{\eta\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\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)\end{gathered}$$

where we have used the fact that

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

and

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

by virtue of Lemma 3.

Similarly,

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

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

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

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

$$E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}] = \mathcal{A}(\boldsymbol{\phi}), \tag{A43}$$

which follows immediately from (A49) and the generalised information matrix equality.

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

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

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

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

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 A4 to show that

$$-\partial E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}]/\partial\boldsymbol{\theta} = E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)|\boldsymbol{\phi}] = E[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0)\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\mathbf{Z}'_{dt}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0]$$
$$-E\left\{\mathbf{W}_s(\boldsymbol{\phi}_0)\left[\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}-1\right]-\frac{2}{(N+2)\kappa_0+2}\left(\frac{\varsigma_t}{N}-1\right)\right]\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\mathbf{Z}'_{dt}(\boldsymbol{\theta}_0)\right|\boldsymbol{\phi}_0\Bigg\}$$
$$=\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)-\mathbf{W}_s(\boldsymbol{\phi}_0)E\left\{\left[\left\{\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}-1\right\}-\frac{2}{(N+2)\kappa_0+2}\left(\frac{\varsigma_t}{N}-1\right)\right]\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\right|\boldsymbol{\phi}_0\Bigg\}\mathbf{Z}_d(\boldsymbol{\theta}_0)$$
$$=\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\text{-}\mathbf{W}_s(\boldsymbol{\phi}_0)E\left[\left\{\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}\text{-}1\right]\text{-}\frac{2}{(N+2)\kappa_0+2}\left(\frac{\varsigma_t}{N}\text{-}1\right)\right\}\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}\text{-}1\right]\right|\boldsymbol{\phi}_0\Bigg]\mathbf{W}'_s(\boldsymbol{\phi}_0)$$
$$=\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)-\mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}'_s(\boldsymbol{\phi}_0)\cdot\left\{\left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta}_0)-1\right]-\frac{4}{N\left[(N+2)\kappa_0+2\right]}\right\}$$
$$=\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)=E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathring{\mathbf{s}}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}]. \tag{A44}$$

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

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

All the remaining asymptotic covariances are obtained analogously. □

Proposition 6

Given the efficiency of $\hat{\boldsymbol{\theta}}_T$ with respect to $\tilde{\boldsymbol{\theta}}_T$, it trivially 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

$$-\partial E[\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}]/\partial\boldsymbol{\theta}'=\mathcal{P}(\boldsymbol{\phi}) \tag{A45}$$

by the information matrix equality and

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

by its generalised analogue.

□

Proposition 7

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

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

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 (A45) 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. □

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. □

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. □

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 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 $\mathbf{Z}_l(\boldsymbol{\phi}_0) \neq \mathbf{0}$, $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}) - \ddot{\mathcal{S}}(\boldsymbol{\phi})$ 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 (A54).

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

$$\mathrm{M}_{ss}(\boldsymbol{\eta}) - \frac{1}{\kappa + 1} = \frac{N\kappa^2}{(\kappa+1)(2\kappa + N\kappa + 2)}$$

and

$$\mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1 + \frac{2\kappa}{(\kappa+1)[(N+2)\kappa + 2]} = -\frac{2\kappa^2}{(\kappa+1)(2\kappa + N\kappa + 2)}.$$

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

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

which can only be 0 under normality. □

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 (A54) 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. □

Proposition 12

Expression (A22) 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.

Proposition 13

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

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

$$\left.\begin{array}{c}\frac{1}{2}\{\partial vec'[\Sigma_t^\circ(\boldsymbol{\vartheta}_c)]/\partial\boldsymbol{\vartheta}_c\}[\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_c)\otimes\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_c)]\\ \frac{1}{2}\vartheta_i^{-1}vec'(\mathbf{I}_N)\end{array}\right\}=\left[\begin{array}{cc}\mathbf{Z}_{\boldsymbol{\vartheta}_c lt}(\boldsymbol{\vartheta}) & \mathbf{Z}_{\boldsymbol{\vartheta}_c st}(\boldsymbol{\vartheta})\\ 0 & \mathbf{Z}_{\vartheta_i st}(\boldsymbol{\vartheta})\end{array}\right]. \quad \text{(A46)}$$

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

$$\left[\begin{array}{c}\mathbf{s}_{\boldsymbol{\vartheta}_c t}(\boldsymbol{\vartheta},\boldsymbol{\eta})\\ s_{\vartheta_i t}(\boldsymbol{\vartheta},\boldsymbol{\eta})\end{array}\right]=\left[\begin{array}{c}\mathbf{Z}_{\boldsymbol{\vartheta}_c lt}(\boldsymbol{\vartheta})\mathbf{e}_{lt}(\boldsymbol{\vartheta},\boldsymbol{\eta})+\mathbf{Z}_{\boldsymbol{\vartheta}_c st}(\boldsymbol{\vartheta})\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta})\\ \mathbf{Z}_{\vartheta_i st}(\boldsymbol{\vartheta})\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta})\end{array}\right], \quad \text{(A47)}$$

where $\mathbf{e}_{lt}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ and $\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta})$ are given in (A6) and (A7), 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

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

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 A1.

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

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

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:

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

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 A4 and (A46) imply that the spherically symmetric semiparametric efficient score corresponding to ϑ_i will be given by

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

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

$$\mathbf{s}_{\vartheta_i t}(\boldsymbol{\vartheta},\mathbf{0}) = \frac{1}{2\vartheta_i}\left[\varsigma_t(\boldsymbol{\vartheta}) - N\right] \quad \text{(A48)}$$

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

$$E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}] = E[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathbf{Z}'_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}] = \mathcal{A}(\boldsymbol{\theta}) \tag{A49}$$

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 A4.

Tedious algebraic manipulations that exploit the block-triangularity of (A46) 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 (11) holds.

Finally, to prove Part 2b note that (A48) 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 (11) holds, the effect of estimating $\boldsymbol{\vartheta}_c$ becomes irrelevant. □

Proposition 14

We can directly work in terms of the $\boldsymbol{\psi}$ 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

$$\begin{bmatrix} \mathbf{s}_{\boldsymbol{\psi}_1 t}(\boldsymbol{\psi},\boldsymbol{\varrho}) \\ \mathbf{s}_{\boldsymbol{\psi}_{ic} t}(\boldsymbol{\psi},\boldsymbol{\varrho}) \\ \mathbf{s}_{\boldsymbol{\psi}_{im} t}(\boldsymbol{\psi},\boldsymbol{\varrho}) \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{\boldsymbol{\psi}_1 lt}(\boldsymbol{\psi})\mathbf{e}_{lt}(\boldsymbol{\psi},\boldsymbol{\varrho}) + \mathbf{Z}_{\boldsymbol{\psi}_1 st}(\boldsymbol{\psi})\mathbf{e}_{st}(\boldsymbol{\psi},\boldsymbol{\varrho}) \\ \mathbf{Z}_{\boldsymbol{\psi}_{ic} st}(\boldsymbol{\psi})\mathbf{e}_{st}(\boldsymbol{\psi},\boldsymbol{\varrho}) \\ \mathbf{Z}_{\boldsymbol{\psi}_{im} lt}(\boldsymbol{\psi})\mathbf{e}_{lt}(\boldsymbol{\psi},\boldsymbol{\varrho}) \end{bmatrix}, \tag{A50}$$

where

$$\left.\begin{array}{l} \mathbf{Z}_{\boldsymbol{\psi}_1 lt}(\boldsymbol{\psi}) = \left\{\partial\boldsymbol{\mu}_t^{\diamond\prime}(\boldsymbol{\psi}_1)/\partial\boldsymbol{\psi}_1 + \partial vec'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_1)]/\partial\boldsymbol{\psi}_1\cdot(\boldsymbol{\psi}_{im}\otimes\mathbf{I}_N)\right\}\boldsymbol{\Sigma}_t^{\diamond-1/2\prime}(\boldsymbol{\psi}_1)\boldsymbol{\Psi}_2^{-1/2\prime}, \\ \mathbf{Z}_{\boldsymbol{\psi}_1 st}(\boldsymbol{\psi}) = \partial vec'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_1)]/\partial\boldsymbol{\psi}_1\cdot[\boldsymbol{\Psi}_2^{1/2}\otimes\boldsymbol{\Sigma}_t^{\diamond-1/2\prime}(\boldsymbol{\psi}_1)\boldsymbol{\Psi}_2^{-1/2\prime}], \\ \mathbf{Z}_{\boldsymbol{\psi}_{im} lt}(\boldsymbol{\psi}) = \boldsymbol{\Psi}_2^{-1/2\prime} = \mathbf{Z}_{\boldsymbol{\psi}_{im} l}(\boldsymbol{\psi}), \\ \mathbf{Z}_{\boldsymbol{\psi}_{ic} st}(\boldsymbol{\psi}) = \partial vec'(\boldsymbol{\Psi}^{1/2})/\partial\boldsymbol{\psi}_{ic}\cdot(\mathbf{I}_N\otimes\boldsymbol{\Psi}_2^{-1/2\prime}) = \mathbf{Z}_{\boldsymbol{\psi}_{ic} s}(\boldsymbol{\psi}), \end{array}\right\} \tag{A51}$$

$\mathbf{e}_{lt}(\boldsymbol{\psi},\boldsymbol{\varrho})$ and $\mathbf{e}_{st}(\boldsymbol{\psi},\boldsymbol{\varrho})$ are given in (B65), with

$$\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}) = \boldsymbol{\Psi}_{ic}^{-1/2}\boldsymbol{\Sigma}_t^{\diamond-1/2}(\boldsymbol{\psi}_c)[\mathbf{y}_t - \boldsymbol{\mu}_t^{\diamond}(\boldsymbol{\psi}_c) - \boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)\boldsymbol{\psi}_{im}]. \tag{A52}$$

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

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

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

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

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

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

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

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

because both $\mathbf{Z}_{\boldsymbol{\psi}_{ic} s}(\boldsymbol{\psi})$ and $\mathbf{Z}_{\boldsymbol{\psi}_{im} l}(\boldsymbol{\psi})$ have full row rank when $\boldsymbol{\Psi}_{ic}$ has full rank in view of the discussion that follows expression (B74).

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

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

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

because $\mathbf{Z}_{\boldsymbol{\psi}_{ic} st}(\boldsymbol{\vartheta}) = \mathbf{Z}_{\boldsymbol{\psi}_{ic} s}(\boldsymbol{\vartheta})$ and $\mathbf{Z}_{\boldsymbol{\psi}_{im} lt}(\boldsymbol{\vartheta}) = \mathbf{Z}_{\boldsymbol{\psi}_{im} l}(\boldsymbol{\vartheta})$ $\forall t$. But if (18) and (17) hold, then the sample averages of $\mathbf{e}_{lt}[\boldsymbol{\psi}_c,\boldsymbol{\psi}_{ic}(\boldsymbol{\psi}_c),\boldsymbol{\psi}_{im}(\boldsymbol{\psi}_c);\mathbf{0}]$ and $\mathbf{e}_{st}[\boldsymbol{\psi}_c,\boldsymbol{\psi}_{ic}(\boldsymbol{\psi}_c),\boldsymbol{\psi}_{im}(\boldsymbol{\psi}_c);\mathbf{0}]$ will be 0, and the same is true of the semiparametric efficient score.

To prove Part 1c note that

$$\left[\begin{array}{c}\mathbf{s}_{\boldsymbol{\psi}_{ic} t}(\boldsymbol{\psi},\mathbf{0})\\ \mathbf{s}_{\boldsymbol{\psi}_{im} t}(\boldsymbol{\psi},\mathbf{0})\end{array}\right] = \left[\begin{array}{cc}\mathbf{0} & \mathbf{Z}_{\boldsymbol{\psi}_{ic} s}(\boldsymbol{\psi})\\ \mathbf{Z}_{\boldsymbol{\psi}_{im} l}(\boldsymbol{\psi}) & \mathbf{0}\end{array}\right]\left[\begin{array}{c}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi})\\ vec[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}) - \mathbf{I}_N]\end{array}\right], \quad \text{(A53)}$$

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 $\mathbf{Z}_{\boldsymbol{\psi}_{ic}s}(\boldsymbol{\psi})$ and $\mathbf{Z}_{\boldsymbol{\psi}_{im}l}(\boldsymbol{\psi})$ have full row rank when $\boldsymbol{\Psi}_{ic}$ has full rank. But as we saw in the proof of Proposition 5, that residual covariance matrix coincides with $\ddot{S}(\boldsymbol{\phi}_0)-\mathcal{A}(\boldsymbol{\theta})\mathcal{B}^{-1}(\boldsymbol{\phi})\mathcal{A}(\boldsymbol{\theta})$.

Tedious algebraic manipulations that exploit the block structure of (A51) and the constancy of $\mathbf{Z}_{\boldsymbol{\psi}_{ic}st}(\boldsymbol{\psi})$ and $\mathbf{Z}_{\boldsymbol{\psi}_{im}lt}(\boldsymbol{\psi})$ show that the different information matrices will be block diagonal when $\mathbf{Z}_{\boldsymbol{\psi}_c l}(\boldsymbol{\psi},\boldsymbol{\varrho})$ and $\mathbf{Z}_{\boldsymbol{\psi}_c s}(\boldsymbol{\psi},\boldsymbol{\varrho})$ are both 0. But those are precisely the necessary and sufficient conditions for $\mathbf{s}_{\boldsymbol{\psi}_c t}(\boldsymbol{\psi},\boldsymbol{\varrho})$ to be equal to $\mathbf{s}_{\boldsymbol{\psi}_c|\boldsymbol{\psi}_{ic},\boldsymbol{\psi}_{im}t}(\boldsymbol{\psi},\boldsymbol{\varrho})$, which is also guaranteed by (16). In this sense, please note that the reparametrisation of $\boldsymbol{\psi}_{ic}$ and $\boldsymbol{\psi}_{im}$ associated with (16) will be such that the Jacobian matrix of $vech[\mathbf{K}^{-1/2}(\boldsymbol{\psi}_c)\boldsymbol{\Psi}_2\mathbf{K}^{-1/2\prime}(\boldsymbol{\psi}_c)]$ and $\mathbf{K}^{-1/2}(\boldsymbol{\psi}_c)\boldsymbol{\psi}_{im}-\mathbf{l}(\boldsymbol{\psi}_c)$ with respect to $\boldsymbol{\psi}$ evaluated at the true values is equal to

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

Finally, to prove Part 2b simply note that (A53) implies the Gaussian PMLE will also satisfy (18) and (17). But since the asymptotic covariance matrices in both cases will be block-diagonal between $\boldsymbol{\psi}_c$ and $\boldsymbol{\psi}_i$ when (16) holds, the effect of estimating $\boldsymbol{\psi}_c$ becomes irrelevant. □

Proposition A1

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

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

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

$$h(\varsigma_t;\boldsymbol{\eta})=\frac{\pi^{N/2}}{\Gamma(N/2)}\varsigma_t^{N/2-1}\exp[c(\boldsymbol{\eta})+g(\varsigma_t,\boldsymbol{\eta})], \tag{A54}$$

whence

$$[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N)-1] = -\frac{2}{N}\left[1+\varsigma_t\cdot\partial\ln h(\varsigma_t;\boldsymbol{\eta})/\partial\varsigma\right]. \tag{A55}$$

On this basis, we can use Lemma 2 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

$$E\left[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N)-1|\boldsymbol{\eta}\right]=0 \tag{A56}$$

in view of (A55). 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 (A1), 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 (A2), (A55) and (A56).

Finally, it is clear from (A3) 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 $\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}_0)$ in (A9) as

$$
\begin{gathered}
-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)\}.
\end{gathered}
$$

Clearly, the first four lines have zero conditional expectation, and the same is true of the sixth line by virtue of (A1). As for the remaining terms, we can write them as

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

whose conditional expectation will be

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

As for $\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\eta}t}(\boldsymbol{\phi}_0)$, it follows from (A10) and (A5) 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 $\mathbf{Z}_{st}(\boldsymbol{\theta}_0)vec(\mathbf{I}_N)E[(\varsigma_t/N)\cdot\partial\delta(\varsigma_t,\boldsymbol{\eta}_0)/\partial\boldsymbol{\eta}'|\boldsymbol{\eta}]$. □

Proposition A2

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 (A17). Further, we will have that

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

for any spherical distribution, with ς_t and $\mathbf{u}_t$ both mutually and serially independent. Then (A18) follows from (A1) and (A2). 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 ($\boldsymbol{\varrho}_0 = \mathbf{0}$) by the (conditional) information matrix equality. $\square$

Proposition A3

It trivially follows from (A17) and (A30) 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 (A24), 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 (A24) as

$$[\mathbf{Z}_{dt}(\boldsymbol{\theta}) - \mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})]\,\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) + \mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})\mathcal{K}(0)\mathcal{K}^{+}(\boldsymbol{\varrho})\,\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0}),$$

then we can use the law of iterated expectations to show that the semiparametric efficient score (A24) 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

$$E\left[\begin{array}{c}\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathcal{K}(0)\mathcal{K}^{+}(\boldsymbol{\varrho})\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\}\\ \times\{\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})'\mathbf{Z}'_{dt}(\boldsymbol{\theta}) - \left[\mathbf{e}'_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathbf{e}'_{dt}(\boldsymbol{\theta},0)\mathcal{K}^{+}(\boldsymbol{\varrho})\mathcal{K}(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta},\boldsymbol{\varrho})\}\end{array}\middle|\,\boldsymbol{\theta},\boldsymbol{\varrho}\right]$$
$$= E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})\mathbf{e}'_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})\mathbf{Z}_{dt}(\boldsymbol{\theta})|\boldsymbol{\theta},\boldsymbol{\varrho}\right]$$
$$-E\left\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})\left[\mathbf{e}'_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathcal{K}^{+}(\boldsymbol{\varrho})\mathcal{K}(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta},\boldsymbol{\varrho})|\boldsymbol{\theta},\boldsymbol{\varrho}\right\}$$
$$-E\left\{\mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathcal{K}(0)\mathcal{K}^{+}(\boldsymbol{\varrho})\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\mathbf{e}_{dt}(\boldsymbol{\phi})'\mathbf{Z}'_{dt}(\boldsymbol{\theta})|\boldsymbol{\theta},\boldsymbol{\varrho}\right\}$$
$$+E\left\{\mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathcal{K}(0)\mathcal{K}^{+}(\boldsymbol{\varrho})\mathbf{e}_{dt}(\boldsymbol{\theta},0)\right]\left[\mathbf{e}'_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathbf{e}'_{dt}(\boldsymbol{\theta},0)\mathcal{K}^{+}(\boldsymbol{\varrho})\mathcal{K}(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta},\boldsymbol{\varrho})|\boldsymbol{\theta},\boldsymbol{\varrho}\right\}$$
$$= \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})\left[\mathcal{M}_{dd}(\boldsymbol{\varrho}) - \mathcal{K}(0)\mathcal{K}^{+}(\boldsymbol{\varrho})\mathcal{K}(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta},\boldsymbol{\varrho})$$

by virtue of (A17), (A30) and the law of iterated expectations. $\square$

Proposition A4

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

$$\begin{aligned}\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta}_0,\mathbf{0}) &= E\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\theta}_0,\mathbf{0})\\ \mathbf{e}_{st}(\boldsymbol{\theta}_0,\mathbf{0})\end{array}\middle|\,\varsigma_t;\boldsymbol{\phi}_0\right] = E\left\{\begin{array}{c}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\\ vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0) - \mathbf{I}_N\right]\end{array}\middle|\,\varsigma_t;\boldsymbol{\phi}_0\right\}\\ &= E\left[\begin{array}{c}\sqrt{\varsigma_t}\mathbf{u}_t\\ vec(\varsigma_t\mathbf{u}_t\mathbf{u}'_t - \mathbf{I}_N)\end{array}\middle|\,\varsigma_t\right] = \left(\frac{\varsigma_t}{N} - 1\right)\left[\begin{array}{c}\mathbf{0}\\ vec(\mathbf{I}_N)\end{array}\right],\end{aligned} \tag{A57}$$

and

$$\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}_0) = E\left[\begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\phi}_0) \\ \mathbf{e}_{st}(\boldsymbol{\phi}_0) \end{array}\middle| \varsigma_t;\boldsymbol{\phi}_0\right]$$

$$= E\left\{\begin{array}{c} \delta[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0) \\ vec\left[\delta[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)-\mathbf{I}_N\right] \end{array}\middle| \varsigma_t;\boldsymbol{\phi}_0\right\}$$

$$= E\left\{\begin{array}{c} \delta(\varsigma_t,\boldsymbol{\eta}_0)\sqrt{\varsigma_t}\mathbf{u}_t \\ vec[\delta(\varsigma_t,\boldsymbol{\eta}_0)\varsigma_t\mathbf{u}_t\mathbf{u}_t'-\mathbf{I}_N] \end{array}\middle| \varsigma_t\right\} = \left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}\text{-}1\right]\left[\begin{array}{c} \mathbf{0} \\ vec(\mathbf{I}_N) \end{array}\right], \tag{A58}$$

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 (A19) to show that

$$E\left[\left(\frac{\varsigma_t}{N}-1\right)^2\middle|\boldsymbol{\eta}\right] = \frac{(N+2)\kappa+2}{N}, \tag{A59}$$

so that

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

We can also use Lemma 2 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 (A55) and (A56), we will have that for any spherically symmetric distribution

$$E\left\{\left(\frac{\varsigma_t}{N}-1\right)\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}-1\right]\middle|\boldsymbol{\eta}\right\} = \frac{2}{N}, \tag{A60}$$

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 A1 immediately implies that

$$E\left\{\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}-1\right]^2\middle|\boldsymbol{\eta}\right\} = \frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta})-1. \tag{A61}$$

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

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

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

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

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 (A26) 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 (A26) 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}\right|\boldsymbol{\phi}\Bigg]\\ &=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}\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})-1\right]-\frac{4}{N\left[(N+2)\kappa+2\right]}\right\}\end{aligned}$$

by virtue of the law of iterated expectations. □

B The general case of non-spherical distributions

B.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

$$\begin{aligned}\frac{\partial\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}&=-\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}-[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})]\frac{\partial vec[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'}\\&=-\{\mathbf{Z}_{lt}'(\boldsymbol{\theta})+[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\},\end{aligned}\tag{B62}$$

where

$$\left.\begin{aligned}\mathbf{Z}_{lt}(\boldsymbol{\theta})=\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\cdot\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})\\\mathbf{Z}_{st}(\boldsymbol{\theta})=\partial vec'[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}\cdot[\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})]\end{aligned}\right\},\tag{B63}$$

it follows that

$$\begin{aligned}\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})&=[\mathbf{Z}_{lt}(\boldsymbol{\theta}),\mathbf{Z}_{st}(\boldsymbol{\theta})]\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\\mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right]=\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}),\\\mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi})&=\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varrho}=\mathbf{e}_{rt}(\boldsymbol{\phi}),\end{aligned}\tag{B64}$$

with

$$\mathbf{e}_{dt}(\boldsymbol{\phi})=\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\\mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right]=\left[\begin{array}{c}-\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*,\\-vec\left\{\mathbf{I}_N+\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\cdot\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\}\end{array}\right].\tag{B65}$$

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 (B62) implies that

$$\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} = -\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\} \quad \text{(B66)}$$

because

$$d\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho}) = -d\{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial \boldsymbol{\varepsilon}^*\}. \quad \text{(B67)}$$

In turn,

$$d\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho}) = -dvec\left[\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \cdot \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right]$$
$$= -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] d\left\{\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right\} - \left\{\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right\} d\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \quad \text{(B68)}$$

implies that

$$\frac{\partial \mathbf{e}_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} = \frac{\partial \mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} = -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} - \left\{\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right\} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$$
$$\left\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} + \left[\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right]\right\} \{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\}. \quad \text{(B69)}$$

Finally, (B67) and (B68) 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

$$\begin{aligned} \mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) &= \mathbf{Z}_{lt}(\boldsymbol{\theta}) \frac{\partial \mathbf{e}_{lt}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} + \mathbf{Z}_{st}(\boldsymbol{\theta}) \frac{\partial \mathbf{e}_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} \\ &\quad + \left[\mathbf{e}_{lt}'(\boldsymbol{\phi}) \otimes \mathbf{I}_p\right] \frac{\partial vec[\mathbf{Z}_{lt}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'} + \left[\mathbf{e}_{st}'(\boldsymbol{\phi}) \otimes \mathbf{I}_p\right] \frac{\partial vec[\mathbf{Z}_{st}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'}, && \text{(B70)} \\ \mathbf{h}_{\boldsymbol{\theta\varrho}t}(\boldsymbol{\phi}) &= \mathbf{Z}_{lt}(\boldsymbol{\theta}) \partial \mathbf{e}_{lt}(\boldsymbol{\phi})/\partial \boldsymbol{\varrho}' + \mathbf{Z}_{st}(\boldsymbol{\theta}) \partial \mathbf{e}_{st}(\boldsymbol{\phi})/\partial \boldsymbol{\varrho}', && \text{(B71)} \\ \mathbf{h}_{\boldsymbol{\varrho\varrho}t}(\boldsymbol{\phi}) &= \partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial \boldsymbol{\varrho} \partial \boldsymbol{\varrho}'. \end{aligned}$$

In this regard, note that since (B67) and (B68) also imply that

$$\begin{aligned} \partial \mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial \boldsymbol{\varrho}' &= -\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varrho}', && \text{(B72)} \\ \partial \mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial \boldsymbol{\varrho}' &= -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varrho}', && \text{(B73)} \end{aligned}$$

respectively, it is clear that

$$\begin{aligned} \mathbf{Z}_{lt}(\boldsymbol{\theta}) \frac{\partial \mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial \boldsymbol{\varrho}'} + \mathbf{Z}_{st}(\boldsymbol{\theta}) \frac{\partial \mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial \boldsymbol{\varrho}'} &= -\{\mathbf{Z}_{lt}(\boldsymbol{\theta}) + \mathbf{Z}_{st}(\boldsymbol{\theta})[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\} \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varrho}'} \\ &= \frac{\partial \boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho})}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varrho}'} \end{aligned}$$

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

$$dvech(\boldsymbol{\Sigma}_t) = [\mathbf{L}_N(\boldsymbol{\Sigma}_t^{1/2} \otimes \mathbf{I}_N)\mathbf{L}_N' + \mathbf{L}_N(\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{1/2})\mathbf{K}_{NN}\mathbf{L}_N']dvech(\boldsymbol{\Sigma}_t^{1/2}), \tag{B74}$$

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

$$\frac{\partial vec'[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}}[\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})] = \frac{1}{2}\frac{\partial vech'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}\mathbf{F}_t\mathbf{L}_N(\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{-1/2})$$

is as follows:

1. From the expression for $\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}$ we can readily obtain $\partial vech'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}$ 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

$$\mathbf{L}_N(\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{1/2}) = \mathbf{L}_N \begin{pmatrix} \boldsymbol{\Sigma}_t^{-1/2} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_t^{-1/2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_t^{-1/2} \end{pmatrix}$$

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 $\partial vech'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}\cdot\mathbf{F}_t$.

B.2 Additional results

Propositions 10.1 and 14, A2.1, A3,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, 13, A2.2 and A4, are specific to the spherically symmetric case. Therefore, the only proposition that really requires a proper generalisation is Proposition A1.

Proposition B5 *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

$$\begin{aligned}
\mathcal{M}_{ll}(\boldsymbol{\varrho}) &= V[\mathbf{e}_{lt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = E\left[\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\right|\boldsymbol{\varrho}\right],\\
\mathcal{M}_{ls}(\boldsymbol{\varrho}) &= E[\mathbf{e}_{lt}(\boldsymbol{\phi})\mathbf{e}_{st}(\boldsymbol{\phi})'|\boldsymbol{\phi}] = E\left[\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\cdot(\boldsymbol{\varepsilon}_t'^*\otimes\mathbf{I}_N)\right|\boldsymbol{\varrho}\right],\\
\mathcal{M}_{ss}(\boldsymbol{\varrho}) &= V[\mathbf{e}_{st}(\boldsymbol{\phi})|\boldsymbol{\phi}] = E\left[(\boldsymbol{\varepsilon}_t^*\otimes\mathbf{I}_N)\cdot\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\cdot(\boldsymbol{\varepsilon}_t^{*\prime}\otimes\mathbf{I}_N)|\boldsymbol{\varrho}\right] - \mathbf{K}_{NN},\\
\mathcal{M}_{lr}(\boldsymbol{\varrho}) &= E[\mathbf{e}_{lt}(\boldsymbol{\phi})\mathbf{e}_{rt}'(\boldsymbol{\phi})|\boldsymbol{\phi}] = -E\left[\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\varrho}\right],\\
\mathcal{M}_{sr}(\boldsymbol{\varrho}) &= E[\mathbf{e}_{st}(\boldsymbol{\phi})\mathbf{e}_{rt}'(\boldsymbol{\phi})|\boldsymbol{\phi}] = -E\left[(\boldsymbol{\varepsilon}_t^*\otimes\mathbf{I}_N)\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\varrho}\right],
\end{aligned}$$

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 (B64) 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

$$E[\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\phi})|z_t, I_{t-1}; \boldsymbol{\phi}] = \mathbf{Z}_{lt}(\boldsymbol{\theta})E\left[\left.\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'}\right| z_t, I_{t-1}; \boldsymbol{\phi}\right] + \mathbf{Z}_{st}(\boldsymbol{\theta})E\left[\left.\frac{\partial \mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'}\right| z_t, I_{t-1}; \boldsymbol{\phi}\right]$$

because

$$E\left[\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})|z_t, I_{t-1}; \boldsymbol{\phi}\right] = -E\left[\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial \boldsymbol{\varepsilon}^* | z_t, I_{t-1}; \boldsymbol{\phi}\right] = \mathbf{0} \tag{B75}$$

and

$$E\left[\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})|z_t, I_{t-1}; \boldsymbol{\phi}\right] = -E\left[vec\{\mathbf{I}_N + \partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial \boldsymbol{\varepsilon}^* \cdot \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\}|z_t, I_{t-1}; \boldsymbol{\phi}\right] = \mathbf{0}. \tag{B76}$$

Expression (B66) 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 (B69) leads to

$$E\left[\left.\frac{\partial \mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'}\right| z_t, I_{t-1}; \boldsymbol{\phi}\right] = E\left[\left\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} + \left[\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right]\right\}\right.$$
$$\times\{\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}'_{st}(\boldsymbol{\theta})\}\Big|\, z_t, I_{t-1}; \boldsymbol{\phi}\Big] = E\left[\left.[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}}\right| \boldsymbol{\phi}\right]\mathbf{Z}'_{lt}(\boldsymbol{\theta})$$
$$+E\left[\left.[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}}[\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\right| z_t, I_{t-1}; \boldsymbol{\phi}\right]\mathbf{Z}'_{st}(\boldsymbol{\theta}) - \mathbf{K}_{NN}\mathbf{Z}'_{st}(\boldsymbol{\theta})$$

because of (B75) and (B76), which in turn implies

$$
\begin{aligned}
& E\left\{\left.\left[\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right][\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\right| z_t, I_{t-1};\boldsymbol{\phi}\right\} \\
= \; & \mathbf{K}_{NN} E\left\{\left.\mathbf{K}_{NN}\left[\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right][\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\right| z_t, I_{t-1};\boldsymbol{\phi}\right\} \\
= \; & \mathbf{K}_{NN} E\left\{\left.\left[\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \otimes \mathbf{I}_N\right][\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\right| z_t, I_{t-1};\boldsymbol{\phi}\right\} \\
= \; & \mathbf{K}_{NN} E\left\{\left.\left[\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N\right]\right| z_t, I_{t-1};\boldsymbol{\phi}\right\} = -\mathbf{K}_{NN}
\end{aligned}
$$

in view of Theorem 3.1 in Magnus (1988).

As a result, the information matrix equality implies that

$$
\begin{aligned}
\mathcal{M}_{ll}(\boldsymbol{\varrho}) &= E\left\{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}\big|\, \boldsymbol{\phi}\right\} \\
\mathcal{M}_{ls}(\boldsymbol{\varrho}) &= E\left\{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime} \cdot [\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\big|\, \boldsymbol{\phi}\right\} \\
\mathcal{M}_{ss}(\boldsymbol{\varrho}) &= E\left\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime} \cdot [\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\big|\, \boldsymbol{\phi}\right\} - \mathbf{K}_{NN}
\end{aligned}
$$

Similarly, equation (B71) 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 (B72) and (B73) 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. □

TABLE 1: Univariate GARCH-M: Parameter estimators.

Parameter		β	γ	δ, ψ_{im}	ϑ_i, ψ_{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 2: 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 $(\psi_{im},\psi_{ic})@(\tilde{\boldsymbol{\theta}}_T,\breve{\eta}_T)$			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 $(\psi_{im},\psi_{ic})@(\tilde{\boldsymbol{\theta}}_T,\breve{\eta}_T)$			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 $(\psi_{im},\psi_{ic})@(\tilde{\boldsymbol{\theta}}_T,\breve{\eta}_T)$			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 $(\psi_{im},\psi_{ic})@(\tilde{\boldsymbol{\theta}}_T,\breve{\eta}_T)$			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 3: 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	
Student t_{12}	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)	
	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)	
Student t_8	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)	
	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)	
DSMN(0.2,0.1)	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)	
	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)	
DLSMN(0.2,0.1,0.5)	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)	
	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.

TABLE 4: Multivariate market model: Empirical rejection rates.

Student t_{12}

	RML=UML				UML=PML			UML=PML			RML=UML & UML=PML		
	$\vartheta_i@(\bar{\boldsymbol{\theta}}_T, \bar{\eta})$				$\vartheta_i@(\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T))$			$(\mathbf{a}, vech(\mathbf{\Omega}))@(\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T)$			$\vartheta_i@(\bar{\boldsymbol{\theta}}_T, \bar{\eta})$		
%	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				UML=PML			UML=PML			RML=UML & UML=PML		
	$\vartheta_i@(\bar{\boldsymbol{\theta}}_T, \bar{\eta})$				$\vartheta_i@(\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T))$			$(\mathbf{a}, vech(\mathbf{\Omega}))@(\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T)$			$\vartheta_i@(\bar{\boldsymbol{\theta}}_T, \bar{\eta})$		
%	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				UML=PML			UML=PML			RML=UML & UML=PML		
	$\vartheta_i@(\bar{\boldsymbol{\theta}}_T, \bar{\eta})$				$\vartheta_i@(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T)$			$(\mathbf{a}, vech(\mathbf{\Omega}))@(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T)$			$\vartheta_i@(\bar{\boldsymbol{\theta}}_T, \bar{\eta}_T)$		
%	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				UML=PML			UML=PML			RML=UML & UML=PML		
	$\vartheta_i@(\bar{\boldsymbol{\theta}}_T, \bar{\eta})$				$\vartheta_i@(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T)$			$(\mathbf{a}, vech(\mathbf{\Omega}))@(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T)$			$\vartheta_i@(\bar{\boldsymbol{\theta}}_T, \bar{\eta}_T)$		
%	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.