Quantitative Economics 12 (2021), 683–742 1759-7331/20210683

Specification tests for non-Gaussian maximum likelihood estimators

GABRIELE FIORENTINI
Department of Statistics, Informatics and Applications, Università di Firenze and RCEA

ENRIQUE SENTANA
CEMFI

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

KEYWORDS. Durbin–Wu–Hausman tests, partial adaptivity, semiparametric estimators, singular covariance matrices, uncertainty and the business cycle.

JEL CLASSIFICATION. C12, C14, C22, C32, C52.

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

Gabriele Fiorentini: gabriele.fiorentini@unifi.it
Enrique Sentana: sentana@cemfi.es
This paper draws heavily on Fiorentini and Sentana (2007). In addition to those explicitly acknowledged there, we would like to thank Dante Amengual, Christian Bontemps and Luca Fanelli for useful comments and suggestions, as well as audiences at CREST, Konstanz, Tokyo, UCLA, UCSD, the University of Liverpool 6th 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 70th ESEM (Lisbon, August 2017), the Bilgi CEFIS conference on "Advances in Econometrics Methods" (Instanbul, March 2018), the Financial Engineering and Risk Management Conference (Shanghai, June 2018), the University of Kent "50 Years of Econometrics at Keynes College" Conference (Canterbury, September 2018), and the TSE Financial Econometrics Conference (Toulouse, May 2019). The coeditor and two anonymous referees have provided very useful feedback. Of course, the usual caveat applies. The second author acknowledges financial support from the Spanish Ministry of Economy, Industry and Competitiveness through grant ECO 2017-89689 and the Santander CEMFI Research Chair.

Expected Shortfall (see Adrian and Brunnermeier (2016) and Acharya, Pedersen, Philippon, and Richardson (2017), respectively), which could be severely mismeasured by assuming normality. Given that empirical researchers are interested in those risk measures for several probability levels, they often specify a parametric leptokurtic distribution, which then they use to estimate their models by maximum likelihood (ML).

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

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

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

The statistical and econometric literature on model specification is huge. In this paper, our focus is the adequacy of the conditional distribution under the maintained assumption that the rest of the model is correctly specified. Even so, there are various ways of assessing it. One possibility is to nest the assumed distribution within a more flexible parametric family in order to conduct a Lagrange Multiplier (LM) test of the nesting restrictions. This is the approach in Mencía and Sentana (2012), who use the generalized hyperbolic family as an instrumental nesting distribution for the multivariate Student t. In contrast, other specification tests do not consider an explicit alternative hypothesis. A case in point are consistent tests based on the difference between the theoretical and empirical cumulative distribution functions of the innovations (Bai (2003) and Bai and

Zhihong (2008)) or their characteristic functions (Bierens and Wang (2012) and Amengual, Carrasco, and Sentana (2020)). An alternative procedure would be the information matrix test of White (1982), which compares some or all of the elements of the expected Hessian and the variance of the score. White (1987) also proposed the application of Newey's (1985) conditional moment test to assesses the martingale difference property of the scores under correct specification. Finally, the general class of moment tests in Newey (1985) and Tauchen (1985) could also be entertained, as Bontemps and Meddahi (2012) illustrate.

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

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

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

In Fiorentini and Sentana (2019), we characterized the mean and variance parameters that distributionally misspecified ML estimators can consistently estimate, and provided simple closed-form consistent estimators for the rest. One of the most interesting results that we obtain in this paper is that the parameters that continue to be consistently estimated by the parametric estimators under distributional misspecification are those which are efficiently estimated by the semiparametric procedures. In contrast, the

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

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 nonnormality of many macroeconomic variables (see Lanne, Meitz, and Saikkonen (2017) and the references therein for recent examples of univariate and multivariate time series models with non-Gaussian innovations). Nevertheless, structural models pose some additional inference challenges, which we discuss separately. Obviously, our approach also applies in cross-sectional models with exogenous regressors, as well as in static ones.

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

2. Durbin–Wu–Hausman tests

2.1 *Wald and score versions*

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

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

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

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

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

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

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

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

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

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

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

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

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

An intuitive way of reinterpreting the asymptotic equivalence between the original DWH test in (1) and the two alternative score versions on the right-hand sides of (3) and (4) is to think of the latter as original DWH tests based on two convenient reparametrizations of $\boldsymbol{\theta}$ obtained through the population version of the first order conditions that give rise to each estimator, namely $\boldsymbol{\pi}_m(\boldsymbol{\theta})=\mathcal{J}_m'(\boldsymbol{\theta})\mathcal{S}_mE[\mathbf{m}_t(\boldsymbol{\theta})]$ and

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

$\boldsymbol{\pi}_n(\boldsymbol{\theta}) = \mathcal{J}_n'(\boldsymbol{\theta})\mathcal{S}_n E[\mathbf{n}_t(\boldsymbol{\theta})]$. 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.[3] The same arguments also allow us to loosely interpret the score versions of the DWH tests as distance metric tests of those moment conditions, as they compare the values of the GMM criteria at the estimator which sets those exactly identified moments to 0 with their values at the alternative estimator. We will discuss more formal links to the classical Wald, Likelihood Ratio (LR) and LM tests in a likelihood context in Section 3.4.

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

2.2 *Numerical invariance to reparametrizations*

Suppose we decide to work with an alternative parametrization 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 $\boldsymbol{\pi}(\cdot)$ from $\boldsymbol{\theta}$ to a new set of parameters $\boldsymbol{\pi}$, with $\text{rank}[\partial\boldsymbol{\pi}'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}] = p = \dim(\boldsymbol{\theta})$ 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*

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

where $\boldsymbol{\Lambda}_{mT}^{\sim}$ and $\boldsymbol{\Lambda}_{\mathfrak{m}T}^{\sim}$, are consistent estimators of the generalized 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 $\bar{\mathfrak{n}}_T(\hat{\boldsymbol{\pi}}_T)$.*

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

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

2.3 *Subsets of parameters*

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

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

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

Proposition 3. *Define*

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

as two sets of p_1 transformed sample moment conditions, where

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

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

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

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

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

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

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

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

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

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

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

2.4 *Multiple simultaneous comparisons*

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

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

In general, the best pairwise comparison, in the sense of having maximum power against a given sequence of local alternatives, would be the one with the highest non-centrality parameter among those tests with the same number of degrees of freedom.[4] But in practice, a researcher might not be able to make the required calculations without knowing the nature of the departure from the null. In those circumstances, a sensible solution would be to simultaneously compare all the alternative estimators. Such a generalization 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 $\hat{\boldsymbol{\theta}}_T^j$, $j = 1, \dots, J$ denote an ordered sequence of asymptotically Gaussian estimators of $\boldsymbol{\theta}$ whose joint asymptotic covariance matrix adopts the following form*:

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

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

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

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

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

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

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

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

3. Application to non-Gaussian likelihood estimators

3.1 *Model specification*

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

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

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

To complete the model, a researcher needs to specify the conditional distribution of $\boldsymbol{\varepsilon}_t^*$. In Supplemental Appendix D (Fiorentini and Sentana (2021)), 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 characterize the distribution of $\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$ (see Supplemental Appendix C for a brief introduction to spherically symmetric distributions).[5] The most prominent example is the standard multivariate normal, which we denote by $\boldsymbol{\eta} = \mathbf{0}$ without loss of generality. Another important example favored by empirical researchers is the standardized multivariate Student t with ν degrees of freedom, or $i.i.d.\ t(\mathbf{0},\mathbf{I}_N,\nu)$ for short. As is well known, the multivariate t approaches the multivariate normal as $\nu \to \infty$, but has generally fatter tails and allows for cross-sectional dependence beyond correlation. For tractability, we define η as $1/\nu$, which will always remain in the finite range $[0, 1/2)$ under our assumptions.[6] Obviously, in the univariate case, any symmetric distribution, including the GED (also known as the Generalized Gaussian distribution), is spherically symmetric, too.[7]

3.2 *Likelihood-based estimators*

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

1. *Restricted ML (RML)*: $\hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}})$, which is such that $\hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}}) = \arg\max_{\boldsymbol{\theta}\in\Theta} L_T(\boldsymbol{\theta},\bar{\boldsymbol{\eta}})$. Its efficiency can be characterized by the $\boldsymbol{\theta}$, $\boldsymbol{\theta}$ block of the information matrix, $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)$, provided that $\bar{\boldsymbol{\eta}} = \boldsymbol{\eta}_0$. Thus, we can interpret $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)$ as the restricted parametric efficiency bound.

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

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 $\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$. This estimator is usually computed by means of one BHHH iteration of the spherically symmetric efficient score starting from a consistent estimator

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

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

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

(see Supplemental Appendix C.5 for further computational details).[8] Associated to it we have the spherically symmetric semiparametric efficiency bound $\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)$.

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

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

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

3.3 *Covariance relationships*

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

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

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

and

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

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

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

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

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

3.4 *Multiple simultaneous comparisons*

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

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

Still, researchers may disregard $\ddot{\boldsymbol{\theta}}_T - \tilde{\boldsymbol{\theta}}_T$ because the semiparametric estimator and the Gaussian estimator are consistent for $\boldsymbol{\theta}_0$ regardless of the conditional distribution, at least as long as the *i.i.d.* assumption holds. For the same reason, they will also disregard

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

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

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

$$\lim_{T\to\infty} V\big[\sqrt{T}\bar{\mathbf{s}}'_{\boldsymbol{\theta}|\boldsymbol{\eta}T}(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta}_0)\big] = \mathcal{P}(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{P}(\boldsymbol{\phi}_0) - \mathcal{P}(\boldsymbol{\phi}_0) \quad \text{and}$$

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

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

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

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

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

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

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

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

while the asymptotic distribution of the trinity of classical tests will be a noncentral distribution with q degrees of freedom and noncentrality parameter

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

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

3.5 *Subsets of parameters*

As in Section 2.3, we may be interested in focusing on a parameter subset either to avoid generalized 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* $\mathbf{y}_t$,

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

$$\lim_{T\to\infty} V\big[\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2\boldsymbol{\eta}T}(\tilde{\boldsymbol{\theta}}_T,\boldsymbol{\eta}_0)\big]$$
$$= \big[\mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)\big]^{-1}\mathcal{C}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)\big[\mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)\big]^{-1} - \big[\mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)\big]^{-1} \quad \textit{and}$$
$$\lim_{T\to\infty} V\big[\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2T}(\hat{\boldsymbol{\theta}}_T,\mathbf{0})\big]$$
$$= \big[\mathcal{A}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)\big]^{-1}\big[\mathcal{C}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0) - \mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)\big]\big[\mathcal{A}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)\big]^{-1}, \quad \textit{where}$$
$$\bar{\mathbf{s}}_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2\boldsymbol{\eta}T}(\boldsymbol{\theta},\boldsymbol{\eta})$$
$$= \bar{\mathbf{s}}_{\boldsymbol{\theta}_1T}(\boldsymbol{\theta},\boldsymbol{\eta}) - \begin{bmatrix}\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\end{bmatrix}\begin{bmatrix}\mathcal{I}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\theta}_2\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\\ \mathcal{I}'_{\boldsymbol{\theta}_2\boldsymbol{\eta}}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\end{bmatrix}^{-1}\begin{bmatrix}\bar{\mathbf{s}}_{\boldsymbol{\theta}_2T}(\boldsymbol{\theta},\boldsymbol{\eta})\\ \bar{\mathbf{s}}_{\boldsymbol{\eta}T}(\boldsymbol{\theta},\boldsymbol{\eta})\end{bmatrix}, \tag{10}$$
$$\mathcal{P}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0)$$
$$= \left\{\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0) - \begin{bmatrix}\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\end{bmatrix}\begin{bmatrix}\mathcal{I}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\theta}_2\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\\ \mathcal{I}'_{\boldsymbol{\theta}_2\boldsymbol{\eta}}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\end{bmatrix}^{-1}\begin{bmatrix}\mathcal{I}'_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\\ \mathcal{I}'_{\boldsymbol{\theta}_1\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\end{bmatrix}\right\}^{-1},$$

while

$$\bar{\mathbf{s}}_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2T}(\boldsymbol{\theta},\mathbf{0}) = \bar{\mathbf{s}}_{\boldsymbol{\theta}_1T}(\boldsymbol{\theta},\mathbf{0}) - \mathcal{A}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\mathcal{A}^{-1}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\bar{\mathbf{s}}_{\boldsymbol{\theta}_2T}(\boldsymbol{\theta},\mathbf{0}), \quad \textit{and}$$
$$\mathcal{A}^{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0) = \big[\mathcal{A}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}_0) - \mathcal{A}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\mathcal{A}^{-1}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\mathcal{A}'_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\phi}_0)\big]^{-1}.$$

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

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

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

and

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

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

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

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

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

3.6 *Choosing the correct number of degrees of freedom*

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

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

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

where

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

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

The first part of this proposition, which generalizes 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 generalizes Result 2 in Gonzalez-Rivera and Drost (1999) and Proposition 6 in Hafner and Rombouts (2007), implies that normality is the only such instance within the spherical family.

For practical purposes, this result implies that a researcher who assumes multivariate normality cannot use DWH tests to assess distributional misspecification. But it also

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

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

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

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

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

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

The first part of this proposition, which generalizes 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 generalization of the multivariate normal that gives rise to some interesting theoretical results,[11] the density of a leptokurtic Kotz distribution has a pole at 0, which is a potential drawback from an empirical point of view.

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

There are also other more subtle but far more pervasive situations in which some, but not all elements of $\boldsymbol{\theta}$ can be estimated as efficiently as if $\boldsymbol{\eta}_0$ were known (see also Lange, Little, and Taylor (1989)), a fact that would be described in the semiparametric literature as partial adaptivity. Effectively, this requires that some elements of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$ be orthogonal to the relevant tangent set after partialing 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,

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

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

often depends on the model parametrization. The following reparametrization provides a general sufficient condition in multivariate dynamic models under sphericity:

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

This parametrizations simply requires the pseudo-standardized residuals

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

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

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

Proposition 13. 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* $\boldsymbol{\varphi}_c$, $\ddot{\boldsymbol{\varphi}}_{cT}$, *is* $\boldsymbol{\varphi}_i$*-adaptive, where* $\boldsymbol{\varphi}_i=(\boldsymbol{\varphi}'_{im},\boldsymbol{\varphi}'_{ic})'$.

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

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

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

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

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

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

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

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

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

3.7 *Maximizing power*

As we discussed in Section 2.1, the local power of a pairwise DWH test depends on the difference in the pseudo-true values of the parameters under misspecification relative to the difference between the covariance matrices under the null. But Proposition 1 in Fiorentini and Sentana (2019) states that in the situation discussed in Proposition 12, $\boldsymbol{\vartheta}_c$ will be consistently estimated when the true distribution of the innovations is spherical but different from the one assumed for estimation purposes, while ϑ_i will be inconsistently estimated. Therefore, rather than losing power by disregarding all the elements of $\boldsymbol{\vartheta}_c$, we will in fact maximize power if we base our DWH tests on the overall scale parameter ϑ_i exclusively. Similarly, Proposition 3 in Fiorentini and Sentana (2019) states that in the context of Proposition 13, $\boldsymbol{\varphi}_c$ will be consistently estimated when the true distribution of the innovations is *i.i.d.* but different from the one assumed for estimation purposes, while $\boldsymbol{\varphi}_{im}$ and $\boldsymbol{\varphi}_{ic}$ will be inconsistently estimated. Consequently, we will maximize power in that case if we base our DWH tests on the mean and covariance parameters of the pseudo standardized residuals $\boldsymbol{\varepsilon}^{\diamond}_t(\boldsymbol{\varphi}_c)$ in (15).

3.8 *Extensions to structural models*

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

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

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

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

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

[13]The LQ decomposition is intimately related to the QR decomposition. Specifically, $\mathbf{Q}'\boldsymbol{\Sigma}_L'$ provides the QR decomposition of the matrix $\mathbf{C}'$,which is uniquely defined if we restrict the diagonal elements of $\boldsymbol{\Sigma}_L$ to be positive (see, e.g., Golub and van Loan (2013) for further details).

[14]See Section 9 of Magnus, Pijls, and Sentana (2021) for a detailed discussion of three ways of explicitly parametrizing a rotation (or special orthogonal) matrix: (i) as the product of Givens matrices that depend on $N(N-1)/2$ Tait–Bryan angles, one for each of the strict upper diagonal elements; (ii) by using the so-called Cayley transform of a skew-symmetric matrix; and (c) by exponentiating a skew-symmetric matrix. Our procedures apply regardless of the chosen parametrization.

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

For simplicity, in the rest of this section we assume that the N structural shocks are cross-sectionally independent with symmetric marginal distributions. One particularly important example will be $\varepsilon_{it}^*|I_{t-1} \sim i.i.d.\ t(0, 1, \nu_i)$. Univariate t distributions are very popular in finance as a way of capturing fat tails while nesting the traditional Gaussian assumption. Their popularity is also on the rise in macroeconomics, as illustrated by Brunnermeier, Palia, Sastry, and Sims (2021).

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

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

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

This result implies that we should base the DWH tests on the comparison of the restricted and unrestricted ML estimators of the elements of $\boldsymbol{\psi}$, their squares or logs, thereby avoiding the need for generalized inverses that would arise if we compared the estimators of the N^2 elements of $\mathbf{c}$ (see Proposition B1.3).[16] As usual, we can obtain two asymptotically equivalent tests by using the scores with respect to $\boldsymbol{\psi}$ instead of the parameter estimators (see Proposition 3). Nevertheless, one should not use any of these tests when one suspects that the innovations are Gaussian not only for the lack of uniformity mentioned after Proposition 10 in Section 3.6, but also because $\boldsymbol{\psi}$ is asymptotically underidentified when two or more shocks are normal.

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

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

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

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

The block diagonality of all the asymptotic covariance matrices immediately implies that we can additively decompose the DWH test that compares all the reduced form parameters into a component that compares the conditional mean parameters and another one that compares the residual covariance matrix $\boldsymbol{\Sigma}$ or its Cholesky decomposition. However, Fiorentini and Sentana (2020) show that if the true joint density of the structural shocks $\boldsymbol{\varepsilon}_t^*$ in (18) is the product of N univariate densities but they are different from the ones assumed for ML estimation purposes, then the restricted and unrestricted non-Gaussian (pseudo) ML estimators of model (18) remain consistent for $\mathbf{a}$ and $\mathbf{j}$ but not for $\boldsymbol{\tau}$ or $\boldsymbol{\psi}$. Thus, the parameters that are efficiently estimated by the unrestricted ML estimator remain once again consistently estimated under distributional misspecification. Although we cannot exploit the consistency of $\mathbf{j}$ to increase the power of the DWH test that compares the ML estimators of the reduced form variance parameters with the Gaussian ones because we cannot separately identify them with a Gaussian pseudo log-likelihood, it makes sense to increase the power of the DWH test that compares the ML estimators of the mean parameters with the Gaussian ones by saving degrees of freedom and focusing on either the drifts in $\boldsymbol{\tau}$ or the unconditional means in $\boldsymbol{\mu}$ even though they do not directly affect the IRFs and FEVDs. Using the results on invariance to reparametrization in Proposition 2, the DWH test of all the mean parameters is asymptotically equivalent whether we parametrize the model in term of $(\boldsymbol{\tau}, \mathbf{a})$ or $(\boldsymbol{\mu}, \mathbf{a})$, and in fact, some of the score versions will be numerically identical. In contrast, the DWH tests that only focus on either $\boldsymbol{\tau}$ or $\boldsymbol{\mu}$ will be different.[18]

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

[18]The intuition is as follows. In the case of the unconditional mean parametrization, the block diagonality of the information matrix not only arises between the conditional mean parameters and the rest, but also between $\boldsymbol{\mu}$ and $\mathbf{a}$, with the same being true for the Gaussian PMLE covariance matrix. As a result,

4. Monte Carlo evidence

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

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

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

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

Random draws of ε_t^* are obtained from four different distributions: two standardized Student t with $\nu = 12$ and $\nu = 8$ degrees of freedom, a standardized symmetric fourth-order Gram–Charlier expansion with an excess kurtosis of 3.2, and another standardized 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 2000 (plus another 100 for initialization) with $\beta = 0.85$, $\alpha = 0.1$, $\tau = 0.05$, and $\omega = 1$, which means that $\delta = \varphi_{im} = 0.05$, $\gamma = 0.1$, and $\vartheta_i = \varphi_{ic} = 1$. These parameter values ensure the strict stationarity of the observed process. Under the null, the large number of Monte Carlo replications implies that the 95% 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 maximizing the log-likelihood function of the Student t distribution with and without fixing the degrees of freedom parameter to 12. We initialize 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 maximized with a quasi-Newton algorithm implemented

the DWH test of the conditional mean parameters can be additively separated between the DWH test of $\boldsymbol{\mu}$, which has all the power, and the DWH test of $\mathbf{a}$, whose asymptotic power is equal to its size. In contrast, neither the information matrix nor the Gaussian sandwich matrix are block diagonal between $\boldsymbol{\tau}$ and $\mathbf{a}$ when we rely on the parametrization in terms of the drifts, which means that the DWH test based on the drifts is not asymptotically independent from the DWH test based the dynamic regression coefficients $\mathbf{a}$. But since both the DWH test of all the mean parameters and the DWH test for $\mathbf{a}$ are the same in both reparametrizations, the DWH test based on $\boldsymbol{\tau}$ must be different from the DWH test for $\boldsymbol{\mu}$. The ordering of the local power of these two tests is unclear.

by means of the NAG library E04LBF routine with the analytical expressions for the score vector and conditional information matrix in Fiorentini, Sentana, and Calzolari (2003).

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

When the true distribution of the standardized 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 reparametrization,[19] seems to work very well.

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

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

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

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

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

[19]Proposition 2 implies that the score tests will be numerically invariant to reparametrizations 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.

Table 1. Univariate GARCH-M: empirical rejection rates.

	RML = UML $\vartheta_i@(\bar{\boldsymbol{\theta}}_T, \bar{\eta})$				UML = PML $\vartheta_i@(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T)$			UML = PML $(\varphi_{im}, \varphi_{ic})@(\tilde{\boldsymbol{\theta}}_T, \breve{\eta}_T)$			RML = UML & UML = PML $\vartheta_i@(\bar{\boldsymbol{\theta}}_T, \bar{\eta}_T)$		
%	DWH1	DWH2	DWH3	LR	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3
Student t_{12}													
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													
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)													
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)													
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

Note: 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 = 2000. Replications = 20,000.

The conditional mean and variance parameters are $\boldsymbol{\theta}' = (\mathbf{a}', \mathbf{b}', \boldsymbol{\omega}')$, where $\boldsymbol{\omega} = \text{vech}(\boldsymbol{\Omega})$ and $\boldsymbol{\Omega} = \boldsymbol{\Omega}^{1/2}\boldsymbol{\Omega}'^{1/2}$. In this case, Fiorentini and Sentana (2019) showed that 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 $\boldsymbol{\Omega}^\circ(\boldsymbol{\varpi}) = \boldsymbol{\Omega}/|\boldsymbol{\Omega}|^{1/N}$ (reparametrization 1) or $\boldsymbol{\varphi}_c = \mathbf{b}$, $\boldsymbol{\varphi}_{im} = \mathbf{a}$ and $\boldsymbol{\varphi}_{ic} = \text{vech}(\boldsymbol{\Phi}_{ic}) = \text{vech}(\boldsymbol{\Omega})$ (reparametrization 2).

We consider four standardized multivariate distributions for $\boldsymbol{\varepsilon}_t^*$, including two multivariate Student t with $\nu = 12$ and $\nu = 8$ degrees of freedom, a discrete scale mixture of two normals (DSMN) with mixing probability 0.2 and variance ratio 10, and an asymmetric, location-scale mixture (DLSMN) with the same parameters but a difference in the mean vectors of the two components $\boldsymbol{\delta} = 0.5\ell_N$, where ℓ_N is a vector of N ones (see Amengual and Sentana (2010) and Supplemental Appendix E.1, respectively, for further details). For each distribution, we generate 20,000 samples of dimension $N = 3$ and length $T = 500$ with $\mathbf{a} = 0.112\ell_3$, $\mathbf{b} = \ell_3$, 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 2 show the results of the size and power assessment of our proposed DWH tests. As in the previous example, the DWH3 version of the test appears to be the best one here too, although not uniformly so. When we compare restricted and unrestricted MLEs, 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 $\text{vech}(\boldsymbol{\Omega})$, 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.

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

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

In the size experiment, ε_{1t}^* and ε_{2t}^* are two independent standardized Student ts with $\eta_1 = 0.15$ and $\eta_2 = 0.10$, respectively, but in the power experiment ε_{1t}^* is drawn from a symmetric DSMN with mixing probability 0.52 and variance ratio 0.06, while ε_{2t}^* follows an asymmetric DLSMN with mixing probability 0.3, variance ratio 0.2 and $\delta = 0.5$. The sample length is $T = 2000$.

TABLE 2. Multivariate market model: empirical rejection rates.

%	RML = UML $\vartheta_i@(\bar{\boldsymbol{\theta}}_T, \bar{\eta})$ DWH1	DWH2	DWH3	LR	UML = PML $\vartheta_i@(\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T)$ DWH1	DWH2	DWH3	UML = PML $(\mathbf{a}, \text{vech}(\boldsymbol{\Omega}))@(\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T)$ DWH1	DWH2	DWH3	RML = UML & UML = PML $\vartheta_i@(\bar{\boldsymbol{\theta}}_T, \bar{\eta})$ DWH1	DWH2	DWH3
							Student t_{12}						
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						
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)						
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)						
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

Note: 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.

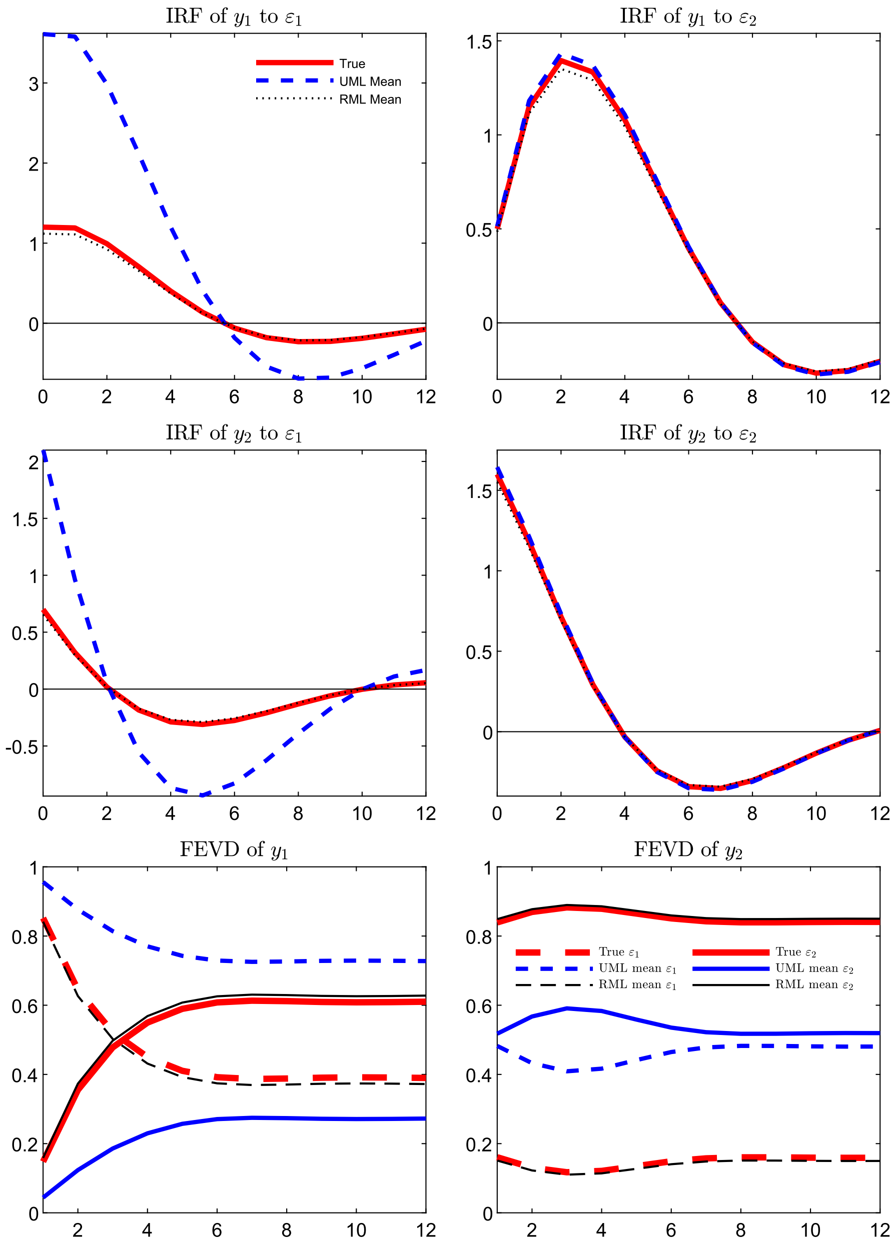

FIGURE 1. IRF and FVED. DGP: Independent DLSMN$_{(\delta_1,\kappa_1,\lambda_1)=(0,0.52,0.06),(\delta_2,\kappa_2,\lambda_2)=(0.3,0.2,0.2)}$.

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

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

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

5. Empirical illustrations

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

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

[20] As expected from Proposition 10, though, size distortions become a serious problem in a separate Monte Carlo exercise in which ε^*_{1t} and ε^*_{2t} are two independent standardized Student t with with 66.6 and 100 degrees of freedom, respectively, which are rather difficult to distinguish from Gaussian random variables in finite samples.

Table 3. Structural VAR(1): empirical rejection rates.

	RML = UML diag($\mathbf{C}$)@($\bar{\boldsymbol{\theta}}_T, \bar{\eta}$)				UML = PML vech($\boldsymbol{\Sigma}$)@($\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T$)			UML = PML $\boldsymbol{\tau}$@($\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T$)			UML = PML ($\boldsymbol{\tau}$, vech($\boldsymbol{\Sigma}$))@($\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T$)			RML = UML & UML = PML (vech($\boldsymbol{\Sigma}$), diag($\mathbf{C}$))@($\bar{\boldsymbol{\theta}}_T, \bar{\eta}$)		
%	DWH1	DWH2	DWH3	LR	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3	DWH1	DWH2	DWH3
Independent Student $t_{(\eta_1=0.15, \eta_2=0.10)}$, Replications 20,000																
1	2.70	2.22	1.11	1.06	3.39	2.06	3.04	1.65	1.16	1.51	3.64	2.09	3.31	4.61	2.76	3.28
5	7.26	6.86	4.98	5.11	6.65	4.17	5.86	5.77	4.87	5.42	7.39	4.85	6.63	8.74	6.18	6.54
10	12.40	11.88	9.76	9.97	9.75	6.51	8.81	10.61	9.78	10.26	11.18	7.87	10.13	12.29	9.47	9.70
Independent DLSMN$_{(\delta_1,\kappa_1,\lambda_1)=(0.52,0.06,0),(\delta_2,\kappa_2,\lambda_2)=(0.3,0.2,0.5)}$, Replications 5000																
1	100	100	100	100	100	100	100	99.84	92.88	99.80	100	100	100	100	100	100
5	100	100	100	100	100	100	100	99.90	97.02	99.86	100	100	100	100	100	100
10	100	100	100	100	100	100	100	99.96	98.14	99.88	100	100	100	100	100	100

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

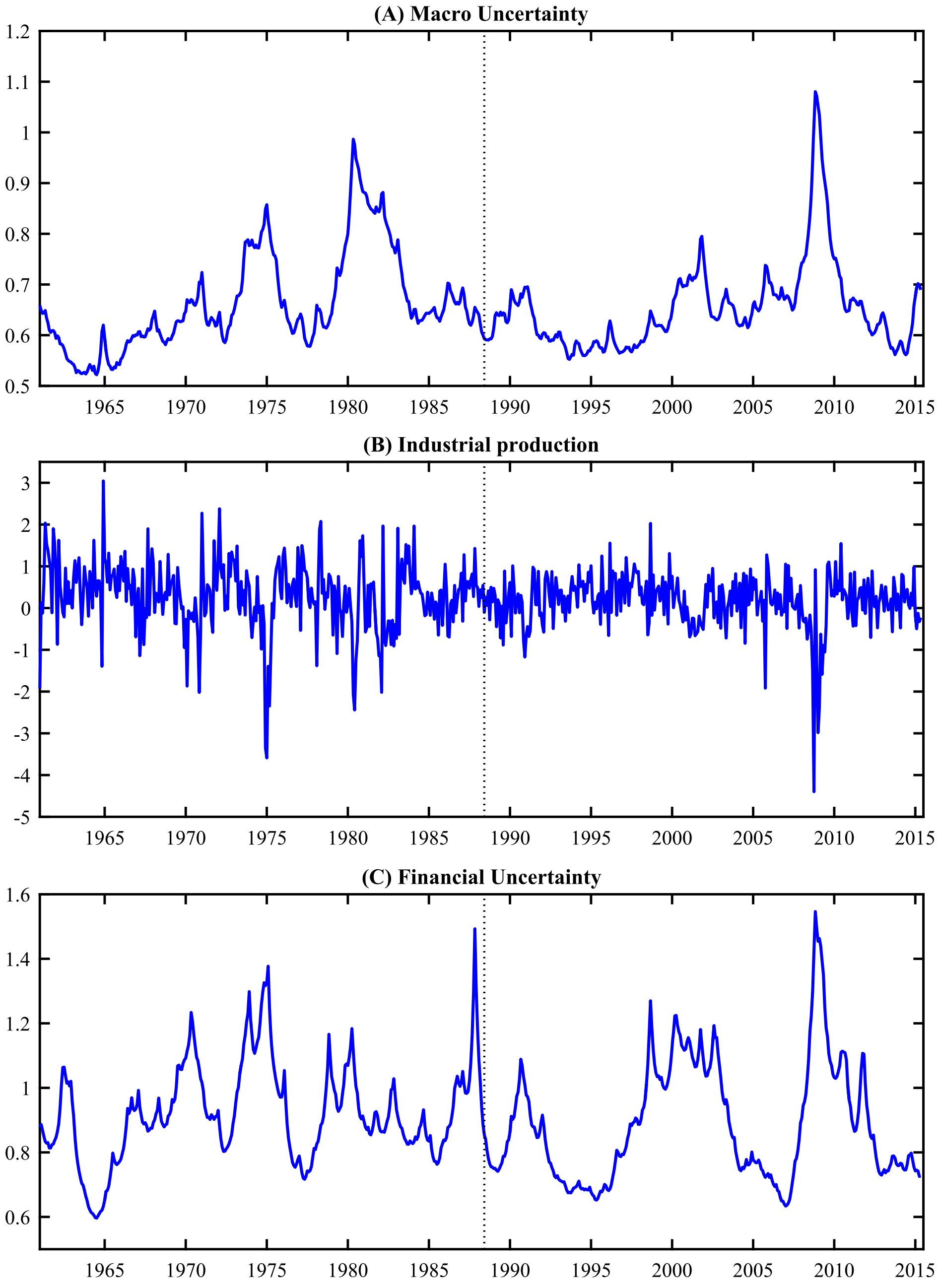

Figure 2. Data.

We adopt the original Var(4) specification in Angelini et al. (2019), which implies that $T = 653$ after initialization of the log-likelihood with 4 presample observations. Our main point of departure is that we assume that the structural innovations follow three independent standardized Student t distributions with ν_i degrees of freedom, which allows us to identify the entire matrix of impact multipliers $\mathbf{C} = \mathbf{J}\boldsymbol{\Psi}$. Thus, the unrestricted

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

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

ML procedure estimates $2N + (p+1)N^2 = 51$ parameters, while the restricted MLE fixes $\nu_1 = \nu_2 = \nu_3 = 8$ (We tried different values for the ν's ranging from 6 to 10 but results were very similar). Finally, the Gaussian PMLE estimates $N(N-1)/2 = 3$ parameters less because it can only identify $\mathbf{CC}' = \mathbf{J}\boldsymbol{\Psi}^2\mathbf{J}' = \boldsymbol{\Sigma}$.

Our PML estimators of the autoregressive matrices coincide with those in Angelini et al. (2019). Further, the restricted and unrestricted MLEs of those parameters are also very similar because the three estimators are consistent under weak conditions, as we explained in Section 3.8. The estimates of the drift, the (scaled) impact multiplier matrix $\mathbf{J}$, the standard deviations of the structural shocks in $\boldsymbol{\Psi}$ and the unconditional variance of the one period ahead forecast errors $\boldsymbol{\Sigma}$ are reported in Table 4. As can be seen, the three estimators of the drift parameters are quite similar for the first two series, while for the last one the sign of the UML and RML estimators is reversed with respect to the PML one. A look at the estimators of $\boldsymbol{\Sigma}$ reveals both an unbalanced scaling of the data, and a low predictability in the rate of growth of the industrial production index. The restricted and unrestricted MLEs of $\mathbf{J}$ are rather similar. In fact, the consistency of the non-Gaussian ML estimators of the matrix $\mathbf{J}$ is indirectly confirmed by the extremely high ($= 0.995$) time series correlation between the (nonstandardized) estimates of each structural shock obtained as $\mathbf{J}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ evaluated at the RMLE and UMLE. In contrast, there is a striking difference in the standard deviation of the third structural shock, which strongly points to distributional misspecification. However, this conjecture needs to be confirmed by our formal DWH test statistics, which account for the sampling variability of the estimators.

The three versions of our DWH tests produce qualitatively similar results. For that reason, in Table 5 we only report the results of the versions that evaluate the score of the more efficient estimators at the less efficient ones (e.g., the unrestricted Student t scores at the Gaussian PMLE). According the Monte Carlo results in the previous section, these are the most conservative ones. As expected, we conclude that the null of correct specification of the structural innovation distributions is clearly rejected. The test statistics that compares the unrestricted ML estimator of the variance of the Wold innovations $\hat{\mathbf{J}}\hat{\boldsymbol{\Psi}}^2\hat{\mathbf{J}}'$ with its PML counterpart $\bar{\boldsymbol{\Sigma}}$ has a tiny p-value. Similarly, if we compare the same estimators of the drift parameters, the p-value of our DWH statistic is 0.001. Given the

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

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

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

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

Test	d.f.	Statistic	p-value
		PML vs. UML	
$\boldsymbol{\tau}@(\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T)$	3	5.65	0.130
$\text{vech}(\boldsymbol{\Sigma})@(\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T)$	6	14.57	0.024
$(\boldsymbol{\tau}, \text{vech}(\boldsymbol{\Sigma}))@(\hat{\boldsymbol{\theta}}_T, \hat{\eta}_T)$	9	20.22	0.017
		UML vs. RML	
$\text{diag}(\mathbf{C})@(\bar{\boldsymbol{\theta}}_T, \bar{\eta})$	3	69.69	0.0
$\boldsymbol{\eta} = \bar{\boldsymbol{\eta}}$	3	37.82	0.0

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

additivity of these two test statistics mentioned at the end of Section 3.8, the p-value of the joint test is virtually zero. As for the comparison between the restricted and unrestricted MLEs of the diagonal elements of $\boldsymbol{\Psi}$, which contain the standard deviations of the structural shocks, the DWH tests massively reject once again. This rejection is confirmed by the asymptotically equivalent LR test of $H_0 : \nu_1 = \nu_2 = \nu_3 = 8$.

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

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

6. Conclusions and directions for further research

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

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

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

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

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

Given the increased popularity of Independent Component Analysis in econometric applications, as illustrated by the SVARs in Section 3.8, specification tests that directly target the maintained assumptions of nonnormality and independence of the structural shocks provide a particularly appropriate complement to our proposed tests (see Amengual, Fiorentini, and Sentana (2021)). We could also extend our theoretical results to a broad class of models for which a pseudo log-likelihood function belonging to the linear exponential family leads to consistent estimators of the conditional mean parameters (see Gouriéroux, Monfort, and Trognon (1984a)). For example, we could use a DWH test to assess the correct distributional specification of Lanne's (2006) multiplicative error model for realized volatility by comparing his ML estimator based on a two-component Gamma mixture with the Gamma-based consistent pseudo ML estimators in Engle and Gallo (2006). Similarly, we could also use the same approach to test the correct specification of the count model for patents in Hausman, Hall, and Griliches (1984) by comparing their ML estimator, which assumes a Poisson model with unobserved gamma heterogeneity, with the consistent pseudo ML estimators in Gouriéroux, Monfort, and Trognon (1984b). All these extensions constitute interesting avenues for further research.

Appendix A: Proofs

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

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

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

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

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

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

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

Therefore,

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

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

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

If we knew $\boldsymbol{\theta}_n$, it would be straightforward to test whether (A5) holds. But since we do not know it, we replace it by its consistent estimator $\tilde{\boldsymbol{\theta}}_T$, which satisfies (A2). To account for the sampling variability that this introduces under the null, we can use again a first-order Taylor expansion around $\boldsymbol{\theta}_0$ of the sample version of (A5) evaluated at $\tilde{\boldsymbol{\theta}}_T$. Given the assumed root-T consistency of $\tilde{\boldsymbol{\theta}}_T$ for $\boldsymbol{\theta}_0$, we can use (A3) to write this expansion as

$$\begin{aligned}
&\mathcal{J}_m'(\tilde{\boldsymbol{\theta}}_T)\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\theta}}_T)\\
&\quad = \mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_0) + \mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) + o_p(1)\\
&\quad = \mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_0)\\
&\qquad - \big[\mathcal{J}_m'(\boldsymbol{\theta}_0)\mathcal{S}_m\mathcal{J}_m(\boldsymbol{\theta}_0)\big]\big[\mathcal{J}_n(\boldsymbol{\theta}_0)\mathcal{S}_n(\boldsymbol{\theta}_0)\mathcal{J}_n'(\boldsymbol{\theta}_0)\big]^{-1}\mathcal{J}_n'(\boldsymbol{\theta}_0)\mathcal{S}_n\sqrt{T}\bar{\mathbf{n}}_T(\boldsymbol{\theta}_0) + o_p(1).
\end{aligned} \tag{A6}$$

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

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

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

Proof of Proposition 2. The Wald-type version of the Hausman test for the original parameters in (1) is infeasible when $\boldsymbol{\Delta}$ is unknown, in which case it must be computed as

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

where $\boldsymbol{\Delta}_T^{\sim}$ denotes a consistent estimator of a generalized inverse of $\boldsymbol{\Delta}$, which does not necessarily coincide with a generalized inverse of a consistent estimator of the asymptotic covariance matrix of $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T)$ because of the potential discontinuities of generalized inverses. Given the assumed regularity of the reparametrization, 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 generalized inverse provided that $\dot{\boldsymbol{\pi}}_T$ is a consistent estimator of $\boldsymbol{\pi}_0$. Therefore, the Wald-type version of the Hausman test for the original parameters will be

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

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

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

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

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

as required.

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

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

Lemma 1 also implies that $\bar{\mathfrak{m}}_T(\boldsymbol{\pi})=\bar{\mathbf{m}}_T[\boldsymbol{\theta}(\boldsymbol{\pi})]$ 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 reparametrization, 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

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

As a consequence, it seems natural to use

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

as a consistent estimator of a generalized inverse of $\boldsymbol{\Lambda}_{\mathfrak{m}}$, 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

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

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

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

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

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

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

Proof of Proposition 3. Again, we focus on the first result, as the second one is entirely analogous. Let us start from the asymptotic equivalence relationship (A7). Given that

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

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

the application of the partitioned inverse formula yields

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

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

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 $\bar{\mathbf{m}}_{1T}^{\perp}(\tilde{\boldsymbol{\theta}}_T, \mathcal{S}_n)$. □

PROOF OF PROPOSITION 4. Given that

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

it follows immediately from (5) that

$$\lim_{T\to\infty} V\left[\sqrt{T}\begin{pmatrix} \hat{\boldsymbol{\theta}}_T^2 - \hat{\boldsymbol{\theta}}_T^1 \\ \hat{\boldsymbol{\theta}}_T^3 - \hat{\boldsymbol{\theta}}_T^2 \\ \vdots \\ \hat{\boldsymbol{\theta}}_T^{J-1} - \hat{\boldsymbol{\theta}}_T^{J-2} \\ \hat{\boldsymbol{\theta}}_T^J - \hat{\boldsymbol{\theta}}_T^{J-1} \end{pmatrix}\right]$$

$$= \begin{bmatrix} \boldsymbol{\Omega}_2 - \boldsymbol{\Omega}_1 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_3 - \boldsymbol{\Omega}_2 & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Omega}_{J-1} - \boldsymbol{\Omega}_{J-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \boldsymbol{\Omega}_J - \boldsymbol{\Omega}_{J-1} \end{bmatrix}, \tag{A14}$$

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

PROOF OF PROPOSITION 5. Given that Propositions C1–C3 in Supplemental Appendix C and Proposition D3 in Supplemental Appendix D derive all the information bounds, we simply need to compute the off-diagonal elements. Let us start with the first row. Straightforward manipulations imply that

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

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

$$\begin{aligned} &E\big[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathring{\mathbf{s}}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}\big] \\ &\quad = E\big[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\big\{\mathbf{e}'_{dt}(\boldsymbol{\phi})\mathbf{Z}'_{dt}(\boldsymbol{\theta}) - \big[\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi}) - \mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^{+}(\kappa)\mathring{\mathcal{K}}(0)\big]\mathbf{Z}'_{d}(\boldsymbol{\phi})\big\}|\boldsymbol{\phi}\big] \end{aligned}$$

$$= E\big[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\phi})\mathbf{Z}_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}\big]$$
$$- E\big\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\big[\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi}) - \mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^{+}(\kappa)\mathring{\mathcal{K}}(0)\big]\mathbf{Z}'_{d}(\boldsymbol{\phi})|\boldsymbol{\phi}\big\}$$
$$= \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}'_s(\boldsymbol{\phi}_0)\cdot\left\{\left[\frac{N+2}{N}\,\mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1\right] - \frac{4}{N\big[(N+2)\kappa_0+2\big]}\right\} = \mathring{\mathcal{S}}(\boldsymbol{\phi}_0)$$

by virtue of the law of iterated expectations, together with expressions (C33), (C34), and (C35) in Supplemental Appendix C. Intuitively, $\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)$ is the variance of the error in the least squares projection of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$ onto the Hilbert space spanned by all the time-invariant functions of $\boldsymbol{\varsigma}_t(\boldsymbol{\theta}_0)$ 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 D3 in Supplemental Appendix D that

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

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

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

thanks to the generalized information equality.

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

$$E\big[\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\phi})\mathring{\mathbf{s}}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}\big]$$
$$= E\big[\big\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi})\mathcal{I}^{-1}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi})\mathbf{e}_{rt}(\boldsymbol{\phi})\big\}$$
$$\times\big\{\mathbf{e}'_{dt}(\boldsymbol{\phi})\mathbf{Z}'_{dt}(\boldsymbol{\theta}) - \big[\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi}) - \mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^{+}(\kappa)\mathring{\mathcal{K}}(0)\big]\mathbf{Z}'_{d}(\boldsymbol{\phi})\big\}|\boldsymbol{\phi}\big]$$
$$= E\big[\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\mathbf{Z}'_{dt}(\boldsymbol{\phi}_0)|\boldsymbol{\phi}\big] - E\big[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi})\mathbf{Z}'_{dt}(\boldsymbol{\phi}_0)|\boldsymbol{\phi}\big]$$
$$+ E\big[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathcal{K}^{+}(\boldsymbol{\varrho}_0)\mathcal{K}(0)\mathbf{Z}'_{d}(\boldsymbol{\phi})|\boldsymbol{\phi}\big]$$

$$
\begin{aligned}
&-\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi})\mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi})E\big[\mathbf{e}_{rt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\phi})\mathbf{Z}'_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}\big]\\
&+\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi})\mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi})E\big[\mathbf{e}_{rt}(\boldsymbol{\phi})\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi})\mathbf{Z}'_{d}(\boldsymbol{\theta})|\boldsymbol{\phi}\big]\\
&-\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi})\mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi})E\big[\mathbf{e}_{rt}(\boldsymbol{\phi})\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^{+}(\kappa)\mathring{\mathcal{K}}(0)\mathbf{Z}'_{d}(\boldsymbol{\phi})|\boldsymbol{\phi}\big]\\
=\;&\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})-\mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}'_s(\boldsymbol{\phi}_0)\cdot\left\{\left[\frac{N+2}{N}\,\mathrm{M}_{ss}(\boldsymbol{\eta}_0)-1\right]-\frac{4}{N\big[(N+2)\kappa_0+2\big]}\right\}=\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)
\end{aligned}
$$

where we have used the fact that

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

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

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

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

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

by the law of iterated expectations. Finally,

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

because of the generalized 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

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

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

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

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

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

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

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 $\ddot{\mathcal{S}}(\boldsymbol{\phi}_0)-\mathcal{A}(\boldsymbol{\theta})\mathcal{B}^{-1}(\boldsymbol{\phi})\times\mathcal{A}(\boldsymbol{\theta})$ 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

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

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

$$\begin{aligned}&-\partial E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}]/\partial\boldsymbol{\theta}\\&\quad=E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)|\boldsymbol{\phi}]\\&\quad=E[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0)\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\mathbf{Z}'_{dt}(\boldsymbol{\theta}_0)\big|\boldsymbol{\phi}_0]\\&\qquad-E\bigg\{\mathbf{W}_s(\boldsymbol{\phi}_0)\bigg[\bigg[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}-1\bigg]\\&\qquad-\frac{2}{(N+2)\kappa_0+2}\bigg(\frac{\varsigma_t}{N}-1\bigg)\bigg]\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\mathbf{Z}'_{dt}(\boldsymbol{\theta}_0)\bigg|\boldsymbol{\phi}_0\bigg\}\\&\quad=\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)-\mathbf{W}_s(\boldsymbol{\phi}_0)E\bigg\{\bigg[\bigg\{\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}-1\bigg\}\\&\qquad-\frac{2}{(N+2)\kappa_0+2}\bigg(\frac{\varsigma_t}{N}-1\bigg)\bigg]\mathbf{e}'_{dt}(\boldsymbol{\phi}_0)\bigg|\boldsymbol{\phi}_0\bigg\}\mathbf{Z}_d(\boldsymbol{\theta}_0)\end{aligned}$$

$$
\begin{aligned}
&= \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0) E\Bigg[\Bigg\{\Bigg[\delta(\varsigma_t, \boldsymbol{\eta}_0)\frac{\varsigma_t}{N} - 1\Bigg] \\
&\quad - \frac{2}{(N+2)\kappa_0 + 2}\left(\frac{\varsigma_t}{N} - 1\right)\Bigg\}\Bigg[\delta(\varsigma_t, \boldsymbol{\eta}_0)\frac{\varsigma_t}{N} - 1\Bigg]\Bigg|\boldsymbol{\phi}_0\Bigg]\mathbf{W}_s'(\boldsymbol{\phi}_0) \\
&= \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}_s'(\boldsymbol{\phi}_0) \cdot \left\{\left[\frac{N+2}{N}\,\mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1\right] - \frac{4}{N\big[(N+2)\kappa_0 + 2\big]}\right\} \\
&= \mathring{\mathcal{S}}(\boldsymbol{\phi}_0) = E\big[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}'(\boldsymbol{\phi})|\boldsymbol{\phi}\big].
\end{aligned}
\tag{A16}
$$

The generalized information matrix equality also implies that

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

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

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

All the remaining asymptotic covariances are obtained analogously. □

Proof of Proposition 6. Given the efficiency of $\hat{\boldsymbol{\theta}}_T$ relative to $\tilde{\boldsymbol{\theta}}_T$, it follows from Lemma 2 in Hausman (1978) that

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

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

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

by the generalized information matrix equality. □

Proof of Proposition 7. The efficiency of $\hat{\boldsymbol{\theta}}_T(\boldsymbol{\eta})$ relative to $\hat{\boldsymbol{\theta}}_T$ and Lemma 2 in Hausman (1978) imply that

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

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

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

by the information matrix equality. □

Proof of Proposition 8. The proof of Proposition 6 immediately implies that

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

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. □

Proof of Proposition 9. The proof of Proposition 7 immediately implies that

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

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. □

Proof of Proposition 10. The proof of the first part is trivial, except perhaps for the fact that $\mathcal{M}_{sr}(\mathbf{0}) = \mathbf{0}$, which follows from Lemma 3 in Supplemental Appendix B because $\mathbf{e}_{st}(\boldsymbol{\theta}_0, \mathbf{0})$ coincides with $\mathbf{e}_{st}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0)$ under normality.

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

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

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

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

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

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), showed that

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

for the Kotz distribution, which immediately implies that

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

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

When $N = 1$, $(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) = 2$ and $\mathrm{vec}(\mathbf{I}_N)\,\mathrm{vec}'(\mathbf{I}_N) = 1$, 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. □

Proof of Proposition 11. Note that $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})-\mathring{\mathcal{S}}(\boldsymbol{\phi})$ is $\mathbf{W}_s(\boldsymbol{\phi})\mathbf{W}_s'(\boldsymbol{\phi})$ times the residual variance in the theoretical regression of $\delta(\varsigma_t,\boldsymbol{\eta}_0)\varsigma_t/N-1$ on $(\varsigma_t/N)-1$. Therefore, given that $\mathbf{W}_s(\boldsymbol{\phi})\neq\mathbf{0}$, $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})-\mathring{\mathcal{S}}(\boldsymbol{\phi})$ can only be 0 if that regression residual is identically 0 for all t. The solution to the resulting differential equation is

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

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

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

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

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

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

which is the product of a rank one matrix times a nonnegative 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. □

Proof of Proposition 12. Given our assumptions on the mapping $\mathbf{r}_s(\cdot)$, 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

$$\mathbf{Z}_{dt}(\boldsymbol{\vartheta}) = \begin{Bmatrix} \vartheta_i^{-1/2}\left[\partial\boldsymbol{\mu}_t'(\boldsymbol{\vartheta}_c)/\partial\boldsymbol{\vartheta}_c\right]\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_c) & \frac{1}{2}\left\{\partial \operatorname{vec}'\left[\boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_c)\right]/\partial\boldsymbol{\vartheta}_c\right\}\left[\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_c)\otimes\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_c)\right] \\ 0 & \frac{1}{2}\vartheta_i^{-1}\operatorname{vec}'(\mathbf{I}_N) \end{Bmatrix}$$

$$= \begin{bmatrix} \mathbf{Z}_{\boldsymbol{\vartheta}_c lt}(\boldsymbol{\vartheta}) & \mathbf{Z}_{\boldsymbol{\vartheta}_c st}(\boldsymbol{\vartheta}) \\ 0 & \mathbf{Z}_{\vartheta_i st}(\boldsymbol{\vartheta}) \end{bmatrix}. \tag{A18}$$

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

$$\begin{bmatrix} \mathbf{s}_{\boldsymbol{\vartheta}_c t}(\boldsymbol{\vartheta},\boldsymbol{\eta}) \\ s_{\vartheta_i t}(\boldsymbol{\vartheta},\boldsymbol{\eta}) \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{\boldsymbol{\vartheta}_c lt}(\boldsymbol{\vartheta})\mathbf{e}_{lt}(\boldsymbol{\vartheta},\boldsymbol{\eta}) + \mathbf{Z}_{\boldsymbol{\vartheta}_c st}(\boldsymbol{\vartheta})\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta}) \\ \mathbf{Z}_{\vartheta_i st}(\boldsymbol{\vartheta})\mathbf{e}_{st}(\boldsymbol{\vartheta},\boldsymbol{\eta}) \end{bmatrix}, \tag{A19}$$

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

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

$$E\left\{\left[\mathbf{Z}_{\boldsymbol{\vartheta}_c lt}(\boldsymbol{\vartheta}) \quad \mathbf{Z}_{\boldsymbol{\vartheta}_c st}(\boldsymbol{\vartheta})\right]\begin{bmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathbf{0} \\ \mathbf{0} & \mathcal{M}_{ss}(\boldsymbol{\eta}) \end{bmatrix}\begin{bmatrix} 0 \\ \mathbf{Z}'_{\vartheta_i st}(\boldsymbol{\vartheta}) \end{bmatrix}\middle|\boldsymbol{\vartheta},\boldsymbol{\eta}\right\}$$

$$= \frac{\left\{2\mathrm{M}_{ss}(\boldsymbol{\eta}) + N\left[\mathrm{M}_{ss}(\boldsymbol{\eta})-1\right]\right\}}{2\vartheta_i}$$

$$\times E\left\{\frac{1}{2}\frac{\partial \operatorname{vec}'\left[\boldsymbol{\Sigma}_t^\circ(\boldsymbol{\vartheta}_c)\right]}{\partial\boldsymbol{\vartheta}_c}\left[\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_c)\otimes\boldsymbol{\Sigma}_t^{\circ-1/2\prime}(\boldsymbol{\vartheta}_c)\right]\middle|\boldsymbol{\vartheta},\boldsymbol{\eta}\right\}\operatorname{vec}(\mathbf{I}_N)$$

$$= \frac{\left\{2\mathrm{M}_{ss}(\boldsymbol{\eta}) + N\left[\mathrm{M}_{ss}(\boldsymbol{\eta})-1\right]\right\}}{2\vartheta_i}\mathbf{Z}_{\boldsymbol{\vartheta}_c s}(\boldsymbol{\vartheta},\boldsymbol{\eta})\operatorname{vec}(\mathbf{I}_N),$$

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

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

$$E\left\{\left[0 \quad \mathbf{Z}_{\vartheta_i st}(\boldsymbol{\vartheta})\right]\begin{bmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathbf{0} \\ \mathbf{0} & \mathcal{M}_{ss}(\boldsymbol{\eta}) \end{bmatrix}\begin{bmatrix} 0 \\ \mathbf{Z}'_{\vartheta_i st}(\boldsymbol{\vartheta}) \end{bmatrix}\middle|\boldsymbol{\vartheta},\boldsymbol{\eta}\right\}$$

$$= \frac{1}{4\vartheta_i^2}\operatorname{vec}'(\mathbf{I}_N)\left[\mathrm{M}_{ss}(\boldsymbol{\eta})(\mathbf{I}_{N^2}+\mathbf{K}_{NN}) + \left[\mathrm{M}_{ss}(\boldsymbol{\eta})-1\right]\right)\operatorname{vec}(\mathbf{I}_N)\operatorname{vec}'(\mathbf{I}_N)\right]\operatorname{vec}(\mathbf{I}_N)$$

$$= \frac{\left\{2\mathrm{M}_{ss}(\boldsymbol{\eta}) + N\left[\mathrm{M}_{ss}(\boldsymbol{\eta})-1\right]\right\}N}{4\vartheta_i^2}.$$

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

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

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

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

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

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

To prove Part 1c note that

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

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

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

because the regression residual

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

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

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

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

Proof of Proposition 13. We can directly work in terms of the $\boldsymbol{\varphi}$ parameters thanks to our assumptions on the mapping $\mathbf{r}_g(\cdot)$. 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

$$\begin{bmatrix}\mathbf{s}_{\varphi_1 t}(\boldsymbol{\varphi},\boldsymbol{\varrho})\\ \mathbf{s}_{\varphi_{ic}t}(\boldsymbol{\varphi},\boldsymbol{\varrho})\\ \mathbf{s}_{\varphi_{im}t}(\boldsymbol{\varphi},\boldsymbol{\varrho})\end{bmatrix}=\begin{bmatrix}\mathbf{Z}_{\varphi_1 lt}(\boldsymbol{\varphi})\mathbf{e}_{lt}(\boldsymbol{\varphi},\boldsymbol{\varrho})+\mathbf{Z}_{\varphi_1 st}(\boldsymbol{\varphi})\mathbf{e}_{st}(\boldsymbol{\varphi},\boldsymbol{\varrho})\\ \mathbf{Z}_{\varphi_{ic}st}(\boldsymbol{\varphi})\mathbf{e}_{st}(\boldsymbol{\varphi},\boldsymbol{\varrho})\\ \mathbf{Z}_{\varphi_{im}lt}(\boldsymbol{\varphi})\mathbf{e}_{lt}(\boldsymbol{\varphi},\boldsymbol{\varrho})\end{bmatrix}, \tag{A22}$$

where

$$\left.\begin{aligned}&\mathbf{Z}_{\varphi_1 lt}(\boldsymbol{\varphi})=\big\{\partial\boldsymbol{\mu}_t^{\diamond\prime}(\boldsymbol{\varphi}_1)/\partial\boldsymbol{\varphi}_1+\partial\mathrm{vec}'\big[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\varphi}_1)\big]/\partial\boldsymbol{\varphi}_1\cdot(\boldsymbol{\varphi}_{im}\otimes\mathbf{I}_N)\big\}\boldsymbol{\Sigma}_t^{\diamond-1/2\prime}(\boldsymbol{\varphi}_1)\boldsymbol{\Phi}_2^{-1/2\prime},\\ &\mathbf{Z}_{\varphi_1 st}(\boldsymbol{\varphi})=\partial\mathrm{vec}'\big[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\varphi}_1)\big]/\partial\boldsymbol{\varphi}_1\cdot\big[\boldsymbol{\Phi}_2^{1/2}\otimes\boldsymbol{\Sigma}_t^{\diamond-1/2\prime}(\boldsymbol{\varphi}_1)\boldsymbol{\Phi}_2^{-1/2\prime}\big],\\ &\mathbf{Z}_{\varphi_{im}lt}(\boldsymbol{\varphi})=\boldsymbol{\Phi}_2^{-1/2\prime}=\mathbf{Z}_{\varphi_{im}l}(\boldsymbol{\varphi}),\\ &\mathbf{Z}_{\varphi_{ic}st}(\boldsymbol{\varphi})=\partial\mathrm{vec}'\big(\boldsymbol{\Phi}^{1/2}\big)/\partial\boldsymbol{\varphi}_{ic}\cdot\big(\mathbf{I}_N\otimes\boldsymbol{\Phi}_2^{-1/2\prime}\big)=\mathbf{Z}_{\varphi_{ic}s}(\boldsymbol{\varphi}),\end{aligned}\right\} \tag{A23}$$

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

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

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

$$\begin{bmatrix}\mathbf{Z}_{\varphi_c l}(\boldsymbol{\varphi},\boldsymbol{\varrho}) & \mathbf{Z}_{\varphi_c s}(\boldsymbol{\varphi},\boldsymbol{\varrho})\end{bmatrix}\begin{bmatrix}\mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho})\\ \mathcal{M}'_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho})\end{bmatrix}\begin{bmatrix}\mathbf{0} & \mathbf{Z}'_{\varphi_{im}l}(\boldsymbol{\varphi})\\ \mathbf{Z}'_{\varphi_{ic}s}(\boldsymbol{\varphi}) & \mathbf{0}\end{bmatrix}$$

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

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

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

$$\begin{bmatrix}\mathbf{0} & \mathbf{Z}_{\boldsymbol{\varphi}_{ic} s}(\boldsymbol{\varphi})\\ \mathbf{Z}_{\boldsymbol{\varphi}_{im} l}(\boldsymbol{\varphi}) & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho})\\ \mathcal{M}_{ls}'(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho})\end{bmatrix}\begin{bmatrix}\mathbf{0} & \mathbf{Z}_{\boldsymbol{\varphi}_{im} l}'(\boldsymbol{\varphi})\\ \mathbf{Z}_{\boldsymbol{\varphi}_{ic} s}'(\boldsymbol{\varphi}) & \mathbf{0}\end{bmatrix}.$$

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

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

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

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

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

$$\begin{bmatrix}\mathbf{0} & \mathbf{Z}_{\boldsymbol{\varphi}_{ic} s}(\boldsymbol{\varphi})\\ \mathbf{Z}_{\boldsymbol{\varphi}_{im} l}(\boldsymbol{\varphi}) & \mathbf{0}\end{bmatrix}\mathcal{K}(0)\mathcal{K}^+(\boldsymbol{\varrho}_0)\begin{Bmatrix}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\varphi})\\ \operatorname{vec}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\varphi})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\varphi})-\mathbf{I}_N]\end{Bmatrix}$$

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

To prove Part 1c, note that

$$\begin{bmatrix}\mathbf{s}_{\boldsymbol{\varphi}_{ic} t}(\boldsymbol{\varphi},\mathbf{0})\\ \mathbf{s}_{\boldsymbol{\varphi}_{im} t}(\boldsymbol{\varphi},\mathbf{0})\end{bmatrix}=\begin{bmatrix}\mathbf{0} & \mathbf{Z}_{\boldsymbol{\varphi}_{ic} s}(\boldsymbol{\varphi})\\ \mathbf{Z}_{\boldsymbol{\varphi}_{im} l}(\boldsymbol{\varphi}) & \mathbf{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\varphi})\\ \operatorname{vec}[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\varphi})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\varphi})-\mathbf{I}_N]\end{bmatrix}, \tag{A25}$$

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

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

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

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

Proof of Proposition 14. The proof builds up on Proposition B1 in Supplemental Appendix B. Assuming covariance stationarity, the relationship vector of drift parameters $\boldsymbol{\tau}$ and the unconditional mean $\boldsymbol{\mu}$ is given by $(\mathbf{I}_N - \mathbf{A}_1 - \cdots - \mathbf{A}_p)\boldsymbol{\mu}$. Hence, the Jacobian from one vector of parameters to the other is

$$\frac{\partial\begin{pmatrix}\boldsymbol{\tau}\\ \mathbf{a}\end{pmatrix}}{\partial(\boldsymbol{\mu}', \mathbf{a}')} = \begin{pmatrix}\mathbf{I}_N - \mathbf{A}_1 - \cdots - \mathbf{A}_p & -\boldsymbol{\mu}'\otimes\mathbf{I}_N & \dots & -\boldsymbol{\mu}'\otimes\mathbf{I}_N\\ \mathbf{0} & \mathbf{I}_{N^2} & \dots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_{N^2}\end{pmatrix}.$$

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

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

so that

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\left. \begin{matrix} \begin{bmatrix} \boldsymbol{\Delta}_N'(\mathbf{I}_N \otimes \mathbf{J}^{-1\prime})(\boldsymbol{\Psi} \otimes \boldsymbol{\Psi}^{-1}) \\ \boldsymbol{\Psi}^{-1}\mathbf{E}_N' \end{bmatrix} \mathbf{E}_N\mathrm{M}_{sr} \\ \mathcal{M}_{rr} \end{matrix} \right\}$$

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

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

Let us now obtain the asymptotic covariance matrix of the restricted ML estimators of $(\mathbf{j}, \boldsymbol{\psi})$ which fix $\boldsymbol{\varrho}$ to its true values. Lemmas 4 and 5 in Supplemental Appendix B contain the inverses of $\mathcal{M}_{ss}$ and

$$\begin{bmatrix}(\boldsymbol{\Psi} \otimes \boldsymbol{\Psi}^{-1})(\mathbf{I}_N \otimes \mathbf{J}^{-1})\boldsymbol{\Delta}_N & \mathbf{E}_N\boldsymbol{\Psi}^{-1}\end{bmatrix},$$

respectively. Thus, the asymptotic covariance matrix of the RMLEs of $(\mathbf{j}, \boldsymbol{\psi})$ will be

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

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

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

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

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

The matrix $\boldsymbol{\Delta}_N'(\boldsymbol{\Psi}^{-1} \otimes \boldsymbol{\Psi})\boldsymbol{\Delta}_N$ is obviously diagonal. In turn, Proposition 5 in Magnus and Sentana (2020) implies that the matrix $\boldsymbol{\Delta}_N'(\mathbf{K}_{NN} + \boldsymbol{\Upsilon})\boldsymbol{\Delta}_N = \boldsymbol{\Delta}_N'\mathbf{K}_{NN}\boldsymbol{\Delta}_N + \boldsymbol{\Delta}_N'\boldsymbol{\Upsilon}\boldsymbol{\Delta}_N$ is the sum of a diagonal matrix $\boldsymbol{\Delta}_N'\boldsymbol{\Upsilon}\boldsymbol{\Delta}_N$ and a symmetric orthogonal matrix $\boldsymbol{\Delta}_N'\mathbf{K}_{NN}\boldsymbol{\Delta}_N$ whose only $N(N-1)$ nonzero elements are 1s in the positions corresponding to the ij and ji

elements of $\mathbf{J}$ for $j > i$. Therefore, although the parameters in the different columns of $\mathbf{J}$ would not be asymptotically orthogonal when $\mathbf{J}_0 = \mathbf{I}_N$, the dependence seems to be limited to pairs of elements $\{\mathbf{J}\}_{ij}$ and $\{\mathbf{J}\}_{ji}$.

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

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

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

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

and the same applies to the southwestern one by symmetry.

Turning now to the eastern block, we get

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

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

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

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

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

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

the second term in the sum reduces to

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

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

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

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

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

The comparison between (A27) and (A28) also indicates that the covariance between the ML estimators of $\boldsymbol{\psi}$ and $\boldsymbol{\varrho}$ is the same regardless of whether $\mathbf{j}$ is estimated or not. The same is true of the correction to the asymptotic covariance matrix of $\boldsymbol{\psi}$ resulting from estimating $\boldsymbol{\varrho}$. In contrast, $\boldsymbol{\Psi}\mathrm{M}_{ss}^{-1}\boldsymbol{\Psi}$ and

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

do not generally coincide unless $\mathbf{J}_0=\mathbf{I}_N$. □

References

Acharya, V., L. H. Pedersen, T. Philippon, and M. Richardson (2017), "Measuring systemic risk." *Review of Financial Studies*, 30, 2–47. [684]

Adrian, T. and M. K. Brunnermeier (2016), "CoVaR." *American Economic Review*, 106, 1705–1741. [684]

Amengual, D., M. Carrasco, and E. Sentana (2020), "Testing distributional assumptions using a continuum of moments." *Journal of Econometrics*, 218, 655–689. [685]

Amengual, D., G. Fiorentini, and E. Sentana (2013), "Sequential estimators of shape parameters in multivariate dynamic models." *Journal of Econometrics*, 177, 233–249. [684, 696, 698, 706, 727]

Amengual, D., G. Fiorentini, and E. Sentana (2021), "Moment tests of independent components." CEMFI Working Paper 2102. [718]

Amengual, D. and E. Sentana (2010), "A comparison of mean–variance efficiency tests." *Journal of Econometrics*, 154, 16–34. [709, 727]

Andrews, D. W. K. (1987), "Asymptotic results for generalized Wald tests." *Econometric Theory*, 3, 348–358. [687]

Angelini, G., E. Bacchiocchi, G. Caggiano, and L. Fanelli (2019), "Uncertainty across volatility regimes." *Journal of Applied Econometrics*, 34, 437–455. [712, 714, 715]

Bai, J. (2003), "Testing parametric conditional distributions of dynamic models." *Review of Economics and Statistics*, 85, 531–549. [684]

Bai, J. and C. Zhihong (2008), "Testing multivariate distributions in GARCH models." *Journal of Econometrics*, 143, 19–36. [684, 685]

Bickel, P. J. (1982), "On adaptive estimation." *Annals of Statistics*, 10, 647–671. [700]

Bierens, H. J. and L. Wang (2012), "Integrated conditional moment tests for parametric conditional distributions." *Econometric Theory*, 28, 328–362. [685]

Bollerslev, T. and J. M. Wooldridge (1992), "Quasi maximum likelihood estimation and inference in dynamic models with time-varying covariances." *Econometric Reviews*, 11, 143–172. [684, 691, 695, 696]

Bontemps, C. and N. Meddahi (2012), "Testing distributional assumptions: A GMM approach." *Journal of Applied Econometrics*, 27, 978–1012. [685]

Brunnermeier, M., D. Palia, K. A. Sastry, and C. A. Sims (2021), "Feedbacks: Financial markets and economic activity." *American Economic Review*, 111 (6), 1845–1879. [704]

Calzolari, G., G. Fiorentini, and E. Sentana (2004), "Constrained indirect estimation." *Review of Economic Studies*, 71, 945–973. [688]

Carriero, A., T. E. Clark, and M. Marcellino (2018), "Measuring uncertainty and its impact on the economy." *Review of Economics and Statistics*, 100, 799–815. [712]

Crowder, M. J. (1976), "Maximum likelihood estimation for dependent observations." *Journal of the Royal Statistical Society B*, 38, 45–53. [695, 697]

Davidson, R. and J. G. MacKinnon (1989), "Testing for consistency using artificial regressions." *Econometric Theory*, 5, 363–384. [687, 696]

Drost, F. C. and C. A. J. Klaassen (1997), "Efficient estimation in semiparametric GARCH models." *Journal of Econometrics*, 80, 193–221. [684, 701, 706]

Drost, F. C., C. A. J. Klaassen, and B. J. M. Werker (1997), "Adaptive estimation in time series models." *Annals of Statistics*, 25, 786–817. [684, 700]

Durbin, J. (1954), "Errors in variables." *Review International Statistical Institute*, 22, 23–32. [685]

Engle, R. F. and G. M. Gallo (2006), "A multiple indicators model for volatility using intra-daily data." *Journal of Econometrics*, 131, 3–27. [718]

Engle, R. F. and G. Gonzalez-Rivera (1991), "Semiparametric ARCH models." *Journal of Business and Economic Statistics*, 9, 345–360. [684]

Fiorentini, G., E. Sentana, and G. Calzolari (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. [698, 699, 707]

Fiorentini, G. and E. Sentana (2007), "On the efficiency and consistency of likelihood estimation in multivariate conditionally heteroskedastic dynamic regression models." CEMFI Working Paper 0713. [683]

Fiorentini, G. and E. Sentana (2019), "Consistent non-Gaussian pseudo maximum likelihood estimators." *Journal of Econometrics*, 213, 321–358. [685, 699, 702, 706, 709, 712, 717]

Fiorentini, G. and E. Sentana (2020), "Discrete mixtures of normals pseudo maximum likelihood estimators of structural vector autoregressions." CEMFI Working Paper 2023. [705]

Fiorentini, G. and E. Sentana (2021), "Supplement to 'Specification tests for non-Gaussian maximum likelihood estimators'." *Quantitative Economics Supplemental Material*, 12, https://doi.org/10.3982/QE1406. [692]

Francq, C. and J.-M. Zakoïan (2010), GARCH *Models: Structure, Statistical Inference and Financial Applications*. Wiley. [699]

Gallant, A. R. and G. Tauchen (1996), "Which moments to match?" *Econometric Theory*, 12, 657–681. [688]

Gillier, G. L. (2005), "A generalized error distribution." http://dx.doi.org/10.2139/ssrn.2265027. [692]

Golub, G. H. and C. F. van Loan (2013), *Matrix Computations*, fourth edition. Johns Hopkins. [703]

Gonzalez-Rivera, G. (1997), "A note on adaptation in GARCH models." *Econometric Reviews*, 16, 55–68. [699]

Gonzalez-Rivera, G. and F. C. Drost (1999), "Efficiency comparisons of maximum-likelihood-based estimators in GARCH models." *Journal of Econometrics*, 93, 93–111. [684, 698]

Gouriéroux, C., A. Monfort, and E. Renault (1993), "Indirect inference." *Journal of Applied Econometrics*, 8, S85–S118. [688]

Gouriéroux, C., A. Monfort, and A. Trognon (1984a), "Pseudo maximum likelihood methods: Theory." *Econometrica*, 52, 681–700. [718]

Gouriéroux, C., A. Monfort, and A. Trognon (1984b), "Pseudo maximum likelihood methods: Applications to Poisson models." *Econometrica*, 52, 701–720. [718]

Hafner, C. M. and J. V. K. Rombouts (2007), "Semiparametric multivariate volatility models." *Econometric Theory*, 23, 251–280. [684, 698]

Hausman, J. (1978), "Specification tests in econometrics." *Econometrica*, 46, 1273–1291. [685, 686, 691, 726]

Hausman, J., B. Hall, and Z. Griliches (1984), "Econometric models for count data with an application to the patents R&D relationship." *Econometrica*, 52, 909–938. [718]

Hausman, J. and W. Taylor (1981), "A generalised specification test." *Economics Letters*, 8, 239–245. [696]

Hodgson, D. J., O. Linton, and K. P. Vorkink (2002), "Testing the capital asset pricing model efficiently under elliptical symmetry: A semiparametric approach." *Journal of Applied Econometrics*, 17, 617–639. [693]

Hodgson, D. J. and K. P. Vorkink (2003), "Efficient estimation of conditional asset pricing models." *Journal of Business and Economic Statistics*, 21, 269–283. [684, 700]

Holly, A. (1982), "A remark on Hausman's specification test." *Econometrica*, 50, 749–759. [696, 704]

Holly, A. (1987), "Specification tests: An overview." In *Advances in Econometrics—Fifth Word Congress* (T. F. Bewley, ed.). Cambridge University Press. Chapter 2. [686, 690, 691]

Holly, A. and A. Monfort (1986), "Some useful equivalence properties of Hausman's test." *Economics Letters*, 20, 39–43. [689]

Jurado, K., S. C. Ludvigson, and S. Ng (2015), "Measuring uncertainty." *American Economic Review*, 105, 1177–1216. [712]

Kotz, S. (1975), "Multivariate distributions at a cross-road." In *Statistical Distributions in Scientific Work, Vol. I* (G. P. Patil, S. Kotz, and J. K. Ord, eds.), 247–270, Reidel. [699]

Lange, K. L., R. J. A. Little, and J. M. G. Taylor (1989), "Robust statistical modeling using the t distribution." *Journal of the American Statistical Association*, 84, 881–896. [699]

Lanne, M. (2006), "A mixture multiplicative error model for realized volatility." *Journal of Financial Econometrics*, 4, 594–616. [718]

Lanne, M., M. Meitz, and P. Saikkonen (2017), "Identification and estimation of non-Gaussian structural vector autoregressions." *Journal of Econometrics*, 196, 288–304. [686, 703, 712]

Ling, S. and M. McAleer (2003), "Asymptotic theory for a vector ARMA-GARCH model." *Econometric Theory*, 19, 280–310. [692]

Linton, O. (1993), "Adaptive estimation in ARCH models." *Econometric Theory*, 9, 539–569. [684, 700]

Ludvigson, S. C., S. Ma, and S. Ng (forthcoming), "Uncertainty and business cycles: Exogenous impulse or endogenous response?" *American Economic Journal: Macroeconomics*. [712]

Magnus, J. R. (1988), *Linear Structures*. Oxford University Press. [728]

Magnus, J. R., H. G. J. Pijls, and E. Sentana (2021), "The Jacobian of the exponential function." *Journal of Economic Dynamics and Control*, 127, 104122. [703]

Magnus, J. R. and E. Sentana (2020), "Zero-diagonality as a linear structure." *Economics Letters*, 196, 109513. [734, 735]

McDonald, J. B. and W. K. Newey (1988), "Partially adaptive estimation of regression models via the generalized *t* distribution." *Econometric Theory*, 4, 428–457. [692]

Mencía, J. and E. Sentana (2012), "Distributional tests in multivariate dynamic models with normal and Student t innovations." *Review of Economics and Statistics*, 94, 133–152. [684]

Newey, W. K. (1985), "Maximum likelihood specification testing and conditional moment tests." *Econometrica*, 53, 1047–1070. [685, 687]

Newey, W. K. and D. L. McFadden (1994), "Large sample estimation and hypothesis testing." In *Handbook of Econometrics, Vol. IV* (R. F. Engle and D. L. McFadden, eds.), 2111–2245, Elsevier. [686, 718]

Newey, W. K. and D. G. Steigerwald (1997), "Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models." *Econometrica*, 65, 587–599. [684]

Phillips, P. C. B. and S. N. Durlauf (1986), "Multiple time series regression with integrated processes." *Review of Economic Studies*, 53, 473–495. [704]

Reiss, P. (1983), "A note on the selection of parameters and estimators in the Hausman specification test." Research Paper 708, Stanford University Graduate School of Business. [687]

Ruud, P. A. (1984), "Tests of specification in econometrics." *Econometric Reviews*, 3, 211–242. [687, 690, 697]

Ruud, P. A. (2000), *An Introduction to Classical Econometric Theory*. Oxford University Press. [689, 691]

Sun, Y. and T. Stengos (2006), "Semiparametric efficient adaptive estimation of asymmetric GARCH models." *Journal of Econometrics*, 127, 373–386. [684, 701]

Tauchen, G. (1985), "Diagnostic testing and evaluation of maximum likelihood models." *Journal of Econometrics*, 30, 415–443. [685]

White, H. (1982), "Maximum likelihood estimation of misspecified models." *Econometrica*, 50, 1–25. [685, 687]

White, H. (1987), "Specification testing in dynamic models." In *Advances in Econometrics—Fifth Word Congress* (T. F. Bewley, ed.). Cambridge University Press. Chapter 1. [685]

White, H. (1994), *Estimation, Inference and Specification Analysis*. Cambridge University Press. [687]

Wu, D.-M. (1973), "Alternative tests of independence between stochastic regressors and disturbances." *Econometrica*, 41, 733–750. [685]

Co-editor Tao Zha handled this manuscript.

Manuscript received 12 August, 2019; final version accepted 13 December, 2020; available online 21 December, 2020.