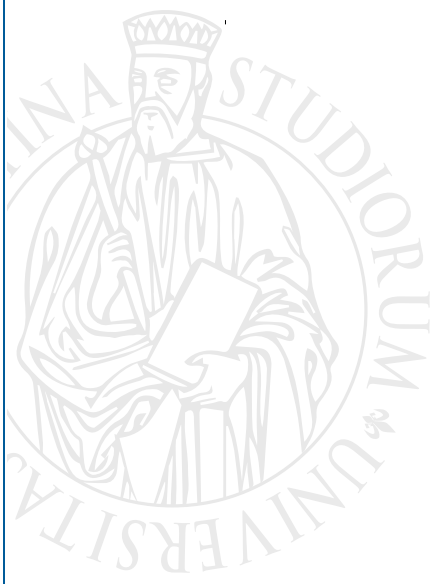
UNIVERSITÀ DEGLI STUDI FIRENZE
DISIA
DIPARTIMENTO DI STATISTICA, INFORMATICA, APPLICAZIONI "GIUSEPPE PARENTI"

Consistent non-Gaussian pseudo maximum likelihood estimators

Gabriele Fiorentini, Enrique Sentana



DISIA WORKING PAPER
2018/01

Consistent non-Gaussian pseudo maximum likelihood estimators*

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

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

Revised: January 2018

Abstract

We characterise the mean and variance parameters that distributionally misspecified maximum likelihood estimators can consistently estimate in multivariate conditionally heteroskedastic dynamic regression models. We also provide simple closed-form consistent estimators for the rest. The inclusion of means and the explicit coverage of multivariate models make our procedures useful not only for GARCH models but also in many empirically relevant macro and finance applications involving VARs and multivariate regressions. We study the statistical properties of our proposed consistent estimators, as well as their efficiency relative to Gaussian pseudo maximum likelihood procedures. Finally, we provide finite sample results through Monte Carlo simulations.

Keywords: Consistency, Efficiency, Misspecification

JEL: C13, C22, C32, C51

*This paper draws heavily on Fiorentini and Sentana (2007). In addition to those explicitly acknowledged there, we would like to thank Dante Amengual, Lorenzo Bretscher and Eric Gautier, as well as audiences at the 7th ICEEE (Messina, January 2017), the University of Liverpool Annual Econometrics Workshop (March 2017), the University of Southampton Finance and Econometrics Workshop (May 2017), the TSE Financial Econometrics Conference (May 2017), the 10^{th} Annual SoFiE Conference (New York, June 2017), the 25^{th} Finance Forum (Barcelona, July 2017) and the 70^{th} ESEM (Lisbon, August 2017) for useful comments and suggestions. Of course, the usual caveat applies. The second author gratefully acknowledges financial support from the Spanish Ministry of Economy & Competitiveness through grant 2014-59262 and the Santander - CEMFI Research Chair on Finance.

1 Introduction

As is well known, the Gaussian pseudo-maximum likelihood (PML) estimators advocated by Bollerslev and Wooldridge (1992) among many others remain root-T consistent for the mean and variance parameters of conditionally heteroskedastic dynamic regression models irrespective of the degree of asymmetry and kurtosis of the conditional distribution of the observed variables, so long as the first two moments are correctly specified and the fourth moments are bounded.

Nevertheless, many empirical researchers prefer to specify a non-Gaussian parametric distribution for the standardised innovations, which they use to estimate the conditional mean and variance parameters by maximum likelihood (ML). The dominant commercially available econometric packages have responded to this demand by offering ML procedures that either jointly estimate the parameters characterising the shape of the assumed distribution or allow the user to fix them to some pre-specified values. In particular, EVIEWS and STATA support Student t and Generalised Error distributions (GED) in univariate models (see the ARCH sections of IHS Global Inc (2015) and StataCorp LP (2015)), while STATA additionally allows for Student t innovations in multivariate ones (see the MARCH section of StataCorp LP (2015)).

However, while such ML estimators (and their Bayesian counterparts) will often yield asymptotically more efficient estimators than Gaussian PML if the assumed conditional distribution is correct, they may end up sacrificing consistency when it is not, as shown by Newey and Steiger-wald (1997) and Gouriéroux, Monfort and Zakoïan (2016). Intuitively, the reason is that mean, variances and covariances are natural location and scale measures for the multivariate normal distribution but not for others, so one cannot generally expect to consistently estimate the mean and covariance matrix of the standardised innovations under distributional misspecification.

For univariate GARCH models with *zero* conditional mean, Francq, Lepage and Zakoïan (2011) and Fan, Qi and Xiu (2014) have proposed modifications of parametric non-Gaussian pseudo ML estimators which achieve consistency even when the assumed distribution is misspecified. The purpose of this paper is to study in detail the statistical properties of the alternative consistent estimators we proposed in Fiorentini and Sentana (2007), whose closed-form expressions in terms of residuals readily generated by the commercial packages make them very easy to code. As we mentioned in Fiorentini and Sentana (2014) and formally prove in Appendix D, our estimators are asymptotically equivalent to the Fan, Qi and Xiu (2014) estimators for the univariate GARCH model with no mean they considered, which in turn are asymptotically equivalent to the Francq, Lepage and Zakoïan (2011) ones.

Nevertheless, our estimators remain consistent in multivariate models with non-zero means. The inclusion of means and the explicit coverage of multivariate contexts make our procedures

useful in many empirically relevant applications beyond ARCH models, which have been the motivating example for most of the existing work. In particular, our results apply to dynamic linear models such as VARs and multivariate regressions, which remain the workhorse in empirical macroeconomics and asset pricing contexts.[1] In addition, our estimators are not affected by the curse of dimensionality because they are effectively sample means of residuals, their squares and cross-products. Obviously, they also apply in univariate contexts as well as in static ones.

Another important differentiating feature of our analysis is that we consider not only ML estimators that fix the shape parameters but also procedures that jointly estimate them. In both cases, we characterise the conditional mean and variance parameters that these procedures can consistently estimate, providing closed-form estimators for the rest. In addition, we study the relative efficiency of these modified procedures vis a vis Gaussian PML estimators.

The rest of the paper is organised as follows. In section 2, we introduce our proposed estimators and study their asymptotic properties under misspecification, paying special attention to their efficiency relative to the Gaussian PML estimators. Then, we evaluate their finite sample properties by means of a Monte Carlo exercise in section 3. Finally we present our conclusions and discuss avenues for further research in section 4. Proofs and auxiliary results are gathered in appendices.

2 Distributional misspecification and parameter consistency

2.1 The estimated model

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

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

where $\boldsymbol{\mu}()$ and $vech\left[\boldsymbol{\Sigma}()\right]$ 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

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

[1] See Lanne, Meitz and Saikkonen (2017) and the references therein for recent examples of VAR models with non-Gaussian innovations.

To complete the model, a researcher needs to specify the conditional distribution of $\boldsymbol{\varepsilon}_t^*$. In Appendix C we study the general case. In view of the options that the dominant commercially available 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 (see appendix A.1 for a brief introduction to spherically symmetric distributions). The most prominent example is the standard multivariate normal, which we denote by $\boldsymbol{\eta} = \mathbf{0}$ without loss of generality. Another important example is a standardised multivariate Student t with ν degrees of freedom, or $i.i.d.\ t(\mathbf{0},\mathbf{I}_N,\nu)$ for short. As is well known, the multivariate t approaches the multivariate normal as $\nu \to \infty$, but has generally fatter tails. For that reason, we define η as $1/\nu$, which will always remain in the finite range $[0, 1/2)$ under our assumptions. Obviously, in the univariate case, any symmetric distribution, including the GED (also known as the Generalised Gaussian distribution), is spherically symmetric too.[2]

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

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

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

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

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

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

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

2.2 Asymptotic properties of the available pseudo maximum likelihood estimators

Let $L_T(\boldsymbol{\phi})$ denote the pseudo log-likelihood function of a sample of size T for the model discussed in the section 2.1, where $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\eta}')'$ are the $p + q$ parameters of interest, which we

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

assume variation free. As we mentioned in the introduction, the most popular commercially available software packages allow users to maximise $L_T(\boldsymbol{\phi})$ with respect to $\boldsymbol{\phi}$. But they also give them the option to fix the shape parameters to some admissible value $\bar{\boldsymbol{\eta}}$. In what follows, we will refer to $\hat{\boldsymbol{\phi}}_T' = (\hat{\boldsymbol{\theta}}_T', \hat{\boldsymbol{\eta}}_T')$ as the joint (or unrestricted) maximum likelihood estimator and to $\hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}})$ as the equality restricted one. An important special case arises when $\bar{\boldsymbol{\eta}} = \mathbf{0}$, in which case $\hat{\boldsymbol{\theta}}_T(\mathbf{0})$ coincides with the Gaussian PML estimator $\tilde{\boldsymbol{\theta}}_T$.

As a benchmark, in Appendix A.3 we provide the asymptotic distribution of these estimators under correct specification. In this section, though, we obtain their distribution under misspecification. We consider several cases, in decreasing order of agreement with the true distribution. We proceed as follows:

1. We transform the original set of conditional mean and variance parameters $\boldsymbol{\theta}$ into another set $\boldsymbol{\pi} = (\boldsymbol{\pi}_c', \boldsymbol{\pi}_i')'$ such that the inconsistencies resulting from misspecification affect the elements of $\boldsymbol{\pi}_i$ but not the rest.

2. We simultaneously estimate both subsets of parameters by (pseudo) maximum likelihood, which effectively allows the estimators of $\boldsymbol{\pi}_i$ to mop up the biases that would otherwise affect the estimators of $\boldsymbol{\pi}_c$. In this sense, it is important to emphasise that a restricted PMLE of $\boldsymbol{\pi}_c$ obtained by fixing $\boldsymbol{\pi}_i$ to its true value would be generally inconsistent. The same applies to GMM estimators that combine the non-Gaussian scores for $\boldsymbol{\pi}_c$ with the Gaussian scores for $\boldsymbol{\pi}_i$, unless the former do not depend on $\boldsymbol{\pi}_i$.

3. We discard the inconsistent pseudo-ML estimators of $\boldsymbol{\pi}_i$, replacing them by closed-form consistent estimators that use the Gaussian scores with respect to $\boldsymbol{\pi}_i$ evaluated at either $\hat{\boldsymbol{\pi}}_c$ or $\hat{\boldsymbol{\pi}}_c(\bar{\boldsymbol{\eta}})$ in a sequential GMM procedure. Given that our proposed estimators of those parameters are effectively sample means of residuals, their squares and cross-products, they are not affected by the curse of dimensionality. Importantly, we rely on standard GMM theory to derive the joint asymptotic distribution of the original estimators and the ones we propose by means of the usual sandwich formula, providing computationally reliable expressions for the expected Jacobian and the asymptotic covariance matrix of the influence functions.

4. We combine the consistent estimators of $\boldsymbol{\pi}_c$ and $\boldsymbol{\pi}_i$ thus obtained to recover consistent estimators of all the original parameters $\boldsymbol{\theta}$, employing the delta method to derive their asymptotic standard errors.

Throughout the paper, we use 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 maintain full generality. Primitive conditions for specific multivariate models can be found for example in Ling and McAleer (2003).

2.2.1 When the true distribution is spherically symmetric

Let us first consider situations in which the true distribution is *i.i.d.* spherical but different from the parametric one assumed for estimation purposes, which will often be chosen for convenience or familiarity. Note that this covers situations in which the conditional distribution is correctly specified, but we fix $\boldsymbol{\eta}$ to some $\bar{\boldsymbol{\eta}}$ which does not coincide with the true value $\boldsymbol{\eta}_0$.

In this case, all the parameters but one can be consistently estimated. To make this statement more precise, it is convenient to introduce the following reparametrisation:

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

$$\left.\begin{array}{c} \boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}_t(\boldsymbol{\vartheta}_c), \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \vartheta_i \boldsymbol{\Sigma}^\circ_t(\boldsymbol{\vartheta}_c) \end{array}\right\} \quad \forall t. \tag{4}$$

Expression (4) simply requires that one can construct pseudo-standardised residuals

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

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

Such a reparametrisation is not unique, since we can always multiply the overall scale parameter ϑ_i by some scalar positive smooth function of $\boldsymbol{\vartheta}_c$, $k(\boldsymbol{\vartheta}_c)$ say, and divide $\boldsymbol{\Sigma}^\circ_t(\boldsymbol{\vartheta}_c)$ by the same function without violating (4) or redefining $\boldsymbol{\vartheta}_c$. As we shall see in Proposition 2 below, a convenient normalisation for the purposes of simplifying some of the expressions would guarantee

$$E[\ln |\boldsymbol{\Sigma}^\circ_t(\boldsymbol{\vartheta}_c)| | \boldsymbol{\phi}_0] = k \quad \forall \boldsymbol{\vartheta}_c. \tag{5}$$

However, this is by no means essential since the estimators of $\boldsymbol{\vartheta}_c$ are numerically invariant to the choice of scaling, so their asymptotic covariance matrix is unaffected.

For the examples in section 2.1, reparametrisation 1 is as follows:

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

$$\left.\begin{array}{c} r_{Mt} = \mu_t(\boldsymbol{\vartheta}_c) + \vartheta_i^{1/2} \sigma^\circ_t(\boldsymbol{\vartheta}_c) \varepsilon^*_t, \\ \mu_t(\boldsymbol{\vartheta}_c) = \delta \sigma^\circ_t(\boldsymbol{\vartheta}_c), \\ \sigma^{\circ 2}_t(\boldsymbol{\vartheta}) = 1 + \gamma r^2_{Mt-1} + \beta \sigma^{\circ 2}_{t-1}(\boldsymbol{\vartheta}_c). \end{array}\right\} \tag{6}$$

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

Imposing (5) in this model would be tricky because we need to obtain

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

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

Multivariate market model We can write model (3) as

$$\mathbf{r}_t = \mathbf{a} + \mathbf{b}r_{Mt} + \vartheta_i^{1/2}\boldsymbol{\Omega}^{\circ 1/2}(\boldsymbol{\varpi})\boldsymbol{\varepsilon}_t^*. \tag{7}$$

The transformed conditional mean and variance parameters are $\boldsymbol{\vartheta}_c' = (\mathbf{a}', \mathbf{b}', \boldsymbol{\varpi}')$ and ϑ_i, where $\boldsymbol{\varpi}$ contains $N(N+1)/2-1$ elements. Following Amengual and Sentana (2010), we can achieve (5) by writing $\vartheta_i = |\boldsymbol{\Omega}|^{1/N}$ and $\boldsymbol{\Omega}^\circ(\boldsymbol{\varpi}) = \boldsymbol{\Omega}/|\boldsymbol{\Omega}|^{1/N}$ so as to achieve $|\boldsymbol{\Omega}^\circ(\boldsymbol{\varpi})| = 1\ \forall\boldsymbol{\varpi}$. Appendix A.5 discusses explicit parametrisations of $\boldsymbol{\Omega}^\circ(\boldsymbol{\varpi})$ that ensure this condition.

For simplicity, we shall define the pseudo-true values of $\boldsymbol{\vartheta}$ and $\boldsymbol{\eta}$ as consistent roots of the expectation of the spherical pseudo log-likelihood score, which under appropriate regularity conditions will maximise the expected value of the pseudo log-likelihood function.

The next proposition extends the first part of Theorem 1 in Newey and Steigerwald (1997) to multivariate dynamic models:

Proposition 1 *If (4) holds, and $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\varphi}_0$, is i.i.d. $s(\mathbf{0},\mathbf{I}_N)$, where $\boldsymbol{\varphi}$ includes $\boldsymbol{\vartheta}$ and the true shape parameters, but the spherical distribution assumed for estimation purposes does not necessarily nest the true density, then the pseudo-true value of the joint ML estimator of $\boldsymbol{\phi} = (\boldsymbol{\vartheta}_c', \vartheta_i, \boldsymbol{\eta})'$, $\boldsymbol{\phi}_\infty$, is such that $\boldsymbol{\vartheta}_{c\infty}$ is equal to the true value $\boldsymbol{\vartheta}_{c0}$.*

This result confirms that a spherically-based PMLE can consistently estimate all the conditional mean and variance parameters except the expected value of

$$\varsigma_t^\circ(\boldsymbol{\vartheta}_c) = [\mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\vartheta}_c)]'\boldsymbol{\Sigma}_t^{\circ -1}(\boldsymbol{\vartheta}_c)[\mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\vartheta}_c)] = \vartheta_i\varsigma_t(\boldsymbol{\vartheta}_c). \tag{8}$$

Figure 1 illustrates the extent of the inconsistency in estimating ϑ_i in a five-dimensional version of model (7) estimated by pseudo maximum likelihood assuming a multivariate Student t with unknown degrees of freedom when the true distribution is a discrete scale mixture of normals as a function of the mixing probability and the ratio of the variances of the two components. As can be seen from the depicted binding function, the relative bias $\vartheta_{i\infty}/\vartheta_{i0}$ can be substantial, especially when the mixture is such that both components are equally likely but one has a much larger variance than the other. Importantly, these relative biases are invariant to the true value of ϑ_i. More importantly, while they depend on the cross-sectional dimension N, they do not depend on the specification of the conditional mean vector or covariance matrix of the model.

In this context, in Fiorentini and Sentana (2007) we proposed to estimate ϑ_i by $\vartheta_{iT}(\hat{\boldsymbol{\vartheta}}_{cT})$, where

$$\vartheta_{iT}(\boldsymbol{\vartheta}_c) = \frac{1}{N}\frac{1}{T}\sum_{t=1}^{T}\varsigma_t^\circ(\boldsymbol{\vartheta}_c). \tag{9}$$

The rationale for this estimator comes from the fact that the Gaussian pseudo score for ϑ_i is simply

$$\mathbf{s}_{\vartheta_i t}(\boldsymbol{\vartheta}, \mathbf{0}) = \frac{1}{2\vartheta_i}\left[\varsigma_t(\boldsymbol{\vartheta}) - N\right] \tag{10}$$

whose expected value when evaluated at $\boldsymbol{\vartheta}_0$ is 0 because the expected value of $\varsigma_t^\circ(\boldsymbol{\vartheta}_{c0})$ in (8) is precisely $N\vartheta_{i0}$, and whose variance is proportional to the theoretical counterpart to Mardia's (1970) coefficient of multivariate excess kurtosis of $\boldsymbol{\varepsilon}_t^*$,

$$\kappa_0 = E(\varsigma_t^2)/[N(N+2)] - 1, \tag{11}$$

where $\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$ and the expectation is taken with respect to the true unconditional distribution of the innovations.

If we regard (10) as an additional influence function, we can rely on standard GMM arguments for just identified models to obtain the asymptotic variance of $\vartheta_{iT}(\hat{\boldsymbol{\vartheta}}_{cT})$, as well as its asymptotic covariances with the pseudo ML estimators $\hat{\boldsymbol{\vartheta}}_T$ and $\hat{\boldsymbol{\eta}}_T$ by means of the usual sandwich formula $\mathcal{C} = \mathcal{A}^{-1}\mathcal{B}\mathcal{A}'^{-1}$, where $\mathcal{A}$ is the expected Jacobian and $\mathcal{B}$ the asymptotic covariance matrix of the sample average of all the influence functions involved (see e.g. Newey and MacFadden (1994) for details). In doing so, though, we must carefully distinguish between ϑ_i, which is the parameter estimated with the misspecified log-likelihood function, and the parameter estimated with the Gaussian score, which we shall refer to $\bar{\vartheta}_i$ to avoid confusion. Specifically,

Proposition 2 *If (4) holds, and $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\varphi}_0$, is i.i.d. $s(\mathbf{0},\mathbf{I}_N)$ with $\kappa_0 < \infty$, where $\boldsymbol{\varphi}$ includes $\boldsymbol{\vartheta}$ and the true shape parameters, but the spherical distribution assumed for estimation purposes does not necessarily nest the true density, then*

1.

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}_{\boldsymbol{\phi}\boldsymbol{\phi}} & \mathbf{0} \\ \mathcal{A}_{\bar{\vartheta}_i\boldsymbol{\phi}} & \mathcal{A}_{\bar{\vartheta}_i\bar{\vartheta}_i} \end{pmatrix} = \begin{pmatrix} \mathcal{A}_{\boldsymbol{\vartheta}_c\boldsymbol{\vartheta}_c} & \mathcal{A}_{\boldsymbol{\vartheta}_c\vartheta_i} & \mathcal{A}_{\boldsymbol{\vartheta}_c\boldsymbol{\eta}} & \mathbf{0} \\ \mathcal{A}'_{\boldsymbol{\vartheta}_c\vartheta_i} & \mathcal{A}_{\vartheta_i\vartheta_i} & \mathcal{A}_{\vartheta_i\boldsymbol{\eta}} & 0 \\ \mathcal{A}'_{\boldsymbol{\vartheta}_c\boldsymbol{\eta}} & \mathcal{A}'_{\vartheta_i\boldsymbol{\eta}} & \mathcal{A}_{\boldsymbol{\eta}\boldsymbol{\eta}} & \mathbf{0} \\ \mathcal{A}_{\bar{\vartheta}_i\boldsymbol{\vartheta}_c} & 0 & \mathbf{0}' & \mathcal{A}_{\bar{\vartheta}_i\bar{\vartheta}_i} \end{pmatrix},$$

$$\mathcal{B} = \begin{pmatrix} \mathcal{B}_{\boldsymbol{\phi}\boldsymbol{\phi}} & \mathcal{B}_{\boldsymbol{\phi}\bar{\vartheta}_i} \\ \mathcal{B}'_{\boldsymbol{\phi}\bar{\vartheta}_i} & \mathcal{B}_{\bar{\vartheta}_i\bar{\vartheta}_i} \end{pmatrix} = \begin{pmatrix} \mathcal{B}_{\boldsymbol{\vartheta}_c\boldsymbol{\vartheta}_c} & \mathcal{B}_{\boldsymbol{\vartheta}_c\vartheta_i} & \mathcal{B}_{\boldsymbol{\vartheta}_c\boldsymbol{\eta}} & \mathcal{B}_{\boldsymbol{\vartheta}_c\bar{\vartheta}_i} \\ \mathcal{B}'_{\boldsymbol{\vartheta}_c\vartheta_i} & \mathcal{B}_{\vartheta_i\vartheta_i} & \mathcal{B}_{\vartheta_i\boldsymbol{\eta}} & \mathcal{B}_{\vartheta_i\bar{\vartheta}_i} \\ \mathcal{B}'_{\boldsymbol{\vartheta}_c\boldsymbol{\eta}} & \mathcal{B}'_{\vartheta_i\boldsymbol{\eta}} & \mathcal{B}_{\boldsymbol{\eta}\boldsymbol{\eta}} & \mathcal{B}_{\boldsymbol{\eta}\bar{\vartheta}_i} \\ \mathcal{B}'_{\boldsymbol{\vartheta}_c\bar{\vartheta}_i} & \mathcal{B}_{\vartheta_i\bar{\vartheta}_i} & \mathcal{B}'_{\boldsymbol{\eta}\bar{\vartheta}_i} & \mathcal{B}_{\bar{\vartheta}_i\bar{\vartheta}_i} \end{pmatrix},$$

with detailed expressions for all the elements in the proof.

2. *If in addition (5) holds, then both $\mathcal{A}$ and $\mathcal{B}$ become block diagonal between $\boldsymbol{\vartheta}_c$ and $(\vartheta_i, \boldsymbol{\eta}, \bar{\vartheta}_i)$.*

This proposition is very general, nesting several previous results in the literature. In particular, it generalises Proposition 5 in Amengual and Sentana (2010), who obtained expressions for $\mathcal{A}_{\phi\phi}$ and $\mathcal{B}_{\phi\phi}$ in a multivariate regression model under exactly the same type of misspecification. Obviously, it also applies under correct specification, in which case the information equality will imply that $\mathcal{A}_{\phi\phi} = \mathcal{B}_{\phi\phi}$ (see Proposition 8 in Appendix A.3).

Importantly, the above results also apply *mutatis mutandi* to restricted spherically-based ML estimators of $\boldsymbol{\vartheta}$ that fix $\boldsymbol{\eta}$ to some a priori chosen value $\bar{\boldsymbol{\eta}}$. In that case, we would simply need to replace $\vartheta_{i\infty}$ by $\vartheta_{i\infty}(\bar{\boldsymbol{\eta}})$ and eliminate the rows and columns corresponding to $\boldsymbol{\eta}$ from the $\mathcal{A}$ and $\mathcal{B}$ matrices. In that way, we would also nest Proposition 1 in Bollerslev and Wooldridge (1992), who focused on the Gaussian PML estimators (see Proposition 9 in Appendix A.4).

Finally, it is important to emphasise that reparametrisation 1 only plays an auxiliary role. After obtaining consistent estimators of the transformed parameters $\boldsymbol{\vartheta}_c$ and ϑ_i with the procedures we propose, it is straightforward to consistently estimate the original parameters $\boldsymbol{\theta}$ by inverting the mapping $\mathbf{r}_s(\boldsymbol{\theta})$. Moreover, the regularity of this mapping guarantees that we can obtain asymptotic standard errors for these consistent estimators by means of the delta method.

2.2.2 When the true distribution is asymmetric

Although the assumption of spherical symmetry is widespread in theoretical and empirical finance, its failure will invalidate the consistency results in Proposition 1. Fortunately, it is possible to find analogous results in the asymmetric case too, but at the cost of restricting further the set of parameters that can be consistently estimated under misspecification. Effectively, the number of parameters that are inconsistently estimated goes from 1 to $N(N+3)/2$, which represents a minimal increase in the univariate case (from 1 to 2). To make this statement more precise, it is convenient to introduce an alternative reparametrisation:

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

$$\left.\begin{array}{l} \boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}_t^{\diamond}(\boldsymbol{\psi}_c) + \boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)\boldsymbol{\psi}_{im} \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)\boldsymbol{\Psi}_{ic}\boldsymbol{\Sigma}_t^{\diamond 1/2\prime}(\boldsymbol{\psi}_c) \end{array}\right\} \quad \forall t. \tag{12}$$

This parametrisations simply requires the pseudo-standardised residuals

$$\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\psi}_c) = \boldsymbol{\Sigma}_t^{\diamond -1/2}(\boldsymbol{\psi}_c)[\mathbf{y}_t - \boldsymbol{\mu}_t^{\diamond}(\boldsymbol{\psi}_c)] \tag{13}$$

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

Again, (12) is not unique, since it continues to hold with the same $\boldsymbol{\psi}_c$ if we replace $\boldsymbol{\Psi}_{ic}$ by $\mathbf{K}^{-1/2}(\boldsymbol{\psi}_c)\boldsymbol{\Psi}_{ic}\mathbf{K}^{-1/2\prime}(\boldsymbol{\psi}_c)$ and $\boldsymbol{\psi}_{im}$ by $\mathbf{K}^{-1/2}(\boldsymbol{\psi}_c)\boldsymbol{\psi}_{im} - \mathbf{l}(\boldsymbol{\psi}_c)$, and adjust $\boldsymbol{\mu}_t^{\diamond}(\boldsymbol{\psi}_c)$ and $\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)$ accordingly, where $\mathbf{l}(\boldsymbol{\psi}_c)$ and $\mathbf{K}(\boldsymbol{\psi}_c)$ are a $N \times 1$ vector and a $N \times N$ positive definite matrix of smooth functions of $\boldsymbol{\psi}_c$, respectively. As we shall see in Proposition 4 below, a convenient normalisation for the purposes of simplifying some of the expressions would be such that:

$$\left.\begin{array}{r} E\left[\partial \boldsymbol{\mu}_t^{\diamond\prime}(\boldsymbol{\psi}_c)/\partial \boldsymbol{\psi}_c \cdot \boldsymbol{\Sigma}_t^{\diamond -1/2}(\boldsymbol{\psi}_c)\middle|\, \boldsymbol{\phi}_0\right] = \mathbf{0} \\ E\left\{\partial vec[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)]/\partial \boldsymbol{\psi}_c \cdot \left[\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\boldsymbol{\psi}_c)\right]\middle|\, \boldsymbol{\phi}_0\right\} = \mathbf{0} \end{array}\right\}. \tag{14}$$

However, this is by no means essential since the estimators of $\boldsymbol{\psi}_c$ are numerically invariant to these location-scale normalisations, so their asymptotic covariance matrix is unaffected.

For the examples in section 2.1, reparametrisation 2 is as follows:

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

$$\left.\begin{array}{c} r_{Mt} = \psi_{im}\mu_t^{\diamond}(\boldsymbol{\psi}_c) + \psi_{ic}^{1/2}\sigma_t^{\diamond}(\boldsymbol{\psi}_c)\varepsilon_t^*, \\ \mu_t^{\diamond}(\boldsymbol{\psi}_c) = \sigma_t^{\diamond}(\boldsymbol{\psi}_c), \\ \sigma_t^{\diamond}(\boldsymbol{\psi}_c) = 1 + \gamma r_{Mt-1}^2 + \beta\sigma_{t-1}^{\diamond 2}(\boldsymbol{\vartheta}_c). \end{array}\right\} \tag{15}$$

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

Multivariate market model We can write model (3) as

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

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

The next proposition provides the multivariate generalisation of Theorem 2 in Newey and Steigerwald (1997):

Proposition 3 *If (12) holds, and $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\varphi}_0$ is i.i.d. $(\mathbf{0}, \mathbf{I}_N)$, where $\boldsymbol{\varphi}$ includes $\boldsymbol{\psi}$ and the true shape parameters, but the distribution assumed for estimation purposes does not necessarily nest the true density, then the pseudo-true value of the joint ML estimator of $\boldsymbol{\phi} = (\boldsymbol{\psi}_c', \boldsymbol{\psi}_i', \boldsymbol{\eta})'$, $\boldsymbol{\phi}_\infty$, is such that $\boldsymbol{\psi}_{c\infty}$ is equal to the true value $\boldsymbol{\psi}_{c0}$.*

If we further assume that the true conditional mean of $\mathbf{y}_t$ is $\mathbf{0}$, and this restriction is imposed in estimation, then $\boldsymbol{\psi}_{im}$ becomes unnecessary, thereby generalising the second part of Theorem 1 in Newey and Steigerwald (1997).

In simple terms, Proposition 3 says that in general, a parametric ML estimator based on a spherically symmetric distribution cannot consistently estimate either the mean or the covariance matrix of the *i.i.d.* pseudo-standardised residuals $\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\psi}_{c0})$ in (13) when the true distribution is not spherically symmetric.

Figures 2A-B illustrate the extent of the inconsistency in estimating ψ_{im} and ψ_{ic} in a univariate model estimated by pseudo maximum likelihood assuming a Student t with unknown degrees of freedom when the true distribution is an admissible fourth-order Gram-Charlier expansion of the standard normal as a function of the skewness and kurtosis coefficients.[3] As can be seen from the depicted binding functions, the relative mean and scale biases $(\psi_{im\infty} - \psi_{im0})/\psi_{ic0}^{1/2}$ and $\psi_{ic\infty}/\psi_{ic0}$, respectively, can be substantial when the skewness increases and especially the kurtosis is large. Importantly, these relative biases are invariant to the true values of ψ_{im} and ψ_{ic}. More importantly, they do not depend on the specification of the rest of the conditional mean or variance of the model.

In this context, in Fiorentini and Sentana (2007) we proposed to estimate $\boldsymbol{\psi}_{im}$ and $\boldsymbol{\psi}_{ic}$ as $\boldsymbol{\psi}_{imT}(\hat{\boldsymbol{\psi}}_{cT})$ and $\boldsymbol{\psi}_{icT}(\hat{\boldsymbol{\psi}}_{cT})$, respectively, where

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

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

Once again, the rationale for these estimators arises from the fact the Gaussian pseudo scores for $\boldsymbol{\psi}_{im}$ and $\boldsymbol{\psi}_{ic}$ are simply:

$$\begin{aligned}\mathbf{s}_{\psi_{im}t}(\boldsymbol{\psi},\mathbf{0}) &= \frac{1}{2}\boldsymbol{\Psi}_{ic}^{-1/2\prime}\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\psi}),\\ \mathbf{s}_{\psi_{ic}t}(\boldsymbol{\psi},\mathbf{0}) &= \frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Psi}_{ic}^{-1/2\prime}\otimes\boldsymbol{\Psi}_{ic}^{-1/2\prime})vec\left\{\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\psi})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi})-\mathbf{I}_N\right\},\end{aligned}$$

where $\mathbf{D}_N$ is the duplication matrix (see Magnus and Neudecker, 1988), whose expected values at $\boldsymbol{\psi}_0$ are 0 because the expected value of

$$\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_i) = \boldsymbol{\Psi}_{ic}^{-1/2}(\boldsymbol{\psi}_{im0} - \boldsymbol{\psi}_{im}) + \boldsymbol{\Psi}_{ic}^{-1/2}\boldsymbol{\Psi}_{ic0}^{1/2}\boldsymbol{\varepsilon}_t^{*}$$

is 0 and the expected value of $\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_i)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_i)$ is $\mathbf{I}_N$ when $\boldsymbol{\psi}_i = \boldsymbol{\psi}_{i0}$.

If we regard $\mathbf{s}_{\psi_{im}t}(\boldsymbol{\psi},\mathbf{0})$ and $\mathbf{s}_{\psi_{ic}t}(\boldsymbol{\psi},\mathbf{0})$ as additional influence functions, we can again rely on standard GMM arguments for just identified models to obtain the asymptotic covariance matrix of $\boldsymbol{\psi}_{imT}(\hat{\boldsymbol{\psi}}_{cT})$ and $\boldsymbol{\psi}_{icT}(\hat{\boldsymbol{\psi}}_{cT})$, as well their asymptotic covariances with the pseudo ML estimators $\hat{\boldsymbol{\psi}}_T$ and $\hat{\boldsymbol{\eta}}_T$ by means of the usual sandwich formula. Nevertheless, we must carefully distinguish between $\boldsymbol{\psi}_i$, which are the parameters estimated with the misspecified log-likelihood function, and the namesake parameters estimated with the Gaussian score, which we shall refer to as $\bar{\boldsymbol{\psi}}_i$ to avoid confusion. Specifically,

[3] Since the magnitudes of the biases do not depend on the sign of the skewness coefficient, we only show the positive side of the admissible region. See Jondeau and Rockinger (2003) for a characterisation of the set of skewness and kurtosis values that give rise to a non-negative density for the fourth-order expansion.

Proposition 4 *If (12) holds, and* $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\varphi}_0$ *is i.i.d.* $(\mathbf{0},\mathbf{I}_N)$ *with bounded fourth moments, where* $\boldsymbol{\varphi}$ *includes* $\boldsymbol{\psi}$ *and the true shape parameters* $\boldsymbol{\rho}$*, but the distribution assumed for estimation purposes does not necessarily nest the true density, then:*

1.

$$\mathcal{A}=\begin{pmatrix} \mathcal{A}_{\phi\phi} & \mathbf{0} \\ \mathcal{A}_{\bar{\psi}_i\phi} & \mathcal{A}_{\bar{\psi}_i\bar{\psi}_i} \end{pmatrix}=\begin{pmatrix} \mathcal{A}_{\psi_c\psi_c} & \mathcal{A}_{\psi_c\psi_{im}} & \mathcal{A}_{\psi_c\psi_{ic}} & \mathcal{A}_{\psi_c\eta} & \mathbf{0} & \mathbf{0} \\ \mathcal{A}'_{\psi_c\psi_{im}} & \mathcal{A}_{\psi_{im}\psi_{im}} & \mathcal{A}_{\psi_{im}\psi_{ic}} & \mathcal{A}_{\psi_{im}\eta} & \mathbf{0} & \mathbf{0} \\ \mathcal{A}'_{\psi_c\psi_{ic}} & \mathcal{A}'_{\psi_{im}\psi_{ic}} & \mathcal{A}_{\psi_{ic}\psi_{ic}} & \mathcal{A}_{\psi_{ic}\eta} & \mathbf{0} & \mathbf{0} \\ \mathcal{A}'_{\psi_c\eta} & \mathcal{A}'_{\psi_{ic}\eta} & \mathcal{A}'_{\psi_{im}\eta} & \mathcal{A}_{\eta\eta} & \mathbf{0} & \mathbf{0} \\ \mathcal{A}_{\bar{\psi}_{im}\psi_c} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathcal{A}_{\bar{\psi}_{im}\bar{\psi}_{im}} & \mathbf{0} \\ \mathcal{A}_{\bar{\psi}_{ic}\psi_c} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathcal{A}_{\bar{\psi}_{ic}\bar{\psi}_{ic}} \end{pmatrix},$$

$$\mathcal{B}=\begin{pmatrix} \mathcal{B}_{\phi\phi} & \mathcal{B}_{\phi\bar{\psi}_i} \\ \mathcal{B}'_{\phi\bar{\psi}_i} & \mathcal{B}_{\bar{\psi}_i\bar{\psi}_i} \end{pmatrix}=\begin{pmatrix} \mathcal{B}_{\psi_c\psi_c} & \mathcal{B}_{\psi_c\psi_{im}} & \mathcal{B}_{\psi_c\psi_{ic}} & \mathcal{B}_{\psi_c\eta} & \mathcal{B}_{\psi_c\bar{\psi}_{im}} & \mathcal{B}_{\psi_c\bar{\psi}_{ic}} \\ \mathcal{B}'_{\psi_c\psi_{im}} & \mathcal{B}_{\psi_{im}\psi_{im}} & \mathcal{B}_{\psi_{im}\psi_{ic}} & \mathcal{B}_{\psi_{im}\eta} & \mathcal{B}_{\psi_{im}\bar{\psi}_{im}} & \mathcal{B}_{\psi_{im}\bar{\psi}_{ic}} \\ \mathcal{B}'_{\psi_c\psi_{ic}} & \mathcal{B}'_{\psi_{im}\psi_{ic}} & \mathcal{B}_{\psi_{ic}\psi_{ic}} & \mathcal{B}_{\psi_{ic}\eta} & \mathcal{B}_{\psi_{ic}\bar{\psi}_{im}} & \mathcal{B}_{\psi_{ic}\bar{\psi}_{ic}} \\ \mathcal{B}'_{\psi_c\eta} & \mathcal{B}'_{\psi_{im}\eta} & \mathcal{B}'_{\psi_{ic}\eta} & \mathcal{B}_{\eta\eta} & \mathcal{B}_{\eta\bar{\psi}_{im}} & \mathcal{B}_{\eta\bar{\psi}_{ic}} \\ \mathcal{B}'_{\psi_c\bar{\psi}_{im}} & \mathcal{B}'_{\psi_{im}\bar{\psi}_{im}} & \mathcal{B}'_{\psi_{ic}\bar{\psi}_{im}} & \mathcal{B}'_{\eta\bar{\psi}_{im}} & \mathcal{B}_{\bar{\psi}_{im}\bar{\psi}_{im}} & \mathcal{B}_{\bar{\psi}_{im}\bar{\psi}_{ic}} \\ \mathcal{B}'_{\psi_c\bar{\psi}_{ic}} & \mathcal{B}'_{\psi_{ic}\bar{\psi}_{im}} & \mathcal{B}'_{\psi_{ic}\bar{\psi}_{ic}} & \mathcal{B}'_{\eta\bar{\psi}_{ic}} & \mathcal{B}'_{\bar{\psi}_{im}\bar{\psi}_{ic}} & \mathcal{B}_{\bar{\psi}_{ic}\bar{\psi}_{ic}} \end{pmatrix},$$

with detailed expressions for all the elements in the proof.

2. *If in addition (14) holds, then both* $\mathcal{A}$ *and* $\mathcal{B}$ *become block diagonal between* $\boldsymbol{\psi}_c$ *and* $(\boldsymbol{\psi}_i,\boldsymbol{\eta},\bar{\boldsymbol{\psi}}_i)$*.*

As in section 2.2.1, the above results also apply *mutatis mutandi* to restricted spherically-based ML estimators of $\boldsymbol{\psi}$ that fix $\boldsymbol{\eta}$ to some a priori chosen value $\bar{\boldsymbol{\eta}}$. In that case, we would simply need to replace $\boldsymbol{\psi}_{im\infty}$ and $\boldsymbol{\psi}_{ic\infty}$ by $\boldsymbol{\psi}_{im\infty}(\bar{\boldsymbol{\eta}})$ and $\boldsymbol{\psi}_{ic\infty}(\bar{\boldsymbol{\eta}})$, respectively, and eliminate the rows and columns corresponding to $\boldsymbol{\eta}$ from the $\mathcal{A}$ and $\mathcal{B}$ matrices.

Once again, we should emphasise that reparametrisation 2 only plays an auxiliary role. After obtaining consistent estimators of the transformed parameters $\boldsymbol{\psi}_c$ and $\boldsymbol{\psi}_i$ with the procedures that we propose, it is straightforward to consistently estimate the original parameters $\boldsymbol{\theta}$ by inverting the mapping $\mathbf{r}_g(\boldsymbol{\theta})$. Moreover, the regularity of this mapping guarantees that we can obtain asymptotic standard errors for these consistent estimators by means of the delta method.

2.2.3 When the shape parameters are inequality restricted

So far, we have maintained the assumption that the shape parameters $\boldsymbol{\eta}$ are freely estimated. In several important cases, though, they will be estimated subject to inequality constraints. In the Student t case, for example, the reciprocal of the degrees of freedom η cannot be negative. This means that what we have called the unrestricted estimator $\hat{\eta}_T$ will in fact be characterised by Kuhn-Tucker (KT) conditions instead of the usual first-order ones. Somewhat surprisingly, such inequality constraints may imply that the whole of $\boldsymbol{\theta}$ will be consistently estimated despite distributional misspecification. The following proposition illustrates our claim:

Proposition 5 *1. Let $\boldsymbol{\phi}_\infty$ denote the pseudo-true values of the parameters $\boldsymbol{\theta}$ and η implied by a multivariate Student t log-likelihood function. If the true coefficient of multivariate excess kurtosis of $\boldsymbol{\varepsilon}_t^*$, κ_0, is not positive, then $\boldsymbol{\theta}_\infty = \boldsymbol{\theta}_0$ and $\eta_\infty = 0$.*

2. If κ_0 is strictly negative, then $\sqrt{T}\hat{\eta}_T = o_p(1)$ and $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) = o_p(1)$.

3. If κ_0 is exactly 0, then $\sqrt{T}\hat{\eta}_T$ will have an asymptotic normal distribution censored from below at 0, and $\tilde{\boldsymbol{\theta}}_T$ will be identical to $\hat{\boldsymbol{\theta}}_T$ with probability approaching 1/2. If in addition

$$\mathcal{A}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = E[[N + 2 - \varsigma_t(\boldsymbol{\theta}_0)]\{\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)|vec'[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)]\}\mathbf{Z}'_{dt}(\boldsymbol{\theta}_0)|\boldsymbol{\varphi}_0] = \mathbf{0}, \quad (18)$$

then $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) = o_p(1)$ the rest of the time.

Intuitively, the reason is that the score with respect to the reciprocal degrees of freedom parameter η evaluated under normality is proportional to the second generalised Laguerre polynomial

$$\varsigma_t^2(\boldsymbol{\theta})/4 - (N+2)\varsigma_t(\boldsymbol{\theta})/2 + N(N+2)/4 \quad (19)$$

in the multivariate Student t case (see Fiorentini, Sentana and Calzolari (2003)). In fact, as far as $\hat{\boldsymbol{\theta}}_T$ is concerned, Proposition 5 is valid not only for the Student t, but also for any pseudo ML estimator based on a symmetric generalised hyperbolic distribution (see Mencía and Sentana (2012) for details). In addition, it is also true for ML estimators based on fourth-order spherically symmetric expansions of the multivariate normal density, as well as on discrete scale mixtures of normals in which the odds ratio of the components is given (see Amengual and Sentana (2011)). More generally, it will be true for any leptokurtic spherical distribution that nests the normal as a limiting case, and which is such that the scores with respect to the shape parameters evaluated under Gaussianity are proportional to (19). In all those cases $\hat{\boldsymbol{\theta}}_T = \tilde{\boldsymbol{\theta}}_T$ whenever $\hat{\boldsymbol{\eta}}_T = \mathbf{0}$, which will occur when the sample coefficient of excess kurtosis of the innovations evaluated at the Gaussian PMLE is non-positive.

2.3 Efficiency comparisons

As explained by Fan, Qi and Xiu (2014), the equality restricted estimators $\hat{\boldsymbol{\psi}}_{cT}(\bar{\boldsymbol{\eta}})$ are not necessarily more efficient than the Gaussian PML estimators under misspecification, the obvious counterexample being an estimator that fixes $\boldsymbol{\eta}$ to a non-zero value when the true distribution is in fact Gaussian. If we knew the true distribution of $\boldsymbol{\varepsilon}_t^*$, but still decided to use the wrong log-likelihood function, we could minimise the asymptotic variance of $\hat{\boldsymbol{\psi}}_{cT}(\bar{\boldsymbol{\eta}})$ with respect to $\boldsymbol{\eta}$. In practice, we could estimate the required expressions by means of sample analogues, with the unknown innovations replaced by estimated ones, and then choose $\boldsymbol{\eta}$ as the minimiser of the estimated asymptotic variance, along the lines of Francq, Lepage and Zakoïan (2011). The comparison between the asymptotic covariance matrix of the resulting "optimised" estimator of $\boldsymbol{\psi}$ and the Gaussian PMLE deserves further investigation.

Given existing software, though, practitioners are more likely to simultaneously estimate $\boldsymbol{\eta}$ by maximum likelihood. In the next two subsections, we will make use of Proposition 2 above and Proposition 12 in Appendix C, which generalises Proposition 4 to non-spherical log-likelihoods, to compare the Gaussian and non-Gaussian PMLEs in detail for the univariate GARCH-M model (2) and the multivariate regression (3).

Univariate GARCH-M In Fiorentini and Sentana (2014), we investigated whether the joint and restricted estimators of $\boldsymbol{\psi}_c$ are more efficient than the Gaussian PML estimators in the univariate GARCH model with no mean considered by Fan, Qi and Xiu (2014) when the distribution used for estimation purposes is a Student t but the true distribution is a GED. Our results indicated that the Gaussian PMLE is always worse than the unrestricted ML estimator that simultaneously estimates η. Those results are in line with the local power comparisons in Fiorentini and Sentana (2010), whose focus is testing for mean and variance predictability in univariate models using non-Gaussian-based classical ML tests. The next proposition extends those results to the univariate GARCH-M model (15) without imposing any symmetry assumption on the distribution used for estimating purposes or the true one, a fact that we emphasise by denoting the shape parameters by $\boldsymbol{\varrho}$.[4]

Proposition 6 *1. Under standard regularity conditions, the asymptotic covariance matrix of the Pseudo ML estimator of $\boldsymbol{\psi}_c$ in model (15) for fixed values of $\boldsymbol{\varrho}$ is given by $S(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\cdot \mathbf{V}^{-1}_{\boldsymbol{\psi}_c}(\boldsymbol{\psi}_{c0};\boldsymbol{\varphi}_0)$, where*

$$S(\boldsymbol{\phi};\boldsymbol{\varphi}) = \frac{\left[\frac{\psi^2_{im\infty}(\boldsymbol{\varrho})}{\psi_{ic\infty}(\boldsymbol{\varrho})}\mathcal{M}^O_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) + \frac{2\psi_{im\infty}(\boldsymbol{\varrho})}{\psi^{1/2}_{ic\infty}(\boldsymbol{\varrho})}\mathcal{M}^O_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) + \mathcal{M}^O_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi})\right]}{\left[\frac{\psi^2_{im\infty}(\bar{\boldsymbol{\eta}})}{\psi_{ic\infty}(\bar{\boldsymbol{\eta}})}\mathcal{M}^H_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) + \frac{2\psi_{im\infty}(\boldsymbol{\varrho})}{\psi^{1/2}_{ic\infty}(\boldsymbol{\varrho})}\mathcal{M}^H_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) + \mathcal{M}^H_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi})\right]^2}, \tag{20}$$

$$\mathbf{V}_{\boldsymbol{\psi}_c}(\boldsymbol{\psi}_c;\boldsymbol{\varphi}) = V\left[\frac{1}{2\sigma^{\diamond 2}_t(\boldsymbol{\psi}_c)}\frac{\partial\sigma^{\diamond 2}_t(\boldsymbol{\psi}_c)}{\partial\boldsymbol{\psi}_c}\middle|\boldsymbol{\varphi}\right], \tag{21}$$

$$-E\left\{\frac{\partial^2\ln f[\epsilon_t[\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})];\boldsymbol{\varrho}]}{\partial\varepsilon\partial\varepsilon}\begin{bmatrix}1 & \epsilon_t[\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\\ \epsilon_t[\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})] & \epsilon^2_t[\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\end{bmatrix}\middle|\boldsymbol{\varphi}_0\right\}$$
$$= \begin{bmatrix}\mathcal{M}^H_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{M}^H_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi})\\ \mathcal{M}^H_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{M}^H_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi})\end{bmatrix} = \mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi}) \tag{22}$$

and

$$V\left\{\begin{pmatrix}\partial\ln f[\epsilon_t[\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})];\boldsymbol{\varrho}]/\partial\varepsilon\\ 1+\epsilon_t[\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\partial\ln f[\epsilon_t[\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})];\boldsymbol{\varrho}]/\partial\varepsilon\end{pmatrix}\middle|\boldsymbol{\varphi}_0\right\}$$
$$= \begin{bmatrix}\mathcal{M}^O_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{M}^O_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi})\\ \mathcal{M}^O_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{M}^O_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi})\end{bmatrix} = \mathcal{M}^O_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi}). \tag{23}$$

[4] An analogous proposition applies when both the assumed distribution and the true one are symmetric. The main difference is that ψ_{im} becomes consistent, so only ψ_{ic} needs to be replaced. Furthermore, some expressions simplify because $\mathcal{M}^O_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) = \mathcal{M}^H_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) = \phi = 0$ under symmetry. We discuss this case in detail at the end of the proof of Proposition 6.

2. In turn, the asymptotic covariance matrix of our consistent estimator of $\boldsymbol{\psi}_i$ is

$$\begin{bmatrix} \psi_{ic0} & \phi\psi_{ic0}^{3/2} \\ \phi\psi_{ic0}^{3/2} & (\varkappa-1)\psi_{ic0}^2 \end{bmatrix} + S(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{c}(\boldsymbol{\psi}_{i0})\mathbf{c}'(\boldsymbol{\psi}_{i0})\mathbf{W}'_{\boldsymbol{\psi}_c}(\boldsymbol{\psi}_{c0};\boldsymbol{\varphi}_0)\mathbf{V}^{-1}_{\boldsymbol{\psi}_c}(\boldsymbol{\psi}_{c0};\boldsymbol{\varphi}_0)\mathbf{W}_{\boldsymbol{\psi}_c}(\boldsymbol{\psi}_{c0};\boldsymbol{\varphi}_0), \tag{24}$$

where ϕ and $\varkappa$ are the coefficients of skewness and kurtosis, respectively, of the true standardised innovations, $\mathbf{c}(\boldsymbol{\psi}_i) = (\ \psi_{im}\ \ 2\psi_{ic}\ )'$ and

$$\mathbf{W}_{\boldsymbol{\psi}_c}(\boldsymbol{\psi}_c;\boldsymbol{\varphi}) = E\left[\frac{1}{2\sigma_t^{\diamond 2}(\boldsymbol{\psi}_c)}\frac{\partial\sigma_t^{\diamond 2}(\boldsymbol{\psi}_c)}{\partial\boldsymbol{\psi}_c}\middle|\boldsymbol{\varphi}\right]. \tag{25}$$

3. When $\boldsymbol{\varrho}$ is jointly estimated, exactly the same expressions apply if we replace $\boldsymbol{\varrho}$ by $\boldsymbol{\varrho}_\infty$.

Somewhat surprisingly, (24) provides an additive decomposition of the asymptotic covariance matrix of our consistent estimators of $\boldsymbol{\psi}_i$, $\boldsymbol{\psi}_{iT}(\hat{\boldsymbol{\psi}}_{cT})$. The first term corresponds to the asymptotic covariance matrix of our estimators of ψ_{im} and ψ_{ic} in (16) and (17) if we knew $\boldsymbol{\psi}_c$, while the second term reflects the additional sampling uncertainty resulting from the estimation of $\boldsymbol{\psi}_c$. Interestingly, though, this second term has rank 1 only.

Importantly, Proposition 6 also gives us the asymptotic covariance matrix of the corresponding Gaussian pseudo-ML estimators by setting $\boldsymbol{\varrho} = \mathbf{0}$. Thus, we can compare the efficiency of the two estimators of $\boldsymbol{\psi}_c$ by simply comparing the scalars (20). For the Gaussian pseudo-ML estimators, the relevant scalar simplifies to

$$S(\boldsymbol{\psi}_0,\mathbf{0};\boldsymbol{\varphi}_0) = \frac{\left(\frac{\psi_{im0}^2}{\psi_{ic0}} + \frac{2\psi_{im0}}{\psi_{ic0}^{1/2}}\phi + \varkappa - 1\right)}{\left(\frac{\psi_{im0}^2}{\psi_{ic0}} + 2\right)^2} \tag{26}$$

regardless of the true distribution. In contrast, we will usually have to resort to numerical quadrature to compute (20) in the general case.

Figure 3 displays the ratio of (20) to (26) for all admissible fourth-order Gram-Charlier expansions of the Gaussian density for a GARCH(1,1)-M model in which $\psi_{im0}/\psi_{ic0}^{1/2} = .05$ when the pseudo log-likelihood is based on the Student t, a design we will revisit in the Monte Carlo section. Although it is not clear a priori how the scaling factor $S(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ vary with $\boldsymbol{\varrho}$, the results clearly show that $\boldsymbol{\varrho}_\infty$ systematically leads to more efficient estimators than $\boldsymbol{\varrho} = \mathbf{0}$, at least for the parametric configuration we have chosen.

Similarly, if we subtract (24) from the asymptotic covariance matrix of the Gaussian PMLE of $\boldsymbol{\psi}_i$, we are left with

$$\mathbf{c}(\boldsymbol{\psi}_{i0})\mathbf{c}'(\boldsymbol{\psi}_{i0})\mathbf{W}'_{\boldsymbol{\psi}_c}(\boldsymbol{\psi}_{c0};\boldsymbol{\varphi}_0)\mathbf{V}^{-1}_{\boldsymbol{\psi}_c}(\boldsymbol{\psi}_{c0};\boldsymbol{\varphi}_0)\mathbf{W}_{\boldsymbol{\psi}_c}(\boldsymbol{\psi}_{c0};\boldsymbol{\varphi}_0)\left[S(\mathbf{0};\boldsymbol{\varrho}_0) - S(\bar{\boldsymbol{\eta}};\boldsymbol{\varrho}_0)\right].$$

Therefore, our proposed estimator of $\boldsymbol{\psi}_i$ will be more efficient than its Gaussian PMLE counterpart if and only if the Pseudo ML estimator of $\boldsymbol{\psi}_c$ is more efficient than the corresponding

Gaussian PMLE. As a result, the evidence presented in Figure 3 implies that our consistent estimators of ψ_{im} and ψ_{ic} will also be more efficient than the Gaussian PMLEs when the true distribution is a Gram-Charlier expansion of the Gaussian density.

These efficiency gains in estimating $\boldsymbol{\psi}_c$ and $\boldsymbol{\psi}_i$ will be inherited by our estimators of the original parameters, whose distribution we can obtain as a straightforward application of the delta method.[5] Specifically, the estimators of $\tau = \psi_{ic}^{-1/2}\psi_{im}$, $\alpha = \psi_{ic}\gamma$ and $\omega = \psi_{ic}$ obtained by combining the non-Gaussian PMLE of γ with our consistent estimators of ψ_{im} and ψ_{ic} will be more efficient that both their Gaussian PML counterparts and "mix and match" estimators that combine the non-Gaussian PMLE of γ with the Gaussian PMLEs of ψ_{im} and ψ_{ic}.

Finally, we have also assessed whether the consistent estimators of $\boldsymbol{\psi}_i$ in (16) and (17) entail any efficiency loss when the distribution assumed for estimation purposes is correct, in which case the MLE of these parameters is fully efficient. Figure 4 displays the asymptotic variances of the MLEs and Gaussian PMLEs of ψ_{im} and ψ_{ic} relative to the asymptotic variance of our consistent estimators for the case in which the true and estimated distribution is a Student t with unknown degrees of freedom. Given that the fourth moment of this distribution diverges to infinity as the number of degrees of freedom converges to 4 from above, the asymptotic efficiency loss of the Gaussian PMLEs of $\boldsymbol{\psi}_i$ can be made arbitrarily large, and the same is true of our consistent estimator of ψ_{ic}. But even in those circumstances, our proposed estimator of this parameter is substantially more efficient than the Gaussian one. In addition, the efficiency loss of (16) and (17) is much smaller for larger, more empirically realistic values of the degrees of freedom, and their advantages over the Gaussian PMLEs persists for longer. Those efficiency losses are not the same for the two parameters, though, being more pronounced for ψ_{ic} than ψ_{im}.

Multivariate regression In the context of the multivariate market model that we have used as our second illustrative example, Amengual and Sentana (2010) compared the efficiency of the non-Gaussian pseudo ML estimator of the conditional mean parameters $\mathbf{a}$ and $\mathbf{b}$ with the Gaussian estimator when the true conditional distribution is spherical and leptokurtic. Specifi-

[5] For example, in the case of a simple ARCH(1) model, the original parametrisation is

$$\sigma_t^2 = \psi_{ic}(1 + \psi_c x_{t-1}^2) = \psi_{ic} + \alpha x_{t-1}^2,$$

so that we need to find the asymptotic distribution of $\alpha(\psi_c, \psi_{ic}) = \psi_{ic}\psi_c$. But since the Jacobian of the transformation is

$$[\ \partial\alpha(\psi_c, \psi_{ic})/\partial\psi_c \quad \partial\alpha(\psi_c, \psi_{ic})/\partial\psi_c\ ] = (\ \psi_{ic} \quad \psi_c\ ),$$

the asymptotic variance of $\tilde{\alpha}_T = \hat{\psi}_c\psi_{ic}(\hat{\psi}_c)$ can be easily obtained as a quadratic form in the joint asymptotic covariance of $\hat{\psi}_{cT}$ and $\psi_{iT}(\hat{\psi}_{cT})$.

cally, their corollary 1 states that the former is more efficient that the latter if and only if

$$\frac{\vartheta_{i\infty}(\boldsymbol{\eta})}{\vartheta_{i0}} \frac{\mathrm{M}_{ll}^{O}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_0)}{\left[\mathrm{M}_{ll}^{H}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_0)\right]^2} < 1. \tag{27}$$

They found that when true distribution is a two-component scale mixture of normals but the distribution used for estimation purposes is a Student t, the pseudo ML estimator that jointly estimates $\boldsymbol{\eta}$ is always strictly more efficient than the Gaussian-based one. Figure 5A illustrates their results. However, they did not compare the efficiency of the estimators of the residual variance parameters, which are also of interest in empirical applications.

Let us parametrise $\boldsymbol{\Omega}^{\circ}$ in terms of the Cholesky decomposition $\boldsymbol{\Omega}_L\boldsymbol{\Omega}_D^{\circ}\boldsymbol{\Omega}_L'$, with $\boldsymbol{\Omega}_L$ unit lower triangular and $\boldsymbol{\Omega}_D^{\circ}$ diagonal. For convenience, we partition $\boldsymbol{\varpi}$ into $\boldsymbol{\varpi}_L = vecl(\boldsymbol{\Omega}_L)$ and $\boldsymbol{\varpi}_D = (\varpi_{D1},\ldots,\varpi_{DN-1})'$, which contains the $N-1$ free parameters that we use to ensure that $|\boldsymbol{\Omega}_D^{\circ}(\boldsymbol{\varpi}_D)| = 1$, as explained in Appendix A.5. Further, we can partition $\boldsymbol{\varpi}_L$ into $\boldsymbol{\varpi}_{L1},\boldsymbol{\varpi}_{L2}\ldots,\boldsymbol{\varpi}_{LN-1}$, of dimension $N-1, N-2,\ldots,1$ respectively, which contain the strict lower triangular elements in each of the columns of the matrix $\boldsymbol{\Omega}_L$. We can then prove that:

Proposition 7 *1. Under standard regularity conditions, the asymptotic covariance matrix of the Pseudo ML estimator of $\boldsymbol{\varpi}$ in model (15) is given by*

$$\frac{\mathrm{M}_{ss}^{O}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_0)}{\left[\mathrm{M}_{ss}^{H}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_0)\right]^2} \tag{28}$$

times the inverse of a $(\frac{1}{2}N^2+\frac{1}{2}N-1)\times(\frac{1}{2}N^2+\frac{1}{2}N-1)$ block diagonal matrix with respect to $\boldsymbol{\varpi}_D,\boldsymbol{\varpi}_{L1},\boldsymbol{\varpi}_{L2}\ldots,\boldsymbol{\varpi}_{LN-1}$which only depends on $\boldsymbol{\varpi}$, whose detailed expression we provide in the proof.

2. The asymptotic variance of our consistent estimator of ϑ_i is given by

$$\frac{N[(N+2)\kappa_0+2]}{N^2}\vartheta_{i0}^2 \tag{29}$$

3. When $\boldsymbol{\eta}$ is jointly estimated, the same expression apply if we replace $\bar{\boldsymbol{\eta}}$ by $\boldsymbol{\eta}_{\infty}$.

Given that this result applies to a Gaussian log-likelihood function too, in which case $\mathrm{M}_{ss}^{O}(\boldsymbol{\vartheta}_0,\mathbf{0};\boldsymbol{\varphi}_0) = 1+\kappa$ and $\mathrm{M}_{ss}^{H}(\boldsymbol{\vartheta}_0,\mathbf{0};\boldsymbol{\varphi}_0) = 1$, the first part of the proposition immediately implies that the non-Gaussian pseudo ML estimator of the variance parameters $\boldsymbol{\varpi}$ will be more efficient than the Gaussian estimator if and only if (28) is less than $1+\kappa$.

Figure 5B shows the ratio of (28) to $(1+\kappa)$ for all possible two-component scale mixture of normals when the assumed distribution is a Student t and η is simultaneously estimated. As can be seen, the Gaussian estimator is systematically dominated, except when the mixture is such that there is a small probability of drawing from a component with very small variance, i.e. the so-called inlier case in Amengual and Sentana (2011).

On the other hand, the second part of the proposition says that our proposed consistent estimator of the overall scale parameter is as asymptotically as efficient as the Gaussian PMLE. The difference with Proposition 6 is that the asymptotic covariance matrices are block diagonal between $\boldsymbol{\vartheta}_c$ and ϑ_i in this model.

Finally, it is also of some interest to assess the efficiency loss in re-estimating ϑ_i when the true distribution is indeed a Student t with unknown degrees of freedom. As in the univariate case, we can make this loss arbitrarily large by choosing the number of degrees of freedom arbitrarily close to 4. However, Figure 6 shows that the efficiency loss is more reasonable for more realistic values. Therefore, it seems to us that this loss is probably worth paying to ensure the consistency of the entire parameter vector.

3 Monte Carlo evidence

In this section, we assess the finite sample performance of the different estimators and testing procedures discussed above by means of some extensive Monte Carlo exercises.

Univariate GARCH-M In our first simulation exercise we consider the univariate GARCH-M model (2). As we saw before, this model can be easily written in terms of reparametrisation 2 with $\boldsymbol{\psi}_c = (\beta, \gamma)'$, ψ_{im} and ψ_{ic}, while in terms of reparametrisation 1 we have $\boldsymbol{\vartheta}_c = (\beta, \gamma, \delta)'$ and ϑ_i.

We generate random draws of ε_t^* from four different distributions: a standard normal, a standardised Student t with $\nu = 10$ degrees of freedom, a standardised symmetric fourth-order Gram-Charlier expansion with an excess kurtosis of 3.2, and another standardised Gram-Charlier expansion with skewness and excess kurtosis coefficients equal to -0.9 and 3.2, respectively. For a given distribution, random draws are obtained with the NAG library G05DDF and G05FFF functions, as detailed in Amengual, Fiorentini and Sentana (2013). In all four cases, we generate 10,000 samples of length 1,000 (plus another 100 for initialisation) with $\beta = 0.85$, $\alpha = 0.1$, $\tau = 0.05$ and $\omega = 1$, which implies that $\delta = \psi_{im} = 0.05$, $\gamma = 0.1$ and $\vartheta_i = \psi_{ic} = 1$. These parameter values ensure strict stationarity of the generating process.

We estimate the model parameters twice: first by Gaussian PML and then by maximising the log-likelihood function of the Student t distribution. In both cases, we initialise the conditional variance processes by setting $\sigma_1^{\circ 2} = (1 + \gamma r^2)/(1 - \beta)$, where $r^2 = \frac{1}{T}\sum_1^T r_t^2$, which corresponds to an estimate of the unconditional variance of $r_t/\omega^{1/2}$. In addition, we compute our closed-form consistent estimators both assuming symmetry of the true distribution, and also allowing for asymmetries. The Gaussian and Student t log-likelihood functions are maximised with a

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

We report the Monte Carlo medians and interquartile ranges of the estimators in Table 1. We also report results for the estimators of $\alpha$and τ in the canonical parametrisation of the model in equation (2). Under normality, the four estimators perform equally well in terms of small sample bias and sampling variability. When the innovations follow a Student t with 10 degrees of freedom, the ML estimators outperform the Gaussian PML estimators, as expected. On the other hand, our consistent estimators are very similar to the MLEs. Under the symmetric GC distribution, our consistent symmetric estimator is the best. In contrast, the Student t-based ML estimator of ω is remarkably biased with a Monte Carlo median equal to 1.4512, in broad agreement with the results displayed in Figure 2B. Of course, this bias is carried forward to the estimators of α and τ in the original parametrisation. Nevertheless, the Student t-based ML estimators of the consistently estimated parameters β, γ and δ are considerably more efficient than the Gaussian PMLE, as expected from Figure 3. Finally, when we draw the innovations from the skewed GC distribution, the Student t-based ML estimator of δ also shows large biases, in agreement with Figure 2A. This time our consistent asymmetric estimator is the best. Somewhat surprisingly, though, the estimator of ω that assumes symmetry does not seem to be much biased, probably because ω is large relative to δ. Once again, the Student t-based ML estimators of the consistently estimated parameters β and γ are more efficient than the Gaussian PMLEs. Overall, the univariate simulation exercise confirms the asymptotic results displayed in Figures 2A, 2B, 3 and 4.[6]

Multivariate market model In our second exercise, we consider the multivariate market model (3). Again, we consider several standardised multivariate distributions for $\boldsymbol{\varepsilon}_t^*$ including a multivariate Gaussian and a Student t with 8 degrees of freedom, a discrete scale mixture of two normals (DSMN) with mixing probability equal to 0.2 and variance ratio equal to 10, and an asymmetric Student t distribution with $\nu = 8$ and $\beta = -1000$ (see Amengual and Sentana (2010) for further details). For each distribution we generate 1,000 samples of dimension $N = 5$ and length $T = 500$ with $\mathbf{a} = .112\boldsymbol{\ell}_5$, $\mathbf{b} = \boldsymbol{\ell}_5$ and $\boldsymbol{\Omega} = \mathbf{D}^{1/2}\mathbf{R}\mathbf{D}^{1/2}$ with $\mathbf{D} = 3.136\ \mathbf{I}_5$ and the off diagonal terms of the correlation matrix $\mathbf{R}$ equal to 0.3, where $\boldsymbol{\ell}_5$ is a vector of fives ones. Finally, we generate the strongly exogenous regressor r_{Mt} in each replication as an $i.i.d.$ normal with an annual mean return of 7% and an annual standard deviation of 16%.

[6] The medians of the estimators of the shape parameter η in the four designs are .0010, .0961, .3592 and .3856, respectively. Those values compare favourably with the corresponding (pseudo) true values: 0, .1, .3615 and .3875. Further, they are precisely estimated, with interquartile ranges .0119, .0389, .0375 and .0386.

The Gaussian PML estimators of $\mathbf{a}$, $\mathbf{b}$ are very easy to obtain using equation by equation OLS. Similarly, the estimated covariance matrix of the OLS residuals yields the closed-form Gaussian PML estimator of $\mathbf{\Omega}$. The Student t-based ML estimator is computationally more demanding because we need to numerically maximise the criterion function with respect to $2N + N(N+1)/2 + 1 = 26$ parameters. For that reason, it is convenient to find very good initial values to start up the numerical maximisation of the joint log-likelihood function. In that regard, we first compute the method of moments estimator of the reciprocal of the degrees of freedom parameter η suggested by Fiorentini, Sentana and Calzolari (2003), which is based on the sample version of the coefficient of multivariate excess kurtosis (11). Next, we obtain a sequential ML estimator by maximising the Student likelihood function with respect to η keeping the other parameters fixed at their Gaussian PML estimates, as in Amengual, Fiorentini and Sentana (2013). Finally, we jointly maximise the Student likelihood function with a quasi-Newton method. Our numerical procedure is fast and reliable.

We report the results of the Monte Carlo experiment in Table 2 for several groups of parameters. Specifically, we exploit the exchangeability of our design to report medians and interquartile ranges of representative elements of the vectors of intercepts $\mathbf{a}$ and slopes $\mathbf{b}$, the global scale parameter $\vartheta_i = |\mathbf{\Omega}|^{1/N}$, and representative elements of the vectors $vecd(\mathbf{\Omega}^\circ)$, $vecl(\mathbf{\Omega}^\circ)$, $vecd(\mathbf{\Omega})$ and $vecl(\mathbf{\Omega})$. Under normality all estimators perform comparably, as expected from Proposition 5. Similarly, when the true distribution of the innovations is a Student t, the ML estimator is the best performer. Nevertheless, our proposed consistent estimator of the global scale parameter ϑ_i fares remarkably well, somewhat better than the corresponding Gaussian PMLE. On the other hand, when the innovations follow a DSMN distribution, the Student t-based PML estimators of ϑ_i and the covariance matrix of the original parametrisation $\mathbf{\Omega}$ are upward biased, confirming the theoretical results in Figure 1. In those circumstances, though, both the symmetric and asymmetric versions of our estimators perform very well, with the former moderately better than the latter, as expected. As already observed in the univariate simulation experiments, the Student t-based PML estimators of $\mathbf{a}$, $\mathbf{b}$, $\mathbf{\Omega}^\circ$ are substantially more efficient than the Gaussian PML counterparts in those circumstances. Finally, when the innovations follow an asymmetric Student t distribution, the symmetric Student t-based MLEs of the intercepts are noticeably biased. In contrast, the version of our estimator consistent under asymmetries and the Gaussian PMLE are the best. Once more, though, the Student t-based MLE of the slopes are not only consistent but they also dominate the Gaussian PMLEs. As expected, both the t-based MLE and our consistent under symmetry estimator of $\mathbf{\Omega}$ are biased. Overall, the multivariate simulation exercise confirms the asymptotic results displayed in Figures 1, 5A, 5B

and 6.[7]

4 Conclusions

We characterise the subset of conditional mean and variance parameters that distributionally misspecified non-Gaussian maximum likelihood estimators can consistently estimate in multivariate conditionally heteroskedastic dynamic regression models. We consider not only ML estimators that fix the parameters characterising the shape of the distribution but also procedures that jointly estimate them.

We then exploit the Gaussian scores of the parameters that are inconsistently estimated by the misspecified log-likelihood to derive simple closed-form consistent estimators for the rest. Our proposed estimators are in effect first and second sample moments of residuals readily generated by most software packages, which make them immune to the curse of dimensionality.

In addition, we show that when the true conditional distribution is either platykurtic or mesokurtic, in the sense that the coefficient of multivariate excess kurtosis is either negative or zero, pseudo ML estimators based on certain leptokurtic spherical distributions, including the multivariate Student t and indeed any symmetric generalised hyperbolic distribution, as well as some discrete scale mixtures and polynomial expansions of the multivariate normal, provide consistent estimators of all the parameters irrespective of the ellipticity of the true distribution because they converge to the Gaussian PML estimators.

It is important to emphasise that the reparametrisations that we consider only play an auxiliary role. After obtaining consistent estimators of all the transformed mean and variance parameters with the procedures that we propose, it is straightforward to consistently estimate the original parameters and to obtain their asymptotic standard errors by means of the delta method.

The inclusion of means and the explicit coverage of multivariate models make our procedures useful in many empirically relevant applications beyond ARCH models, which have been the motivating example for most of the existing work. In particular, our results apply to dynamic linear models such as VARs and multivariate regressions, which remain the workhorse in empirical macroeconomics and asset pricing contexts.

We study the statistical properties of our proposed consistent estimators. We also assess their efficiency relative to Gaussian pseudo maximum likelihood for two empirically relevant examples: a univariate GARCH-M and a multivariate market model. In accordance with earlier

[7] The medians of the estimators of the shape parameter η in the four designs are 0, .1240, .3344 and .2023, respectively. Those values compare favourably with the corresponding (pseudo) true values: 0, .125, .3344 and .2014. Further, they are precisely estimated, with interquartile ranges .0043, .0233, .0297 and .0322.

results in Amengual and Sentana (2010) and Fiorentini and Sentana (2010, 2014), it seems that our modified estimators are usually more efficient than their Gaussian PML counterparts, at least when the pseudo log-likelihood function is based on a Student t distribution whose shape parameter is simultaneously estimated. Those efficiency gains should translate into more precise estimators of transformations of the model parameters of empirical interest, such as impulse response functions in VAR contexts or Sharpe ratios and optimal mean variance weights in portfolio allocation ones, as well as more powerful tests.

In a detailed Monte Carlo experiment we confirm that the ML estimators of the transformed parameters we single out in our theoretical analysis are biased when the true distribution does not coincide with the one assumed for estimation purposes. Nevertheless, our simulation results also indicate that our proposed methods yield consistent estimators for all the parameters, and with lower Monte Carlo dispersion than their Gaussian counterparts.

It is difficult to find empirically relevant examples of models for which reparametrisation 1 does not hold, so our spherically symmetric results can be directly applied to most static and dynamic models. Our first example also shows that reparametrisation 2 applies seamlessly to univariate GARCH-M models, including sophisticated asymmetric alternatives such as the one in Sun and Stengos (2006). In turn, our second example confirms that this reparametrisation can also be readily applied to multivariate regression models. Unfortunately, the same is not generally true in multivariate GARCH models when the true distribution is asymmetric even if the conditional mean is 0. The constant conditional correlation (CCC) model of Bollerslev (1990), which assumes that $\boldsymbol{\Sigma}_t(\boldsymbol{\psi}_c, \boldsymbol{\psi}_{ic}) = \mathbf{S}_t(\boldsymbol{\psi}_c)\mathbf{R}\mathbf{S}_t(\boldsymbol{\psi}_c)$, where $\mathbf{S}_t$ is a positive diagonal matrix, $\boldsymbol{\psi}_{ic} = vecl(\mathbf{R})$ and $\mathbf{R}$ a correlation matrix, provides an important exception.[8] In most other models, though, we may need to artificially augment the original parametrisation with $\boldsymbol{\psi}_{ic}$ and $\boldsymbol{\psi}_{im}$ even though we know that $\boldsymbol{\psi}_{im0} = \mathbf{0}$ and $\boldsymbol{\psi}_{ic0} = vech(\mathbf{I}_N)$, which might lead to a substantial efficiency cost. Furthermore, in doing so, we must guarantee that the parameters $\boldsymbol{\psi}_c$ remain identified (see Newey and Steigerwald (1997) and Gouriéroux, Monfort and Zakoïan (2016) for a detailed discussion of these issues in univariate and multivariate models, respectively). Assessing the efficiency costs of estimating those overparametrised models relative to using Gaussian PML estimators in the original model would constitute a valuable addition.

In a univariate context with bounded fourth moments, Meddahi and Renault (1998) proposed optimal GMM estimators that combine the Gaussian scores with an optimal weighting matrix, which generally differs from the one implicit in the Gaussian PMLE. It would also be interesting to study the efficiency properties of our procedures relative to a multivariate generalisation of

[8] Ling and McAleer's (2003) generalisation of the CCC model and example 1 in Hafner and Rombouts' (2007) are other examples of multivariate models that can also be directly written using analogous reparametrisations.

theirs.

When the true innovations have unbounded fourth moments, the variance of the Gaussian scores for scale will be unbounded too, and the asymptotic distribution of our consistent estimators will be non-standard, a property shared with the Gaussian PMLE and the Meddahi and Renault (1998) procedures (see Hall and Yao (2003)). Alternative "robust" consistent estimators such as multivariate versions of the ones mentioned by Andrews (2014), Francq and Zakoïan (2014) and Ling and Zhu (2014) would prove useful in those circumstances.

A comparison of the sequential estimators of the shape parameters discussed in Amengual, Fiorentini and Sentana (2013), which keep $\boldsymbol{\theta}$ fixed at the Gaussian PMLEs, with an analogous sequential procedure which instead keeps them fixed at the consistent estimators we have studied in this paper would be worthwhile too.

Finally, one of the reasons why practitioners prefer to use non-Gaussian distributions for estimating GARCH models is that they are often not only interested in the conditional variance of the process, but also in other features of the conditional distribution. For example, they might be interested in its quantiles, which are required for the computation of commonly used risk management measures such as V@R, or the probability of the joint occurrence of several negative events, which is relevant for systemic risk measures. In contrast, the existing literature, including our paper, focuses mostly on parameter estimation. An evaluation of the consequences that the different estimation procedures which we consider have for such empirically relevant functionals of the conditional distribution constitutes another fruitful avenue for future research.

References

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

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

Amengual, D. and Sentana (2011): "Inference in multivariate dynamic models with elliptical innovations", mimeo, CEMFI.

Andrews, B. (2014): "Comment", *Journal of Business and Economic Statistics* 32, 191-193.

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

Bollerslev, T. (1990): "Modelling the coherence in short-run nominal exchange rates: a multivariate Generalized ARCH model", *Review of Economics and Statistics* 72, 498-505.

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

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

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

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

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

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

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

Fiorentini, G., and Sentana, E. (2010): "New testing approaches for mean-variance predictability", mimeo, CEMFI.

Fiorentini, G. and Sentana, E. (2014): "Comment", *Journal of Business and Economic Statistics* 32, 193-198.

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

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

Francq, C. and Zakoïan, J-M. (2014): "Comment", *Journal of Business and Economic Statistics* 32, 198-201.

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

Gouriéroux, C., Monfort, A. and Zakoïan, J.-M. (2016): "Pseudo-maximum likelihood and Lie groups of linear transformations", mimeo, CREST.

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

Hall, P. and Yao, Q. (2003): "Inference in ARCH and GARCH models with heavy-tailed errors", *Econometrica* 71, 285-317.

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

Jondeau, E. and Rockinger, M. (2003): "Conditional volatility, skewness and kurtosis: Existence, persistence and comovements", *Journal of Economics Dynamics and Control* 27, 1699-1737.

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

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

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

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

Ling, S. and Zhu, K. (2014): "Comment", *Journal of Business and Economic Statistics* 32, 202-203.

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

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

Mardia, K.V. (1970): "Measures of multivariate skewness and kurtosis with applications", *Biometrika* 57, 519-530.

Meddahi, N. and E. Renault (1998). "Quadratic M-estimators for ARCH-type processes", CIRANO Working paper 98s-29.

Mencía, J. and Sentana, E. (2012): "Distributional tests in multivariate dynamic models

with Normal and Student t innovations", *Review of Economics and Statistics* 94, 133-152.

NAG (2001): *NAG Fortran 77 Library Mark 19 Reference Manual.*

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

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

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

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

Appendices

A Auxiliary results

A.1 Some useful distribution results

A spherically symmetric random vector of dimension N, $\boldsymbol{\varepsilon}_t^\circ$, is fully characterised in Theorem 2.5 (iii) of Fang, Kotz and Ng (1990) as $\boldsymbol{\varepsilon}_t^\circ = e_t \mathbf{u}_t$, where $\mathbf{u}_t$ is uniformly distributed on the unit sphere surface in $\mathbb{R}^N$, and e_t is a non-negative random variable independent of $\mathbf{u}_t$, whose distribution determines the distribution of $\boldsymbol{\varepsilon}_t^\circ$. The variables e_t and $\mathbf{u}_t$ are referred to as the generating variate and the uniform base of the spherical distribution. Assuming that $E(e_t^2) < \infty$, we can standardise $\boldsymbol{\varepsilon}_t^\circ$ by setting $E(e_t^2) = N$, so that $E(\boldsymbol{\varepsilon}_t^\circ) = \mathbf{0}$, $V(\boldsymbol{\varepsilon}_t^\circ) = \mathbf{I}_N$. Specifically, if $\boldsymbol{\varepsilon}_t^\circ$ is distributed as a standardised multivariate Student t random vector of dimension N with ν_0 degrees of freedom, then $e_t = \sqrt{(\nu_0 - 2)\zeta_t/\xi_t}$, where ζ_t is a chi-square random variable with N degrees of freedom, and ξ_t is an independent Gamma variate with mean $\nu_0 > 2$ and variance $2\nu_0$. If we further assume that $E(e_t^4) < \infty$, then the coefficient of multivariate excess kurtosis κ_0, which is given by $E(e_t^4)/[N(N+2)] - 1$, will also be bounded. For instance, $\kappa_0 = 2/(\nu_0 - 4)$ in the Student t case with $\nu_0 > 4$, and $\kappa_0 = 0$ under normality. In this respect, note that since $E(e_t^4) \geq E^2(e_t^2) = N^2$ by the Cauchy-Schwarz inequality, with equality if and only if $e_t = \sqrt{N}$ so that $\boldsymbol{\varepsilon}_t^\circ$ is proportional to $\mathbf{u}_t$, then $\kappa_0 \geq -2/(N+2)$, the minimum value being achieved in the uniformly distributed case.

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

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

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

where $\mathbf{K}_{mn}$ is the commutation matrix of orders m and n (see e.g. Magnus and Neudecker (1987)).

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

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

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

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

while

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

where

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

and

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

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

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

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

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

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

where

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

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

Note that $\partial\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, $\partial^2 d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ and $\partial^2\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ depend on the dynamic model specification, while $\partial^2 g(\varsigma,\eta)/(\partial\varsigma)^2$, $\partial^2 g(\varsigma,\eta)/\partial\varsigma\partial\boldsymbol{\eta}'$ and $\partial g(\varsigma,\eta)/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'$ depend on the specific spherical distribution assumed for estimation purposes (see Fiorentini, Sentana and Calzolari (2003) for expressions for $\delta(\varsigma_t,\boldsymbol{\eta})$, $c(\boldsymbol{\eta})$, $g(\varsigma_t,\boldsymbol{\eta})$ and its derivatives in the multivariate Student t case, Amengual and Sentana (2010) for the Kotz distribution (see Kotz (1975)) and discrete scale mixture of normals, and Amengual, Fiorentini and Sentana (2013) for polynomial expansions).

A.3 Asymptotic distribution under correct specification

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

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

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

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

Proposition 1 in Fiorentini and Sentana (2007), which generalises Propositions 3 in Lange, Little and Taylor (1989), 1 in Fiorentini, Sentana and Calzolari (2003) and 5.2 in Hafner and Rombouts (2007), provides detailed expressions for $\mathcal{M}(\boldsymbol{\phi})$. We reproduce it here to facilitate its comparison to Proposition 2:

Proposition 8 *If $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}$ is i.i.d. $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta})$ with density* $\exp[c(\boldsymbol{\eta})+g(\varsigma_t,\boldsymbol{\eta})]$*, then*

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

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

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

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

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

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

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

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

A.4 Gaussian pseudo maximum likelihood estimators

Let $\tilde{\boldsymbol{\theta}}_T=\arg\max_{\boldsymbol{\theta}}L_T(\boldsymbol{\theta},\mathbf{0})$ denote the Gaussian PML estimator of $\boldsymbol{\theta}$. As we mentioned in the introduction, $\tilde{\boldsymbol{\theta}}_T$ remains root-T consistent for $\boldsymbol{\theta}_0$ under correct specification of $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ even though the true conditional distribution of $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}_0$ is neither Gaussian nor spherical, provided that it has bounded fourth moments. The proof is based on the fact that in those circumstances, the pseudo log-likelihood score, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})$, is also a vector martingale difference sequence when evaluated at $\boldsymbol{\theta}_0$, a property that inherits from $\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})$. This property is preserved even when the standardised innovations, $\boldsymbol{\varepsilon}_t^*$, are not stochastically independent of I_{t-1}. The asymptotic distribution of the PML estimator of $\boldsymbol{\theta}$ is stated in the following result, which specialises Proposition 1 in Bollerslev and Wooldridge (1992) to models with *i.i.d.* innovations with shape parameters $\boldsymbol{\rho}$:

Proposition 9 *Assume that the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied.*

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

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

and

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

where

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

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

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

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

which only depends on the true distribution through the population coefficient of multivariate excess kurtosis κ_0*.*

A.5 Explicit parametrisation of the residual covariance matrix

Let us start with the simplest possible example in which $\boldsymbol{\Omega}$ is assumed diagonal. In that case, we can easily achieve $|\boldsymbol{\Omega}_D^\circ(\boldsymbol{\varpi})| = 1$ by writing $\omega_{jj} = \vartheta_i \exp(\varpi_{Dj})$ for $j = 1, \ldots, N-1$ and $\omega_{NN} = \vartheta_i \exp\left(-\sum_{j=1}^{N-1} \varpi_{Dj}\right)$. Thus, the Jacobian of $vecd(\boldsymbol{\Omega})$ with respect to ϑ_i will be

$$vecd(\boldsymbol{\Omega}_D^\circ) = [\exp(\varpi_{D1}), \ldots, \exp(\varpi_{DN-1}), \exp\left(-\textstyle\sum_{j=1}^{N-1} \varpi_{Dj}\right)]'$$

while the one with respect to $\boldsymbol{\varpi}_D = (\varpi_1, \ldots, \varpi_{N-1})'$ will be ϑ_i times

$$\frac{\partial vecd(\boldsymbol{\Omega}_D^\circ)}{\partial \boldsymbol{\varpi}_D'} = \begin{bmatrix} diag[\exp(\varpi_{D1}), \ldots, \exp(\varpi_{DN-1})] \\ -\exp\left(-\sum_{j=1}^{N-1} \varpi_{Dj}\right) \ell_{N-1}' \end{bmatrix}, \tag{A19}$$

where ℓ_{N-1} denotes a vector of $N-1$ ones and $diag(\boldsymbol{\varpi}_D)$ a square diagonal matrix with the elements of the vector $\boldsymbol{\varpi}_D$ along the main diagonal. Obviously, in the special case of $\boldsymbol{\Omega}$ scalar, then $\boldsymbol{\Omega}_D^\circ = \mathbf{I}_N$ and $\boldsymbol{\varpi}_D$ drops out.

Let us now move to the case in which, other than being positive (semi)definite, $\boldsymbol{\Omega}$ is completely unrestricted. Let $\boldsymbol{\Omega} = \boldsymbol{\Omega}_L \boldsymbol{\Omega}_D \boldsymbol{\Omega}_L'$ denote the Cholesky factorisation of the matrix $\boldsymbol{\Omega}$, with $\boldsymbol{\Omega}_D$ diagonal and $\boldsymbol{\Omega}_L$ unit lower triangular. Given that $|\boldsymbol{\Omega}_L| = 1$, we will have that $|\boldsymbol{\Omega}| = |\boldsymbol{\Omega}_D|$ so we can ensure $|\boldsymbol{\Omega}^\circ(\boldsymbol{\varpi})| = 1$ by parametrising $\boldsymbol{\Omega}_D$ as in the diagonal case above.

Using the product rule for differentials, we get that

$$d\boldsymbol{\Omega} = \mathbf{d}\boldsymbol{\Omega}_L \cdot \boldsymbol{\Omega}_D \boldsymbol{\Omega}_L' + \boldsymbol{\Omega}_L \cdot d\boldsymbol{\Omega}_D \cdot \boldsymbol{\Omega}_L' + \boldsymbol{\Omega}_L \boldsymbol{\Omega}_D \cdot d\boldsymbol{\Omega}_L',$$

whence

$$\begin{aligned} dvec(\boldsymbol{\Omega}) &= (\boldsymbol{\Omega}_L \boldsymbol{\Omega}_D \otimes \mathbf{I}_N) dvec(\boldsymbol{\Omega}_L) + (\boldsymbol{\Omega}_L \otimes \boldsymbol{\Omega}_L) dvec(\boldsymbol{\Omega}_D) + (\mathbf{I}_N \otimes \boldsymbol{\Omega}_L \boldsymbol{\Omega}_D) dvec(\boldsymbol{\Omega}_L') \\ &= [(\boldsymbol{\Omega}_L \boldsymbol{\Omega}_D \otimes \mathbf{I}_N) + (\mathbf{I}_N \otimes \boldsymbol{\Omega}_L \boldsymbol{\Omega}_D) \mathbf{K}_{NN}] dvec(\boldsymbol{\Omega}_L) + (\boldsymbol{\Omega}_L \otimes \boldsymbol{\Omega}_L) dvec(\boldsymbol{\Omega}_D) \\ &= (\mathbf{I}_N + \mathbf{K}_{NN})(\boldsymbol{\Omega}_L \boldsymbol{\Omega}_D \otimes \mathbf{I}_N) dvec(\boldsymbol{\Omega}_L) + (\boldsymbol{\Omega}_L \otimes \boldsymbol{\Omega}_L) dvec(\boldsymbol{\Omega}_D). \end{aligned}$$

Let $\mathbf{S}_N$ the unique $\frac{1}{2}N(N-1)\times N^2$ matrix which transforms $vecl(\mathbf{\Omega}_L)$ into $vec(\mathbf{\Omega}_L-\mathbf{I}_N)$ as $vec(\mathbf{\Omega}_L-\mathbf{I}_N)=\mathbf{S}'_N vecl(\mathbf{\Omega}_L)$, where $vecl(\mathbf{\Omega}_L)$ is the $\frac{1}{2}N(N-1)\times 1$ vector that contains the elements in the strict lower triangle of $\mathbf{\Omega}_L$ stacked by columns (see Magnus (1988)). Given that $dvec(\mathbf{\Omega}_L)=dvec(\mathbf{\Omega}_L-\mathbf{I}_N)=\mathbf{S}'_N dvecl(\mathbf{\Omega}_L)$, we can finally write

$$dvec(\mathbf{\Omega})=(\mathbf{I}_N+\mathbf{K}_{NN})(\mathbf{\Omega}_L\mathbf{\Omega}_D\otimes\mathbf{I}_N)\mathbf{S}'_N dvecl(\mathbf{\Omega}_L)+(\mathbf{\Omega}_L\otimes\mathbf{\Omega}_L)\mathbf{E}'_N dvecd(\mathbf{\Omega}_D),$$

where $\mathbf{E}_N$ is the unique diagonalisation matrix which transforms $vecd(\mathbf{\Omega}_D)$ into $vec(\mathbf{\Omega}_D)$ as $vec(\mathbf{\Omega}_D)=\mathbf{E}'_N vecd(\mathbf{\Omega}_D)$ (see again Magnus (1988)). Using this expression we can trivially prove that

$$\frac{\partial vec(\mathbf{\Omega})}{\partial vecl'(\mathbf{\Omega}_L)} = (\mathbf{I}_N+\mathbf{K}_{NN})(\mathbf{\Omega}_L\mathbf{\Omega}_D\otimes\mathbf{I}_N)\mathbf{S}'_N, \tag{A20}$$

$$\frac{\partial vec(\mathbf{\Omega})}{\partial vecd'(\mathbf{\Omega}_D)} = (\mathbf{\Omega}_L\otimes\mathbf{\Omega}_L)\mathbf{E}'_N. \tag{A21}$$

Then, we can obtain the Jacobian of $vec(\mathbf{\Omega})$ with respect to ϑ_i and $\boldsymbol{\varpi}_D$ from the expressions for the Jacobian of $vecd(\mathbf{\Omega}^{\circ}_D)$ in the diagonal case.

Finally, let us study the fairly common situation in which $\mathbf{\Omega}$ is estimated subject to the exact single factor structure $\mathbf{cc}'+\mathbf{\Upsilon}$, where $\mathbf{c}$ is an $N\times 1$ vector and $\mathbf{\Upsilon}$ a diagonal matrix. Assuming that $\mathbf{\Upsilon}$ is positive definite, we can always parametrise $\mathbf{\Omega}$ as

$$\mathbf{\Upsilon}^{1/2}(\mathbf{c}^*\mathbf{c}^{*\prime}+\mathbf{I}_N)\mathbf{\Upsilon}^{1/2}, \tag{A22}$$

where $\mathbf{c}^*=\mathbf{\Upsilon}^{-1/2}\mathbf{c}$. Given that the eigenvalues of $\mathbf{c}^*\mathbf{c}^{*\prime}+\mathbf{I}_N$ are $1+\mathbf{c}^{*\prime}\mathbf{c}^*$ (once) and 1 ($N-1$ times), then $|\mathbf{\Omega}|=|\mathbf{\Upsilon}|\cdot(1+\mathbf{c}^{*\prime}\mathbf{c}^*)$. As a result, if we write $\upsilon_{jj}=\vartheta_i(1+\mathbf{c}^{*\prime}\mathbf{c}^*)^{-1/N}\exp(\varpi_{Dj})$ for $j=1,\ldots,N-1$ and $\upsilon_{NN}=\vartheta_i(1+\mathbf{c}^{*\prime}\mathbf{c}^*)^{-1/N}\exp\left(-\sum_{j=1}^{N-1}\varpi_{Dj}\right)$, we will ensure that $|\mathbf{\Omega}|=\vartheta_i^N$ as required.

As for the Jacobian matrices, it follows from (A22) that

$$d\mathbf{\Omega}=d\mathbf{\Upsilon}^{1/2}(\mathbf{c}^*\mathbf{c}^{*\prime}+\mathbf{I}_N)\mathbf{\Upsilon}^{1/2}+\mathbf{\Upsilon}^{1/2}\cdot d\mathbf{c}^*\cdot\mathbf{c}^{*\prime}\mathbf{\Upsilon}^{1/2}+\mathbf{\Upsilon}^{1/2}\mathbf{c}^*\cdot d\mathbf{c}^{*\prime}\cdot\mathbf{\Upsilon}^{1/2}+\mathbf{\Upsilon}^{1/2}(\mathbf{c}^*\mathbf{c}^{*\prime}+\mathbf{I}_N)\cdot d\mathbf{\Upsilon}^{1/2},$$

whence

$$\begin{aligned}
dvec(\mathbf{\Omega}) &= [\mathbf{\Upsilon}^{1/2}(\mathbf{c}^*\mathbf{c}^{*\prime}+\mathbf{I}_N)\otimes\mathbf{I}_N]dvec(\mathbf{\Upsilon}^{1/2})+(\mathbf{\Upsilon}^{1/2}\mathbf{c}^*\otimes\mathbf{\Upsilon}^{1/2})d\mathbf{c}^* \\
&+(\mathbf{\Upsilon}^{1/2}\otimes\mathbf{\Upsilon}^{1/2}\mathbf{c}^*)d\mathbf{c}^*+[\mathbf{I}_N\otimes\mathbf{\Upsilon}^{1/2}(\mathbf{c}^*\mathbf{c}^{*\prime}+\mathbf{I}_N)]dvec(\mathbf{\Upsilon}^{1/2}) \\
&= [(\mathbf{\Upsilon}^{1/2}\mathbf{c}^*\otimes\mathbf{\Upsilon}^{1/2})+(\mathbf{\Upsilon}^{1/2}\otimes\mathbf{\Upsilon}^{1/2}\mathbf{c}^*)]d\mathbf{c}^* \\
&+\frac{1}{2}\{[\mathbf{\Upsilon}^{1/2}(\mathbf{c}^*\mathbf{c}^{*\prime}+\mathbf{I}_N)\otimes\mathbf{I}_N]+[\mathbf{I}_N\otimes\mathbf{\Upsilon}^{1/2}(\mathbf{c}^*\mathbf{c}^{*\prime}+\mathbf{I}_N)]\}\mathbf{E}'_N\mathbf{\Upsilon}^{-1/2}dvecd(\mathbf{\Upsilon}),
\end{aligned}$$

where we have exploited the fact that

$$dvec(\mathbf{\Upsilon}^{1/2})=\mathbf{E}'_N dvecd(\mathbf{\Upsilon}^{1/2})=\mathbf{E}'_N\mathbf{\Upsilon}^{-1/2}dvecd(\mathbf{\Upsilon}).$$

The derivatives of $\boldsymbol{\gamma}_{jj}$ with respect to to ϑ_i and ϖ_{Dk} are simply $(1+\mathbf{c}^{*\prime}\mathbf{c}^{*})^{-1/N}$ times the corresponding derivatives in the diagonal case we discussed above. Thus, the only remaining derivatives will be

$$\frac{\partial \upsilon_{jj}}{\partial c_k^*} = \frac{2\vartheta_i \upsilon_{jj}}{N(1+\mathbf{c}^{*\prime}\mathbf{c}^{*})^{(N+1)/N}} c_k^*.$$

B Proofs

Proposition 1

We can directly work in terms of the $\boldsymbol{\vartheta}$ parameters thanks to our assumptions on the mapping $\mathbf{r}_s(.)$. Let us initially keep $\boldsymbol{\eta}$ fixed to some admissible value so as to focus on the spherically symmetric score vector for $\boldsymbol{\vartheta}$ in (A4). Given that 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

$$\begin{aligned}\mathbf{Z}_{dt}(\boldsymbol{\vartheta}) = \begin{bmatrix} \mathbf{Z}_{\boldsymbol{\vartheta}_c lt}(\boldsymbol{\vartheta}) & \mathbf{Z}_{\boldsymbol{\vartheta}_c st}(\boldsymbol{\vartheta}) \\ 0 & \mathbf{Z}_{\vartheta_i st}(\boldsymbol{\vartheta}) \end{bmatrix} &= \begin{cases} \vartheta_i^{-1/2}\left[\partial \boldsymbol{\mu}_t'(\boldsymbol{\vartheta}_c)/\partial \boldsymbol{\vartheta}_c\right] \boldsymbol{\Sigma}_t^{\circ -1/2\prime}(\boldsymbol{\vartheta}_c) \\ 0 \end{cases} \\ &= \begin{cases} \frac{1}{2}\{\partial vec'[\Sigma_t^{\circ}(\boldsymbol{\vartheta}_c)]/\partial \boldsymbol{\vartheta}_c\}[\boldsymbol{\Sigma}_t^{\circ -1/2\prime}(\boldsymbol{\vartheta}_c) \otimes \boldsymbol{\Sigma}_t^{\circ -1/2\prime}(\boldsymbol{\vartheta}_c)] \\ \frac{1}{2}\vartheta_i^{-1} vec'(\mathbf{I}_N) \end{cases}. \end{aligned} \tag{B23}$$

Thus, the conditional mean and variance parameter scores will be

$$\begin{aligned}\mathbf{s}_{\boldsymbol{\vartheta}_c t}(\boldsymbol{\vartheta},\boldsymbol{\eta}) &= \vartheta_i^{-1/2}\frac{\partial \boldsymbol{\mu}_t'(\boldsymbol{\vartheta}_c)}{\partial \boldsymbol{\vartheta}_c}\boldsymbol{\Sigma}_t^{\circ -1/2\prime}(\boldsymbol{\vartheta}_c)\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]\boldsymbol{\varepsilon}_t^*(\boldsymbol{\vartheta}) \\ &+\frac{1}{2}\frac{\partial vec'[\Sigma_t^{\circ}(\boldsymbol{\vartheta}_c)]}{\partial \boldsymbol{\vartheta}_c}[\boldsymbol{\Sigma}_t^{\circ -1/2\prime}(\boldsymbol{\vartheta}_c) \otimes \boldsymbol{\Sigma}_t^{\circ -1/2\prime}(\boldsymbol{\vartheta}_c)] vec\left\{\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]\cdot \boldsymbol{\varepsilon}_t^*(\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\vartheta}) - \mathbf{I}_N\right\} \end{aligned} \tag{B24}$$

and

$$s_{\vartheta_i t}(\boldsymbol{\vartheta},\boldsymbol{\eta}) = \frac{1}{2\vartheta_i}\left\{\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]\varsigma_t(\boldsymbol{\vartheta}) - N\right\}. \tag{B25}$$

But since

$$\boldsymbol{\varepsilon}_t^*(\boldsymbol{\vartheta}_{c0},\vartheta_i) = \sqrt{1/\vartheta_i}\boldsymbol{\Sigma}_t^{\circ -1/2}(\boldsymbol{\vartheta}_{c0})[\mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\vartheta}_{c0})] = \sqrt{\vartheta_{i0}/\vartheta_i}\boldsymbol{\varepsilon}_t^* = \lambda^{-1/2}\boldsymbol{\varepsilon}_t^*,$$

so that

$$\varsigma_t(\boldsymbol{\vartheta}_{c0},\vartheta_i) = (\vartheta_{i0}/\vartheta_i)\varsigma_t = \lambda^{-1}\varsigma_t,$$

we will have that

$$\mathbf{e}_{lt}(\boldsymbol{\vartheta}_{c0},\vartheta_i,\boldsymbol{\eta}) = \delta(\lambda^{-1}\varsigma_t,\boldsymbol{\eta})\lambda^{-1/2}\boldsymbol{\varepsilon}_t^* = \delta(\lambda^{-1}\varsigma_t,\boldsymbol{\eta})\lambda^{-1/2}\sqrt{\varsigma_t}\mathbf{u}_t, \tag{B26}$$

$$\mathbf{e}_{st}(\boldsymbol{\vartheta}_{c0},\vartheta_i,\boldsymbol{\eta}) = vec\left[\delta(\lambda^{-1}\varsigma_t,\boldsymbol{\eta})\lambda^{-1}\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime} - \mathbf{I}_N\right] = vec\left[\delta(\lambda^{-1}\varsigma_t,\boldsymbol{\eta})\lambda^{-1}\varsigma_t\mathbf{u}_t\mathbf{u}_t' - \mathbf{I}_N\right]. \tag{B27}$$

Then, it follows that $E[\mathbf{e}_{lt}(\boldsymbol{\vartheta}_{c0},\vartheta_i,\boldsymbol{\eta})|I_{t-1};\boldsymbol{\varphi}_0] = \mathbf{0}$ regardless of ϑ_i and $\boldsymbol{\eta}$ because of the serial and mutual independence of ς_t and $\mathbf{u}_t$, and the fact that $E(\mathbf{u}_t) = \mathbf{0}$. Similarly,

$$E[\mathbf{e}_{st}(\boldsymbol{\vartheta}_{c0},\vartheta_i,\boldsymbol{\eta})|I_{t-1};\boldsymbol{\varphi}_0] = E[\delta(\lambda^{-1}\varsigma_t,\eta)\lambda^{-1}(\varsigma_t/N) - 1\big|\,\boldsymbol{\varphi}_0]\cdot vec(\mathbf{I}_N)$$

because of the serial and mutual independence of ς_t and $\mathbf{u}_t$, and the fact that $E(\mathbf{u}_t\mathbf{u}_t') = N^{-1}\mathbf{I}_N$.

If we define $\vartheta_{i\infty}(\boldsymbol{\eta}) > 0$ as the value of ϑ_i that satisfies the moment condition

$$E\left[\delta[\lambda_\infty^{-1}(\boldsymbol{\eta})\varsigma_t, \boldsymbol{\eta}]\lambda_\infty^{-1}(\boldsymbol{\eta})(\varsigma_t/N) - 1 \middle| \boldsymbol{\varphi}_0\right] = 0, \tag{B28}$$

with

$$\lambda_\infty(\boldsymbol{\eta}) = \vartheta_{i\infty}(\boldsymbol{\eta})/\vartheta_{i0}, \tag{B29}$$

then it is straightforward to show that

$$E\{\mathbf{s}_{\boldsymbol{\vartheta}t}[\boldsymbol{\vartheta}_{c0}, \vartheta_{i\infty}(\boldsymbol{\eta}), \boldsymbol{\eta}]|I_{t-1}; \boldsymbol{\varphi}_0\} = \mathbf{0}, \tag{B30}$$

which confirms that $\boldsymbol{\vartheta}_{c0}$ and $\vartheta_{i\infty}(\boldsymbol{\eta})$ will be the pseudo-true values of the parameters corresponding to a restricted PML estimator that keeps $\boldsymbol{\eta}$ fixed. Thus, we can understand $\lambda_\infty(\boldsymbol{\eta})$ in (B29) as the "relative asymptotic bias" in estimating ϑ_i.

If we define $\boldsymbol{\eta}_\infty$ as the value of $\boldsymbol{\eta}$ that satisfies the moment condition

$$E\{\mathbf{s}_{\boldsymbol{\eta}t}[\boldsymbol{\vartheta}_{c0}, \vartheta_{i\infty}(\boldsymbol{\eta}_\infty), \boldsymbol{\eta}_\infty]|\boldsymbol{\varphi}_0\} = \mathbf{0}, \tag{B31}$$

which we assume lies in the interior of the admissible parameter space, then it is clear that $\boldsymbol{\vartheta}_{c0}, \vartheta_{i\infty} = \vartheta_{i\infty}(\boldsymbol{\eta}_\infty)$ and $\boldsymbol{\eta}_\infty$ will be the pseudo-true values of the parameters corresponding to the unrestricted PMLE that jointly estimates $\boldsymbol{\eta}$, and $\lambda_\infty = \vartheta_{i\infty}/\vartheta_{i0}$ the corresponding "relative asymptotic bias".

Proposition 2

To obtain the asymptotic distribution of the unrestricted pseudo ML estimators $\hat{\boldsymbol{\vartheta}}_T$ and $\hat{\boldsymbol{\eta}}_T$, we need the asymptotic covariance matrix of the average scores as well as the expected value of the average Hessian matrix evaluated at the pseudo true values $\boldsymbol{\phi}_\infty' = (\boldsymbol{\vartheta}_{c0}', \vartheta_{i\infty}, \boldsymbol{\eta}_\infty')$. Given that $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}_\infty)$ only depends on $\varsigma_t(\boldsymbol{\vartheta}_{c0}, \vartheta_{i\infty})$, which is *i.i.d.* over time, it follows that

$$E[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] = \mathbf{0}, \tag{B32}$$

which in conjunction with (B30) proves the martingale difference nature of the misspecified spherical score evaluated at the pseudo-true values. As a result, we only need the contemporaneous covariance matrix of the component of the score corresponding to the t^{th} observation, which in turn depends on the contemporaneous covariance matrix of $\mathbf{e}_{dt}(\boldsymbol{\phi}_\infty)$ and $\mathbf{e}_{rt}(\boldsymbol{\phi}_\infty)$.

If we re-write $\mathbf{e}_{dt}(\boldsymbol{\phi}_\infty)$ as in (B26) and (B27), it immediately follows that

$$\begin{aligned}
E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)\mathbf{e}_{lt}'(\boldsymbol{\phi}_\infty)] &= E\left\{\delta^2(\lambda_\infty^{-1}\varsigma_t, \boldsymbol{\eta}_\infty)\lambda_\infty^{-1}\varsigma_t\mathbf{u}_t\mathbf{u}_t'\right\} \\
&= E[\delta^2(\lambda_\infty^{-1}\varsigma_t, \boldsymbol{\eta}_\infty)\lambda_\infty^{-1}(\varsigma_t/N)]\mathbf{I}_N = \mathrm{M}_{ll}^{O}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\mathbf{I}_N, \\
E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)\mathbf{e}_{st}'(\boldsymbol{\phi}_\infty)] &= E\left\{\delta(\lambda_\infty^{-1}\varsigma_t, \boldsymbol{\eta}_\infty)\lambda_\infty^{-1/2}\sqrt{\varsigma_t}\mathbf{u}_t \cdot vec'\left[\delta(\lambda_\infty^{-1}\varsigma_t, \boldsymbol{\eta}_\infty)\lambda_\infty^{-1}\varsigma_t\mathbf{u}_t\mathbf{u}_t' - \mathbf{I}_N\right]\right\} = \mathbf{0}
\end{aligned}$$

by virtue of (A1), and

$$
\begin{aligned}
E[\mathbf{e}_{st}(\boldsymbol{\phi}_0)\mathbf{e}'_{st}(\boldsymbol{\phi}_0)] &= E\left\{vec\left[\delta(\lambda_\infty^{-1}\varsigma_t,\boldsymbol{\eta}_\infty)(\lambda_\infty^{-1}\varsigma_t/N)N\mathbf{u}_t\mathbf{u}'_t-\mathbf{I}_N\right]\right. \\
&\left.\times vec'\left[\delta(\lambda_\infty^{-1}\varsigma_t,\boldsymbol{\eta}_\infty)(\lambda_\infty^{-1}\varsigma_t/N)N\mathbf{u}_t\mathbf{u}'_t-\mathbf{I}_N\right]\right\} \\
&= E\left[\delta(\lambda_\infty^{-1}\varsigma_t,\boldsymbol{\eta}_\infty)(\lambda_\infty^{-1}\varsigma_t/N)\right]^2\frac{N}{N+2}[(\mathbf{I}_{N^2}+\mathbf{K}_{NN})+vec\,(\mathbf{I}_N)\,vec'\,(\mathbf{I}_N)] \\
&-2E\left[\delta(\lambda_\infty^{-1}\varsigma_t,\boldsymbol{\eta}_\infty)(\lambda_\infty^{-1}\varsigma_t/N)\right]vec\,(\mathbf{I}_N)\,vec'\,(\mathbf{I}_N)+vec\,(\mathbf{I}_N)\,vec'\,(\mathbf{I}_N) \\
&= \frac{N}{(N+2)}E\left[\delta(\lambda_\infty^{-1}\varsigma_t,\boldsymbol{\eta}_\infty)(\lambda_\infty^{-1}\varsigma_t/N)\right]^2(\mathbf{I}_{N^2}+\mathbf{K}_{NN}) \\
&+\left\{\frac{N}{(N+2)}E\left[\delta(\lambda_\infty^{-1}\varsigma_t,\boldsymbol{\eta}_\infty)(\lambda_\infty^{-1}\varsigma_t/N)\right]^2-1\right\}vec\,(\mathbf{I}_N)\,vec'\,(\mathbf{I}_N)] \\
&= \mathrm{M}_{ss}^{O}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\,(\mathbf{I}_{N^2}+\mathbf{K}_{NN})+[\mathrm{M}_{ss}^{O}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)-1]vec(\mathbf{I}_N)vec'(\mathbf{I}_N)
\end{aligned}
$$

by virtue of (A2) and (B28).

Moreover, it is clear from (A3) that $\mathbf{e}_{rt}(\boldsymbol{\phi}_\infty)$ will be a function of ς_t but not of $\mathbf{u}_t$, which immediately implies that $E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{rt}(\boldsymbol{\phi}_\infty)]=\mathbf{0}$ and

$$
\begin{aligned}
E[\mathbf{e}_{st}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{rt}(\boldsymbol{\phi}_\infty)] &= E\left\{vec\left[\delta(\lambda_\infty^{-1}\varsigma_t,\boldsymbol{\eta}_\infty)\lambda_\infty^{-1}\varsigma_t\cdot\mathbf{u}_t\mathbf{u}'_t-\mathbf{I}_N\right]\mathbf{e}'_{rt}(\boldsymbol{\phi}_\infty)\right\} \\
&= vec(\mathbf{I}_N)E\left\{\left[\delta(\lambda_\infty^{-1}\varsigma_t,\boldsymbol{\eta}_\infty)(\lambda_\infty^{-1}\varsigma_t/N)-1\right]\mathbf{e}'_{rt}(\boldsymbol{\phi}_\infty)\right\}=vec(\mathbf{I}_N)\mathrm{M}_{sr}^{O}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0).
\end{aligned}
$$

If we combine these expressions with (B23) and apply the law of iterated expectations, after some algebraic manipulations we obtain

$$
\begin{aligned}
\mathcal{B}_{\boldsymbol{\vartheta}_c\boldsymbol{\vartheta}_c}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\mathbf{s}_{\boldsymbol{\vartheta}_c t}(\boldsymbol{\phi}_\infty)\mathbf{s}'_{\boldsymbol{\vartheta}_c t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0]=\frac{\mathrm{M}_{ll}^{O}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)}{\vartheta_{i\infty}}E\left[\frac{\partial\boldsymbol{\mu}'_t(\boldsymbol{\vartheta}_{c0})}{\partial\boldsymbol{\vartheta}_c}\boldsymbol{\Sigma}_t^{\circ-1}(\boldsymbol{\vartheta}_{c0})\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\vartheta}_{c0})}{\partial\boldsymbol{\vartheta}'_c}\middle|\boldsymbol{\varphi}_0\right] \\
&+\frac{\mathrm{M}_{ss}^{O}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)}{2}E\left[\frac{\partial vec'[\Sigma_t^{\circ}(\boldsymbol{\vartheta}_{c0})]}{\partial\boldsymbol{\vartheta}_c}[\boldsymbol{\Sigma}_t^{\circ-1}(\boldsymbol{\vartheta}_{c0})\otimes\boldsymbol{\Sigma}_t^{\circ-1}(\boldsymbol{\vartheta}_{c0})]\frac{\partial vec[\Sigma_t^{\circ}(\boldsymbol{\vartheta}_{c0})]}{\partial\boldsymbol{\vartheta}'_c}\middle|\boldsymbol{\varphi}_0\right] \\
&+\frac{\mathrm{M}_{ss}^{O}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)-1}{4}E\left[\frac{\partial vec'[\Sigma_t^{\circ}(\boldsymbol{\vartheta}_{c0})]}{\partial\boldsymbol{\vartheta}_c}vec[\boldsymbol{\Sigma}_t^{\circ-1}(\boldsymbol{\vartheta}_{c0})]vec'[\boldsymbol{\Sigma}_t^{\circ-1}(\boldsymbol{\vartheta}_{c0})]\frac{\partial vec[\Sigma_t^{\circ}(\boldsymbol{\vartheta}_{c0})]}{\partial\boldsymbol{\vartheta}'_c}\middle|\boldsymbol{\varphi}_0\right], \quad \text{(B33)}
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{B}_{\boldsymbol{\vartheta}_c\vartheta_i}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\mathbf{s}_{\boldsymbol{\vartheta}_c t}(\boldsymbol{\phi}_\infty)s_{\vartheta_i t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0]=\frac{\mathrm{M}_{ss}^{O}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)(N+2)-N}{2\vartheta_{i\infty}}\mathbf{W}_{\boldsymbol{\vartheta}_c}(\boldsymbol{\vartheta}_{c0};\boldsymbol{\varphi}_0), \\
\mathcal{B}_{\vartheta_i\vartheta_i}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[s^2_{\vartheta_i t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0]=\frac{N[(N+2)\mathrm{M}_{ss}^{O}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)-N]}{4\vartheta^2_{i\infty}},
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{B}_{\boldsymbol{\vartheta}_c\boldsymbol{\eta}}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\mathbf{s}_{\boldsymbol{\vartheta}_c t}(\boldsymbol{\phi}_\infty)\mathbf{s}'_{\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0]=\mathbf{W}_{\boldsymbol{\vartheta}_c}(\boldsymbol{\vartheta}_{c0};\boldsymbol{\varphi}_0)\mathrm{M}_{sr}^{O}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0), \\
\mathcal{B}_{\vartheta_i\boldsymbol{\eta}}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[s_{\vartheta_i t}(\boldsymbol{\phi}_\infty)\mathbf{s}'_{\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0]=\frac{N}{2\vartheta_{i\infty}}\mathrm{M}_{sr}^{O}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{W}_{\boldsymbol{\vartheta}}(\boldsymbol{\vartheta}_0;\boldsymbol{\varphi}_0) &= \mathbf{Z}_d(\boldsymbol{\vartheta}_0)[\mathbf{0}',vec'(\mathbf{I}_N)]'=E[\mathbf{Z}_{dt}(\boldsymbol{\vartheta}_0)|\boldsymbol{\varphi}_0][\mathbf{0}',vec'(\mathbf{I}_N)]' \\
&= E\left\{\frac{1}{2}\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)\right]}{\partial\boldsymbol{\vartheta}}vec[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}_0)]\middle|\boldsymbol{\varphi}_0\right\}=E[\mathbf{W}_{\boldsymbol{\vartheta}t}(\boldsymbol{\vartheta}_0)|\boldsymbol{\varphi}_0]=-E\left[\frac{\partial d_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\middle|\boldsymbol{\varphi}_0\right], \quad \text{(B34)}
\end{aligned}
$$

$$\mathrm{M}_{ll}^{O}(\boldsymbol{\phi};\boldsymbol{\varphi}) = E\left\{\delta^2[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]\cdot[\varsigma_t(\boldsymbol{\vartheta})/N]\big|\,\boldsymbol{\varphi}\right\}$$
$$\mathrm{M}_{ss}^{O}(\boldsymbol{\phi};\boldsymbol{\varphi}) = N(N+2)^{-1}\left[1+V\left\{\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]\cdot[\varsigma_t(\boldsymbol{\vartheta})/N]|\,\boldsymbol{\varphi}\right\}\right],$$
$$\mathrm{M}_{sr}^{O}(\boldsymbol{\phi};\boldsymbol{\varphi}) = E\left[\left\{\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]\cdot[\varsigma_t(\boldsymbol{\vartheta})/N]-1\right\}\mathbf{e}_{rt}'(\boldsymbol{\phi})\big|\,\boldsymbol{\varphi}\right],$$

and

$$\mathcal{B}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) = E[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}_\infty)\mathbf{s}_{\boldsymbol{\eta}t}'(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] = \mathrm{M}_{rr}^{O}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0).$$

To obtain the expected value of the Hessian, it is convenient to write $\mathbf{h}_{\boldsymbol{\vartheta}\boldsymbol{\vartheta}t}(\boldsymbol{\phi}_\infty)$ in (A9) as

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

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

$$\begin{aligned}
&-\delta(\lambda_\infty^{-1}\varsigma_t,\boldsymbol{\eta}_\infty)\mathbf{Z}_{lt}(\boldsymbol{\vartheta}_\infty)\mathbf{Z}_{lt}'(\boldsymbol{\vartheta}_\infty)-2\partial\delta(\lambda_\infty^{-1}\varsigma_t,\boldsymbol{\eta}_\infty)/\partial\varsigma\cdot\mathbf{Z}_{lt}(\boldsymbol{\vartheta}_\infty)\lambda_\infty^{-1}\varsigma_t\mathbf{u}_t\mathbf{u}_t'\mathbf{Z}_{lt}'(\boldsymbol{\vartheta}_\infty)\\
&-2\mathbf{Z}_{st}(\boldsymbol{\vartheta}_\infty)\mathbf{Z}_{st}'(\boldsymbol{\vartheta}_\infty)-2\partial\delta(\lambda_\infty^{-1}\varsigma_t,\boldsymbol{\eta}_\infty)/\partial\varsigma\cdot(\lambda_\infty^{-1}\varsigma_t)^2\mathbf{Z}_{st}(\boldsymbol{\vartheta}_\infty)vec(\mathbf{u}_t\mathbf{u}_t')vec'(\mathbf{u}_t\mathbf{u}_t')\mathbf{Z}_{st}'(\boldsymbol{\vartheta}_\infty),
\end{aligned}$$

whose conditional expectation will be

$$\begin{aligned}
&-E[\delta(\lambda_\infty^{-1}\varsigma_t,\boldsymbol{\eta}_\infty)+2(\lambda_\infty^{-1}\varsigma_t/N)\cdot\partial\delta(\lambda_\infty^{-1}\varsigma_t,\boldsymbol{\eta}_\infty)/\partial\varsigma]\cdot\mathbf{Z}_{lt}(\boldsymbol{\vartheta}_\infty)\mathbf{Z}_{lt}'(\boldsymbol{\vartheta}_\infty)-2\mathbf{Z}_{st}(\boldsymbol{\vartheta}_\infty)\mathbf{Z}_{st}'(\boldsymbol{\vartheta}_\infty)\\
&-\frac{2NE[(\lambda_\infty^{-1}\varsigma_t/N)^2\cdot\partial\delta(\lambda_\infty^{-1}\varsigma_t,\boldsymbol{\eta}_\infty)/\partial\varsigma]}{(N+2)}\mathbf{Z}_{st}(\boldsymbol{\vartheta}_\infty)[(\mathbf{I}_{N^2}\otimes\mathbf{K}_{NN})+vec(\mathbf{I}_N)vec'(\mathbf{I}_N)]\mathbf{Z}_{st}'(\boldsymbol{\vartheta}_\infty)\\
=\;&-\mathrm{M}_{ll}^{H}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}_{lt}(\boldsymbol{\vartheta}_\infty)\mathbf{Z}_{lt}'(\boldsymbol{\vartheta}_\infty)\\
&-\mathbf{Z}_{st}(\boldsymbol{\vartheta}_\infty)\left\{\mathrm{M}_{ss}^{H}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)(\mathbf{I}_{N^2}+\mathbf{K}_{NN})+[\mathrm{M}_{ss}^{H}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)-1]vec(\mathbf{I}_N)vec'(\mathbf{I}_N)\right\}\mathbf{Z}_{st}'(\boldsymbol{\vartheta}_\infty),
\end{aligned}$$

where

$$\mathrm{M}_{ll}^{H}(\boldsymbol{\phi};\boldsymbol{\varphi}) = E\left\{2\partial\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]/\partial\varsigma\cdot[\varsigma_t(\boldsymbol{\vartheta})/N]+\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]|\,\boldsymbol{\varphi}\right\},$$
$$\mathrm{M}_{ss}^{H}(\boldsymbol{\phi};\boldsymbol{\varphi}) = N(N+2)^{-1}E\left\{2\partial\delta[\varsigma_t(\boldsymbol{\vartheta}),\boldsymbol{\eta}]/\partial\varsigma\cdot\varsigma_t^2(\boldsymbol{\vartheta})/[N(N+2)]\big|\,\boldsymbol{\varphi}\right\}+1.$$

As for $\mathbf{h}_{\boldsymbol{\vartheta}\boldsymbol{\eta}t}(\boldsymbol{\phi}_\infty)$, it follows from (A10) and (A5) that we can write it as

$$\begin{aligned}
&\{\mathbf{Z}_{lt}(\boldsymbol{\vartheta}_\infty)\boldsymbol{\varepsilon}_t^*(\boldsymbol{\vartheta}_\infty) + \mathbf{Z}_{st}(\boldsymbol{\vartheta}_\infty)vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\vartheta}_\infty)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\vartheta}_\infty)\right]\} \cdot \partial\delta(\lambda_\infty^{-1}\varsigma_t, \boldsymbol{\eta}_\infty)/\partial\boldsymbol{\eta}' \\
= \;& [\mathbf{Z}_{lt}(\boldsymbol{\vartheta}_\infty)\mathbf{u}_t\lambda_\infty^{-1/2}\sqrt{\varsigma_t} + \mathbf{Z}_{st}(\boldsymbol{\vartheta}_\infty)vec(\mathbf{u}_t\mathbf{u}_t')\lambda_\infty^{-1}\varsigma_t \cdot \partial\delta(\lambda_\infty^{-1}\varsigma_t, \boldsymbol{\eta}_\infty)/\partial\boldsymbol{\eta}',
\end{aligned}$$

whose conditional expected value will be

$$\mathbf{Z}_{st}(\boldsymbol{\vartheta}_\infty)vec(\mathbf{I}_N)E[(\lambda_\infty^{-1}\varsigma_t/N) \cdot \partial\delta(\lambda_\infty^{-1}\varsigma_t, \boldsymbol{\eta}_\infty)/\partial\boldsymbol{\eta}'] = -\mathbf{Z}_{st}(\boldsymbol{\vartheta}_\infty)vec(\mathbf{I}_N)\mathrm{M}_{sr}^H(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0),$$

with

$$\mathrm{M}_{sr}^H(\boldsymbol{\phi}; \boldsymbol{\varphi}) = -E\left\{[\varsigma_t(\boldsymbol{\vartheta})/N] \cdot \partial\delta[\varsigma_t(\boldsymbol{\vartheta}), \boldsymbol{\eta}]/\partial\boldsymbol{\eta}|\boldsymbol{\varphi}\right\}.$$

Replacing once again $\mathbf{Z}_{lt}(\boldsymbol{\vartheta}_\infty)$ and $\mathbf{Z}_{st}(\boldsymbol{\vartheta}_\infty)$ by the relevant expressions in (B23) and applying the law of iterated expectations, we obtain

$$\begin{aligned}
\mathcal{A}_{\boldsymbol{\vartheta}_c\boldsymbol{\vartheta}_c}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = -E[\mathbf{h}_{\boldsymbol{\vartheta}_c\boldsymbol{\vartheta}_c t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \frac{\mathrm{M}_{ll}^H(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)}{\vartheta_{i\infty}} E\left[\frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\vartheta}_{c0})}{\partial\boldsymbol{\vartheta}_c}\boldsymbol{\Sigma}_t^{\circ -1}(\boldsymbol{\vartheta}_c)\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\vartheta}_{c0})}{\partial\boldsymbol{\vartheta}_c'}\middle|\boldsymbol{\varphi}_0\right] \\
&+\frac{\mathrm{M}_{ss}^H(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)}{2} E\left[\frac{\partial vec'[\Sigma_t^\circ(\boldsymbol{\vartheta}_{c0})]}{\partial\boldsymbol{\vartheta}_c}[\boldsymbol{\Sigma}_t^{\circ -1}(\boldsymbol{\vartheta}_{c0}) \otimes \boldsymbol{\Sigma}_t^{\circ -1}(\boldsymbol{\vartheta}_{c0})]\frac{\partial vec[\Sigma_t^\circ(\boldsymbol{\vartheta}_{c0})]}{\partial\boldsymbol{\vartheta}_c'}\middle|\boldsymbol{\varphi}_0\right] \\
&+\frac{\mathrm{M}_{ss}^H(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)-1}{4} E\left[\frac{\partial vec'[\Sigma_t^\circ(\boldsymbol{\vartheta}_{c0})]}{\partial\boldsymbol{\vartheta}_c}vec[\boldsymbol{\Sigma}_t^{\circ -1}(\boldsymbol{\vartheta}_{c0})]vec'[\boldsymbol{\Sigma}_t^{\circ -1}(\boldsymbol{\vartheta}_{c0})]\frac{\partial vec[\Sigma_t^\circ(\boldsymbol{\vartheta}_{c0})]}{\partial\boldsymbol{\vartheta}_c'}\middle|\boldsymbol{\varphi}_0\right], \quad \text{(B35)}
\end{aligned}$$

$$\begin{aligned}
\mathcal{A}_{\boldsymbol{\vartheta}_c\vartheta_i}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) &= -E[\mathbf{h}_{\boldsymbol{\vartheta}_c\vartheta_i t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] = \frac{\mathrm{M}_{ss}^H(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)(N+2) - N}{2\vartheta_{i\infty}}\mathbf{W}_{\boldsymbol{\vartheta}_c}(\boldsymbol{\vartheta}_{c0}; \boldsymbol{\varphi}_0), \\
\mathcal{A}_{\vartheta_i\vartheta_i}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) &= -E[h_{\vartheta_i\vartheta_i t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] = \frac{N[(N+2)\mathrm{M}_{ss}^H(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) - N]}{4\vartheta_{i\infty}^2},
\end{aligned}$$

$$\begin{aligned}
\mathcal{A}_{\boldsymbol{\vartheta}_c\boldsymbol{\eta}}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) &= -E[\mathbf{h}_{\boldsymbol{\vartheta}_c\boldsymbol{\eta}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] = \mathbf{W}_{\boldsymbol{\vartheta}_c}(\boldsymbol{\vartheta}_{c0}; \boldsymbol{\varphi}_0)\mathrm{M}_{sr}^H(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0), \\
\mathcal{A}_{\vartheta_i\boldsymbol{\eta}}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) &= -E[h_{\vartheta_i\boldsymbol{\eta}t}(\boldsymbol{\phi}_\infty)\mathbf{s}_{\boldsymbol{\eta}t}'(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] = \frac{N}{2\vartheta_{i\infty}}\mathrm{M}_{sr}^H(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0),
\end{aligned}$$

and

$$\mathcal{A}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = -E[\mathbf{h}_{\boldsymbol{\eta}\boldsymbol{\eta}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] = \mathrm{M}_{rr}^H(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0).$$

Let us now turn to our consistent estimator of ϑ_i in (9). The fact that the Gaussian pseudo score for this parameter is an influence function that only depends on $\boldsymbol{\vartheta}_c$ and $\bar{\vartheta}_i$ trivially implies that

$$\frac{\partial s_{\vartheta_i t}(\boldsymbol{\vartheta}_c, \bar{\vartheta}_i; \mathbf{0})}{\partial\vartheta_i} = \mathbf{0} \text{ and } \frac{\partial s_{\vartheta_i t}(\boldsymbol{\vartheta}_c, \bar{\vartheta}_i; \mathbf{0})}{\partial\boldsymbol{\eta}} = \mathbf{0}.$$

For analogous reasons,

$$\frac{\partial\mathbf{s}_{\boldsymbol{\vartheta}_c t}(\boldsymbol{\vartheta}_c, \vartheta_i; \boldsymbol{\eta})}{\partial\bar{\vartheta}_i} = \mathbf{0}, \;\; \frac{\partial s_{\vartheta_i t}(\boldsymbol{\vartheta}_c, \vartheta_i; \boldsymbol{\eta})}{\partial\bar{\vartheta}_i} = 0 \text{ and } \frac{\partial\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\vartheta}_c, \vartheta_i; \boldsymbol{\eta})}{\partial\bar{\vartheta}_i} = \mathbf{0}.$$

We will also have that

$$\frac{\partial s_{\vartheta_i t}(\boldsymbol{\vartheta}_c, \bar{\vartheta}_i; \mathbf{0})}{\partial \boldsymbol{\vartheta}_c} = \mathbf{h}'_{\boldsymbol{\vartheta}_c \vartheta_i t}(\boldsymbol{\vartheta}, \mathbf{0}) = -\frac{1}{\vartheta_i^{3/2}} \frac{\partial \boldsymbol{\mu}'_t(\boldsymbol{\vartheta}_c)}{\partial \boldsymbol{\vartheta}_c} \boldsymbol{\Sigma}_t^{\circ -1/2\prime}(\boldsymbol{\vartheta}_c) \boldsymbol{\varepsilon}_t^*(\boldsymbol{\vartheta})$$
$$-\frac{1}{2\vartheta_i} \frac{\partial vec'[\Sigma_t^\circ(\boldsymbol{\vartheta}_c)]}{\partial \boldsymbol{\vartheta}_c} [\boldsymbol{\Sigma}_t^{\circ -1/2\prime}(\boldsymbol{\vartheta}_c) \otimes \boldsymbol{\Sigma}_t^{\circ -1/2\prime}(\boldsymbol{\vartheta}_c)] vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\vartheta})\right],$$
$$\frac{\partial s_{\vartheta_i t}(\boldsymbol{\vartheta}_c, \bar{\vartheta}_i; \mathbf{0})}{\partial \vartheta_i} = h_{\vartheta_i \vartheta_i t}(\boldsymbol{\vartheta}, \mathbf{0}) = \frac{1}{\bar{\vartheta}_i^2} [\varsigma_t(\boldsymbol{\vartheta}) - N] - \frac{N}{2\bar{\vartheta}_i^2}.$$

But $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_0) = \boldsymbol{\varepsilon}_t$ because we are evaluating these two expressions at consistent estimators of both $\boldsymbol{\vartheta}_c$ and ϑ_i, whence we can prove that

$$\begin{aligned} \mathcal{A}_{\boldsymbol{\vartheta}_c \bar{\vartheta}_i}(\boldsymbol{\vartheta}_0, \mathbf{0}; \boldsymbol{\varphi}_0) &= \mathcal{A}_{\boldsymbol{\vartheta}_c \vartheta_i}(\boldsymbol{\vartheta}_0, \mathbf{0}; \boldsymbol{\varphi}_0) = \mathbf{W}_{\boldsymbol{\vartheta}_c}(\boldsymbol{\vartheta}_{c0}; \boldsymbol{\varphi}_0) \frac{1}{\vartheta_{i0}}, \\ \mathcal{A}_{\bar{\vartheta}_i \bar{\vartheta}_i}(\boldsymbol{\vartheta}_0, \mathbf{0}; \boldsymbol{\varphi}_0) &= \mathcal{A}_{\vartheta_i \vartheta_i}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = \frac{N}{2\vartheta_{i0}^2}. \end{aligned} \tag{B36}$$

Finally, we need to find out the asymptotic variance of the sample average of $s_{\vartheta_i t}(\boldsymbol{\vartheta}_0, \mathbf{0})$ as well as its asymptotic covariance with the sample averages of $\mathbf{s}_{\boldsymbol{\vartheta}_c t}(\boldsymbol{\phi}_\infty)$, $s_{\vartheta_i t}(\boldsymbol{\phi}_\infty)$ and $\mathbf{s}_{\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty)$, which coincide with contemporaneous variance and covariances of these influence functions because they are all martingale difference sequences.

The definition of the coefficient of multivariate excess kurtosis in (11) immediately implies that

$$\mathcal{B}_{\bar{\vartheta}_i \bar{\vartheta}_i}(\boldsymbol{\varphi}_0) = E[s^2_{\bar{\vartheta}_i t}(\boldsymbol{\vartheta}_0, \mathbf{0}) | \boldsymbol{\varphi}_0] = \frac{N[(N+2)\kappa_0 + 2]}{4\theta_{i0}^2}. \tag{B37}$$

Tedious algebraic manipulations also show that

$$\begin{aligned} \mathcal{B}_{\boldsymbol{\vartheta}_c \bar{\vartheta}_i}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) &= E[\mathbf{s}_{\boldsymbol{\vartheta}_c t}(\boldsymbol{\phi}_\infty) s_{\vartheta_i t}(\boldsymbol{\vartheta}_0, \mathbf{0}) | \boldsymbol{\varphi}_0] = \frac{N}{2\vartheta_{i0}} \mathbf{W}_{\boldsymbol{\vartheta}_c}(\boldsymbol{\vartheta}_{c0}; \boldsymbol{\varphi}_0) \mathrm{M}^O_{s\bar{s}}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0), \\ \mathcal{B}_{\vartheta_i \bar{\vartheta}_i}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) &= E[s_{\vartheta_i t}(\boldsymbol{\phi}_\infty) s_{\vartheta_i t}(\boldsymbol{\vartheta}_0, \mathbf{0}) | \boldsymbol{\varphi}_0] = \frac{N^2}{4\vartheta_{i0}\vartheta_{i\infty}} \mathrm{M}^O_{s\bar{s}}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0), \\ \mathcal{B}_{\boldsymbol{\eta} \bar{\vartheta}_i}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) &= E[\mathbf{s}_{\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty) s_{\vartheta_i t}(\boldsymbol{\vartheta}_0, \mathbf{0}) | \boldsymbol{\varphi}_0] = \frac{N}{2\vartheta_{i0}} \mathrm{M}^O_{r\bar{s}}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \end{aligned}$$

with

$$\mathrm{M}^O_{s\bar{s}}(\boldsymbol{\phi}; \boldsymbol{\varphi}) = E\left[\left\{ \delta[\varsigma_t(\boldsymbol{\vartheta}), \boldsymbol{\eta}] \frac{\varsigma_t(\boldsymbol{\vartheta})}{N} - 1 \right\} \left(\frac{\varsigma_t}{N} - 1 \right) \middle| \boldsymbol{\varphi}_0 \right],$$
$$\mathrm{M}^O_{r\bar{s}}(\boldsymbol{\phi}; \boldsymbol{\varphi}_0) = E\left[\mathbf{e}_{rt}(\boldsymbol{\phi}) \left(\frac{\varsigma_t}{N} - 1 \right) \middle| \boldsymbol{\varphi}_0 \right].$$

Finally, it follows from the above expressions that the condition for block-diagonality of $\mathcal{A}$ and $\mathcal{B}$ between $\boldsymbol{\vartheta}_c$ and $(\vartheta_i, \boldsymbol{\eta}, \bar{\vartheta}_i)$ is $\mathbf{W}_{\boldsymbol{\vartheta}_c}(\boldsymbol{\vartheta}_{c0}; \boldsymbol{\varphi}_0) = \mathbf{0}$ regardless of the values of $\vartheta_{i\infty}$ and $\boldsymbol{\eta}_\infty$ because $\mathbf{W}_{\boldsymbol{\vartheta}_c t}(\boldsymbol{\vartheta}_{c0}, \vartheta_{i\infty})$ does not depend of those parameters in view of (B23). □

Proposition 3

We can directly work in terms of the $\boldsymbol{\psi}$ parameters thanks to our assumptions on the mapping $\mathbf{r}_g(.)$. Let us initially keep $\boldsymbol{\eta}$ fixed to some admissible value. It immediately follows from reparametrisation 2 that

$$\begin{aligned}
\partial\boldsymbol{\mu}_t(\boldsymbol{\psi})/\partial\boldsymbol{\psi}_c' &= \partial\boldsymbol{\mu}_t^{\diamond}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}_c' + (\boldsymbol{\psi}_{im}'\otimes\mathbf{I}_N)\partial vec[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)]/\partial\boldsymbol{\psi}_c', \\
\partial\boldsymbol{\mu}_t(\boldsymbol{\psi})/\partial\boldsymbol{\psi}_{im}' &= \boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c), \\
\partial\boldsymbol{\mu}_t(\boldsymbol{\psi})/\partial\boldsymbol{\psi}_{ic}' &= \mathbf{0},
\end{aligned}$$

and

$$\begin{aligned}
\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\psi})]/\partial\boldsymbol{\psi}_c' &= (\mathbf{I}_{N^2}+\mathbf{K}_{NN})[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)\boldsymbol{\Psi}_{ic}\otimes\mathbf{I}_N]\partial vec[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)]/\partial\boldsymbol{\psi}_c', \\
\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\psi})]/\partial\boldsymbol{\psi}_{im}' &= \mathbf{0}, \\
\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\psi})]/\partial\boldsymbol{\psi}_{ic}' &= [\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)\otimes\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)]\mathbf{D}_N.
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbf{Z}_{\boldsymbol{\psi}_c lt}(\boldsymbol{\psi}) &= \frac{\partial\boldsymbol{\mu}_t^{\diamond\prime}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}_c} + \frac{\partial vec'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)]}{\partial\boldsymbol{\psi}_c}(\boldsymbol{\psi}_{im}'\otimes\mathbf{I}_N)\boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\boldsymbol{\psi}_c)\boldsymbol{\Psi}_{ic}^{-1/2\prime}, \\
\mathbf{Z}_{\boldsymbol{\psi}_{im} lt}(\boldsymbol{\psi}) &= \boldsymbol{\Psi}_{ic}^{-1/2\prime}, \\
\mathbf{Z}_{\boldsymbol{\psi}_{ic} lt}(\boldsymbol{\psi}) &= \mathbf{0},
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{Z}_{\boldsymbol{\psi}_c st}(\boldsymbol{\psi}) &= \frac{\partial vec'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)]}{\partial\boldsymbol{\psi}_c}[\boldsymbol{\Psi}_{ic}^{1/2}\otimes\boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\boldsymbol{\psi}_c)\boldsymbol{\Psi}_{ic}^{-1/2\prime})(\mathbf{I}_{N^2}+\mathbf{K}_{NN}), \\
\mathbf{Z}_{\boldsymbol{\psi}_{im} st}(\boldsymbol{\psi}) &= \mathbf{0}, \\
\mathbf{Z}_{\boldsymbol{\psi}_{ic} st}(\boldsymbol{\psi}) &= \mathbf{D}_N'(\boldsymbol{\Psi}_{ic}^{-1/2\prime}\otimes\boldsymbol{\Psi}_{ic}^{-1/2\prime}).
\end{aligned}$$

As a result,

$$\begin{aligned}
\mathbf{s}_{\psi_c t}(\boldsymbol{\psi},\boldsymbol{\eta}) = \left[\frac{\partial\boldsymbol{\mu}_t^{\diamond\prime}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}_c} + \frac{\partial vec'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)]}{\partial\boldsymbol{\psi}_c}(\boldsymbol{\psi}_{im}'\otimes\mathbf{I}_N)\boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\boldsymbol{\psi}_c)\boldsymbol{\Psi}_{ic}^{-1/2\prime}\right]\delta[\varsigma_t(\boldsymbol{\psi}),\boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}) \\
+\frac{\partial vec'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)]}{\partial\boldsymbol{\psi}_c}[\boldsymbol{\Psi}_{ic}^{1/2}\otimes\boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\boldsymbol{\psi}_c)\boldsymbol{\Psi}_{ic}^{-1/2\prime})vec\left\{\delta[\varsigma_t(\boldsymbol{\psi}),\boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi})-\mathbf{I}_N\right\}, \\
\mathbf{s}_{\psi_{ic} t}(\boldsymbol{\psi},\boldsymbol{\eta}) = \frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Psi}_{ic}^{-1/2\prime}\otimes\boldsymbol{\Psi}_{ic}^{-1/2\prime})vec\left\{\delta[\varsigma_t(\boldsymbol{\psi}),\boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi})-\mathbf{I}_N\right\}, \\
\mathbf{s}_{\psi_{im} t}(\boldsymbol{\psi},\boldsymbol{\eta}) = \frac{1}{2}\boldsymbol{\Psi}_{ic}^{-1/2\prime}\delta[\varsigma_t(\boldsymbol{\psi}),\boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}).
\end{aligned}$$

Let $\boldsymbol{\psi}_{im\infty}(\boldsymbol{\eta})$ and $\boldsymbol{\psi}_{ic\infty}(\boldsymbol{\eta}) = vech[\boldsymbol{\Psi}_{ic\infty}(\boldsymbol{\eta})]$, with $\boldsymbol{\Psi}_{ic\infty}(\boldsymbol{\eta})$ p.d., denote the solution to the implicit system of $N(N+3)/2$ equations

$$\left.\begin{aligned}
E[\mathbf{s}_{\boldsymbol{\psi}_{im} t}(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic},\boldsymbol{\eta})|\boldsymbol{\varphi}_0] = \mathbf{0} \\
E[\mathbf{s}_{\boldsymbol{\psi}_{ic} t}(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic},\boldsymbol{\eta})|\boldsymbol{\varphi}_0] = \mathbf{0}
\end{aligned}\right\}. \tag{B38}$$

The time-invariance of $\mathbf{Z}_{\boldsymbol{\psi}_{im}st}(\boldsymbol{\psi})$ and $\mathbf{Z}_{\boldsymbol{\psi}_{ic}lt}(\boldsymbol{\psi})$ implies that $\boldsymbol{\psi}_{im\infty}(\boldsymbol{\eta})$ and $\boldsymbol{\psi}_{ic\infty}(\boldsymbol{\eta})$ will also solve the alternative system of $N(N+3)/2$ equations

$$\left.\begin{array}{c} E\{\delta[\varsigma_t(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic},\boldsymbol{\eta}),\boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic},\boldsymbol{\eta})|\boldsymbol{\varphi}_0\}=\mathbf{0} \\ E\{vech\,\{\delta[\varsigma_t(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic},\boldsymbol{\eta}),\boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic},\boldsymbol{\eta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic},\boldsymbol{\eta})-\mathbf{I}_N\}\,|\boldsymbol{\varphi}_0\}=\mathbf{0} \end{array}\right\}.$$

Given that

$$\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi})=\boldsymbol{\Psi}_{ic}^{-1/2}\boldsymbol{\Sigma}_t^{\diamond-1/2}(\boldsymbol{\psi}_c)[\mathbf{y}_t-\boldsymbol{\mu}_t^{\diamond}(\boldsymbol{\psi}_c)-\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)\boldsymbol{\psi}_{im}]=\boldsymbol{\Psi}_{ic}^{-1/2}[\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\psi}_c)-\boldsymbol{\psi}_{im}], \tag{B39}$$

with $\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\psi}_c)$ defined in (13), so that

$$\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\psi}_{c0})=\boldsymbol{\psi}_{im0}+\boldsymbol{\Psi}_{ic0}^{1/2}\boldsymbol{\varepsilon}_t^*,$$

we can immediately see that the pseudo standardised residuals $\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\psi}_{c0})$ will be $i.i.d.(\boldsymbol{\psi}_{im},\boldsymbol{\Psi}_{ic})$ conditional on I_{t-1}. Moreover, instantaneous transformations of $\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\psi}_{c0})$ such as

$$\begin{array}{c}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic})=\boldsymbol{\Psi}_{ic}^{-1/2}(\boldsymbol{\psi}_{im0}-\boldsymbol{\psi}_{im})+\boldsymbol{\Psi}_{ic}^{-1/2}\boldsymbol{\Psi}_{ic0}^{1/2}\boldsymbol{\varepsilon}_t^*, \\ \varsigma_t(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic})=\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic}),\end{array}$$

$\mathbf{e}_{lt}(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic},\boldsymbol{\eta})$ and $\mathbf{e}_{st}(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic},\boldsymbol{\eta})$ will also be $i.i.d.$ As a result, the law of iterated expectations implies that

$$E\{\mathbf{s}_{\boldsymbol{\psi}_c t}[\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im\infty}(\boldsymbol{\eta}),\boldsymbol{\psi}_{ic\infty}(\boldsymbol{\eta}),\boldsymbol{\eta}]|I_{t-1};\boldsymbol{\varphi}_0\}=\mathbf{0}, \tag{B40}$$

which confirms that $\boldsymbol{\psi}_{c0}$, $\boldsymbol{\psi}_{im\infty}(\boldsymbol{\eta})$ and $\boldsymbol{\psi}_{ic\infty}(\boldsymbol{\eta})$ will indeed be the pseudo-true values corresponding to a restricted PML estimator that keeps $\boldsymbol{\eta}$ fixed.

If we define $\boldsymbol{\eta}_\infty$ as the solution to the q equations

$$E\{\mathbf{s}_{\boldsymbol{\eta}t}[\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im\infty}(\boldsymbol{\eta}),\boldsymbol{\psi}_{ic\infty}(\boldsymbol{\eta}),\boldsymbol{\eta}]|\boldsymbol{\varphi}_0\}=\mathbf{0}$$

which we assume lies in the interior of the admissible parameter space, then it is clear that $\boldsymbol{\psi}_{c0}$, $\boldsymbol{\psi}_{im\infty}=\boldsymbol{\psi}_{im\infty}(\boldsymbol{\eta}_\infty)$, $\boldsymbol{\psi}_{ic\infty}=\boldsymbol{\psi}_{ic\infty}(\boldsymbol{\eta}_\infty)$ and $\boldsymbol{\eta}_\infty$ will be the pseudo-true values of the parameters corresponding to an unrestricted non-Gaussian PMLE that also estimates $\boldsymbol{\eta}$. □

Proposition 4

This proposition is a special case of Proposition 12, so we omit its proof.

Proposition 5

The consistency of the Gaussian PML derives from the fact that $E[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,0)|I_{t-1};\boldsymbol{\varphi}_0]=\mathbf{0}$. Thus, if the pseudo-true value of η, η_∞ say, is 0, then the Student t based pseudo-true values of the conditional mean and variance parameters, $\boldsymbol{\theta}_\infty$ say, will coincide with their true values $\boldsymbol{\theta}_0$

by the law of iterated expectations. But since η is estimated subject to the inequality constraint $\eta \geq 0$, the population KT conditions that define η_∞ will be

$$E[s_{\eta t}(\boldsymbol{\theta}_\infty, \eta_\infty)|\boldsymbol{\varphi}_0] + \upsilon_{\eta\infty} = 0; \quad \eta_\infty \geq 0; \quad \upsilon_{\eta\infty} \geq 0; \quad \eta_\infty \cdot \upsilon_{\eta\infty} = 0,$$

where $\upsilon_{\eta\infty}$ is the pseudo-true value of the KT multiplier, and the expectation is taken with respect to the true unconditional distribution of the observations (see Calzolari, Fiorentini and Sentana (2004)). Hence, $\eta_\infty = 0$ if and only if $E[s_{\eta t}(\boldsymbol{\theta}_0, 0)|\boldsymbol{\varphi}_0] \leq 0$.

Fiorentini, Sentana and Calzolari (2003) show that in the multivariate Student t case $s_{\eta t}(\boldsymbol{\theta}_0, 0)$ it is proportional to the second generalised Laguerre polynomial (19). Given that $\varsigma_t(\boldsymbol{\theta}_0) = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$, we can write

$$\begin{aligned} s_{\eta t}(\boldsymbol{\theta}_0, 0) &= \frac{N(N+2)}{4} - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}_0) + \frac{1}{4}\varsigma_t^2(\boldsymbol{\theta}_0) \\ &= \frac{N(N+2)}{4}\left[\frac{(\boldsymbol{\varepsilon}_{\mathbf{t}}^{*\prime}\boldsymbol{\varepsilon}_t^*)^2}{N(N+2)} - 1\right] + \frac{N+2}{2}[(\boldsymbol{\varepsilon}_{\mathbf{t}}^{*\prime}\boldsymbol{\varepsilon}_t^*) - N]. \end{aligned}$$

But since we have normalised the innovations so that $E(\boldsymbol{\varepsilon}_{\mathbf{t}}^*\boldsymbol{\varepsilon}_t^{*\prime}|I_{t-1}; \boldsymbol{\varphi}_0) = \mathbf{I}_N$, then

$$N = tr(\mathbf{I}_N) = tr[E(\boldsymbol{\varepsilon}_{\mathbf{t}}^*\boldsymbol{\varepsilon}_t^{*\prime}|I_{t-1}; \boldsymbol{\varphi}_0)] = E[tr(\boldsymbol{\varepsilon}_{\mathbf{t}}^*\boldsymbol{\varepsilon}_t^{*\prime})|I_{t-1}; \boldsymbol{\varphi}_0] = E(\boldsymbol{\varepsilon}_{\mathbf{t}}^{*\prime}\boldsymbol{\varepsilon}_t^*|I_{t-1}; \boldsymbol{\varphi}_0)$$

by the linearity of the expectation and trace operators. Therefore, it immediately follows that

$$\upsilon_{\eta\infty} = \min\{0, -E[s_{\eta t}(\boldsymbol{\theta}_0, 0)|\boldsymbol{\varphi}_0]\} = \min\left\{0, -\frac{N(N+2)}{4}\kappa_0\right\}$$

in view of the definition of κ_0 in (11). Therefore, $\eta_\infty = 0$ if and only if $\kappa_0 \leq 0$.

To prove the second and third parts, we can use Propositions 1 and 2 in Calzolari, Fiorentini and Sentana (2004) if we regard the Student t based estimator $\hat{\boldsymbol{\phi}}_T$ as the "inequality restricted" PML estimator of $\boldsymbol{\phi}$, and the Gaussian-based estimator $\tilde{\boldsymbol{\phi}}_T = (\tilde{\boldsymbol{\theta}}_T, 0)$ as its "equality restricted" counterpart, both of which share not only the pseudo-true values $(\boldsymbol{\theta}_0, 0, \upsilon_{\eta\infty})$ when $\kappa_0 \leq 0$, but also the modified pseudo-score $\mathbf{m}_t(\boldsymbol{\theta}_0, 0, \upsilon_{\eta\infty}) = s_{\boldsymbol{\phi}t}(\boldsymbol{\theta}_0, 0) + \mathbf{e}_{p+1} \cdot \upsilon_{\eta\infty}$, where $\mathbf{e}_{p+1}$ is the $(p+1)^{th}$ column of $\mathbf{I}_{p+1}$, as well as the expected value of the average Hessian $\mathcal{A}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = -E[\bar{\mathbf{h}}_{\boldsymbol{\phi}\boldsymbol{\phi}T}(\boldsymbol{\phi}_0)|\boldsymbol{\varphi}_0]$.

Specifically, Proposition 1 in Calzolari, Fiorentini and Sentana (2004) implies here that

$$\upsilon_{\eta\infty} \cdot \sqrt{T}\hat{\eta}_T = o_p(1),$$

while their Proposition 2 implies that

$$\begin{aligned} \begin{bmatrix} \mathcal{A}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{A}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\ \mathcal{A}'_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{A}_{\eta\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \end{bmatrix} \sqrt{T}\begin{pmatrix} \hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0 \\ \hat{\eta}_T \end{pmatrix} + \mathbf{e}_{p+1}\sqrt{T}(\hat{\upsilon}_{\eta T} - \upsilon_{\eta\infty}) \\ -\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_0, 0, \upsilon_{\eta\infty}) = o_p(1), \\ \begin{bmatrix} \mathcal{A}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{A}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\ \mathcal{A}'_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{A}_{\eta\eta}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \end{bmatrix} \sqrt{T}\begin{pmatrix} \tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0 \\ 0 \end{pmatrix} + \mathbf{e}_{p+1}\sqrt{T}(\tilde{\upsilon}_{\eta T} - \upsilon_{\eta\infty}) \\ -\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_0, 0, \upsilon_{\eta\infty}) = o_p(1), \end{aligned}$$

where $\hat{\upsilon}_{\eta T}$ and $\tilde{\upsilon}_{\eta T}$ are the sample versions of the KT and Lagrange multipliers associated to the constraint $\eta = 0$. As a consequence,

$$\left[\begin{array}{cc} \mathcal{A}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \mathcal{A}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ \mathcal{A}'_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \mathcal{A}_{\eta\eta}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \end{array} \right] \sqrt{T} \left(\begin{array}{c} \hat{\boldsymbol{\theta}}_T - \tilde{\boldsymbol{\theta}}_T \\ \hat{\eta}_T \end{array} \right) + \mathbf{e}_{p+1}\sqrt{T}(\hat{\upsilon}_{\eta T} - \tilde{\upsilon}_{\eta T}) = o_p(1).$$

Part 2 immediately follows from the fact that $\upsilon_{\eta\infty} > 0$ when $\kappa_0 < 0$. Similarly, the first statement of Part 3 follows from the fact that $\upsilon_{\eta\infty} = 0$ when $\kappa_0 = 0$. As for the condition (18), which derives directly from the expression for $\mathbf{h}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi})$ in Fiorentini, Sentana and Calzolari (2003) evaluated at $(\boldsymbol{\theta}_0, 0)$, its role is to guarantee that $\mathcal{A}_{\boldsymbol{\theta}\eta}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) = \mathbf{0}$. In this sense, it is worth mentioning that condition (18) will be satisfied for instance if $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}_0$ is *i.i.d.* $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ with $\kappa_0 = 0$ irrespective of whether or not it is Gaussian because in that case

$$E\{[N + 2 - \varsigma_t(\boldsymbol{\theta}_0)]\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)|I_{t-1};\boldsymbol{\theta}_0,\boldsymbol{\eta}_0] = E[(N + 2 - \varsigma_t)\sqrt{\varsigma_t}\mathbf{u}_t|\boldsymbol{\eta}_0] = \mathbf{0}$$

by the serial and mutual independence of ς_t and $\mathbf{u}_t$, and the fact that $E(\mathbf{u}_t) = \mathbf{0}$, while

$$\begin{aligned} E\{[N + 2 - \varsigma_t(\boldsymbol{\theta}_0)]\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)|I_{t-1},\boldsymbol{\phi}_0\} &= E[(N + 2 - \varsigma_t)\varsigma_t\mathbf{u}_t\mathbf{u}_t'|\boldsymbol{\eta}_0] \\ &= N^{-1}E[(N + 2 - \varsigma_t)\varsigma_t|\boldsymbol{\eta}_0]\mathbf{I}_N = \mathbf{0} \end{aligned}$$

by the definition of κ_0 and the fact that $E(\mathbf{u}_t\mathbf{u}_t') = N^{-1}\mathbf{I}_N$. □

Proposition 6

The proof essentially applies the results in the proof of Proposition 12 to model (15). Specifically, expressions (C64) and (C66) become

$$\mathbf{Z}_{lt}(\boldsymbol{\psi}) = \frac{\partial\mu_t(\boldsymbol{\psi})/\partial\boldsymbol{\psi}}{\psi_{ic}^{1/2}\sigma_t^\diamond(\boldsymbol{\psi}_c)} = \frac{1}{\psi_{ic}^{1/2}\sigma_t^\diamond(\boldsymbol{\psi}_c)} \left[\begin{array}{c} \frac{1}{2}\psi_{im}\sigma_t^{\diamond -1}(\boldsymbol{\psi}_c)\partial\sigma_t^{\diamond 2}(\boldsymbol{\psi}_c)/\partial\boldsymbol{\psi}_c \\ \sigma_t^\diamond(\boldsymbol{\psi}_c) \\ 0 \end{array} \right] = \left[\begin{array}{c} \psi_{im}\psi_{ic}^{-1/2}\mathbf{W}_{\boldsymbol{\psi}_c t}(\boldsymbol{\psi}_c) \\ \psi_{ic}^{-1/2} \\ 0 \end{array} \right],$$

$$\mathbf{Z}_{st}(\boldsymbol{\psi}) = \frac{\partial\sigma_t^2(\boldsymbol{\psi})/\partial\boldsymbol{\psi}}{2\psi_{ic}\sigma_t^{\diamond 2}(\boldsymbol{\psi}_c)} = \frac{1}{2\psi_{ic}\sigma_t^{\diamond 2}(\boldsymbol{\psi}_c)} \left[\begin{array}{c} \psi_{ic}\partial\sigma_t^{\diamond 2}(\boldsymbol{\psi}_c)/\partial\boldsymbol{\psi}_c \\ 0 \\ \sigma_t^{\diamond 2}(\boldsymbol{\psi}_c) \end{array} \right] = \left[\begin{array}{c} \mathbf{W}_{\boldsymbol{\psi}_c t}(\boldsymbol{\psi}_c;\boldsymbol{\varphi}) \\ 0 \\ \frac{1}{2}\psi_{ic}^{-1} \end{array} \right]$$

and

$$\begin{aligned} e_{lt}(\boldsymbol{\psi},\boldsymbol{\varrho}) &= -\frac{\partial \ln f[\epsilon_t(\boldsymbol{\psi});\boldsymbol{\rho}]}{\partial\varepsilon}, \\ e_{st}(\boldsymbol{\psi},\boldsymbol{\varrho}) &= -\left\{1 + \epsilon_t(\boldsymbol{\psi})\frac{\partial \ln f[\epsilon_t(\boldsymbol{\psi});\boldsymbol{\rho}]}{\partial\varepsilon}\right\}, \end{aligned}$$

respectively, where

$$\epsilon_t(\boldsymbol{\psi}) = \frac{\epsilon_t^\diamond(\boldsymbol{\psi}_c) - \psi_{im}}{\psi_{ic}^{1/2}} = \frac{x_t}{\psi_{ic}^{1/2}\sigma_t^\diamond(\boldsymbol{\psi}_c)} - \frac{\psi_{im}}{\psi_{ic}^{1/2}} = \frac{x_t - \psi_{im}\sigma_t^\diamond(\boldsymbol{\psi}_c)}{\psi_{ic}^{1/2}\sigma_t^\diamond(\boldsymbol{\psi}_c)} \tag{B41}$$

and

$$\mathbf{W}_{\boldsymbol{\psi}_c t}(\boldsymbol{\psi}_c) = \frac{1}{2\sigma_t^{\diamond 2}(\boldsymbol{\psi}_c)} \frac{\partial \sigma_t^{\diamond 2}(\boldsymbol{\psi}_c)}{\partial \boldsymbol{\psi}_c}. \tag{B42}$$

Then, a direct application of (C65) yields

$$\mathbf{s}_{\boldsymbol{\psi} t}(\boldsymbol{\phi}) = \left[\begin{array}{cc} \mathbf{Z}_{lt}(\boldsymbol{\psi}) & \mathbf{Z}_{st}(\boldsymbol{\psi}) \end{array} \right] \left[\begin{array}{c} e_{lt}(\boldsymbol{\psi}, \boldsymbol{\varrho}) \\ e_{st}(\boldsymbol{\psi}, \boldsymbol{\varrho}) \end{array} \right] = \left[\begin{array}{c} \mathbf{W}_t(\boldsymbol{\psi}_c)\mathbf{r}'(\boldsymbol{\psi}_i) \\ \boldsymbol{\Delta}(\psi_{ic}) \end{array} \right] \left[\begin{array}{c} e_{lt}(\boldsymbol{\psi}, \boldsymbol{\varrho}) \\ e_{st}(\boldsymbol{\psi}, \boldsymbol{\varrho}) \end{array} \right], \tag{B43}$$

where

$$\mathbf{r}(\boldsymbol{\psi}_i) = \left(\begin{array}{cc} \psi_{im}\psi_{ic}^{-1/2} & 1 \end{array} \right)'$$

and

$$\boldsymbol{\Delta}(\psi_{ic}) = \left(\begin{array}{cc} \psi_{ic}^{-1/2} & 0 \\ 0 & \frac{1}{2}\psi_{ic}^{-1} \end{array} \right).$$

Let us now define $\boldsymbol{\psi}_i(\boldsymbol{\varrho}) = [\psi_{im\infty}(\boldsymbol{\varrho}), \psi_{ic\infty}(\boldsymbol{\varrho})]$ as the values of ψ_{im} and ψ_{ic} that simultaneously solve the equations

$$E\left[\frac{\partial \ln f\{\epsilon_t[\boldsymbol{\psi}_{c0}, \boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]; \boldsymbol{\rho}\}}{\partial \varepsilon} \middle| \boldsymbol{\varphi}_0 \right] = 0, \tag{B44}$$

$$E\left[1 + \epsilon_t[\boldsymbol{\psi}_{c0}, \boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})] \frac{\partial \ln f\{\epsilon_t[\boldsymbol{\psi}_{c0}, \boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]; \boldsymbol{\rho}\}}{\partial \varepsilon} \middle| \boldsymbol{\varphi}_0 \right] = 0. \tag{B45}$$

In what follows, we shall refer to the ratio

$$\lambda_\infty(\boldsymbol{\varrho}) = \psi_{ic\infty}(\boldsymbol{\varrho})/\psi_{ic0} \tag{B46}$$

as the "relative scale bias" in estimating ψ_{ic0}, and to

$$\mu_\infty(\boldsymbol{\varrho}) = \frac{\psi_{im\infty}(\boldsymbol{\varrho}) - \psi_{im0}}{\psi_{ic\infty}^{1/2}(\boldsymbol{\varrho})} \tag{B47}$$

as the "relative mean bias" in estimating ψ_{im0}, so that

$$\begin{aligned} \epsilon_t[\boldsymbol{\psi}_{c0}, \psi_{im\infty}(\boldsymbol{\varrho}), \psi_{ic\infty}(\boldsymbol{\varrho})] &= \frac{\epsilon_t^\diamond(\boldsymbol{\psi}_{c0})}{\psi_{ic\infty}^{1/2}(\boldsymbol{\varrho})} - \frac{\psi_{im\infty}(\boldsymbol{\varrho})}{\psi_{ic\infty}^{1/2}(\boldsymbol{\varrho})} \\ = \sqrt{\frac{\psi_{ic0}}{\psi_{ic\infty}(\boldsymbol{\varrho})}}\varepsilon_t^* - \left[\frac{\psi_{im\infty}(\boldsymbol{\varrho}) - \psi_{im0}}{\psi_{ic\infty}^{1/2}(\boldsymbol{\varrho})} \right] &= \lambda_\infty^{-1/2}(\boldsymbol{\varrho})\varepsilon_t^* - \mu_\infty(\boldsymbol{\varrho}). \end{aligned} \tag{B48}$$

We will also make extensive use of $\mathbf{W}_{\boldsymbol{\psi}_c}(\boldsymbol{\psi}_c; \boldsymbol{\varphi}) = E[\mathbf{W}_{\boldsymbol{\psi}_c t}(\boldsymbol{\psi}_c)|\boldsymbol{\varphi}]$ and $\mathbf{V}_{\boldsymbol{\psi}_c}(\boldsymbol{\psi}_c; \boldsymbol{\varphi}) = V[\mathbf{W}_{\boldsymbol{\psi}_c t}(\boldsymbol{\psi}_c)|\boldsymbol{\varphi}]$, which are defined in (25) and (21), respectively, which we will shorten to $\mathbf{W}$ and $\mathbf{V}$ for the sake of brevity.

Given (C83), the expected Hessian is

$$\begin{aligned} \mathcal{A}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) &= E\{\mathbf{h}_{\boldsymbol{\psi}\boldsymbol{\psi} t}[\boldsymbol{\psi}_{c0}, \boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho}), \boldsymbol{\varrho}]|\boldsymbol{\varphi}_0\} = \left\{ \begin{array}{cc} \mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_i}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\ \mathcal{A}'_{\boldsymbol{\psi}_c\boldsymbol{\psi}_i}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{A}_{\boldsymbol{\psi}_i\boldsymbol{\psi}_i}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \end{array} \right\} \\ = E\left[\left\{ \begin{array}{c} \mathbf{W}_{\boldsymbol{\psi}_c t}(\boldsymbol{\psi}_c)\mathbf{r}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})] \\ \boldsymbol{\Delta}[\boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho})] \end{array} \right\} \mathcal{M}_{dd}^H(\boldsymbol{\phi}; \boldsymbol{\varphi}) \left\{ \begin{array}{cc} \mathbf{r}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathbf{W}'_{\boldsymbol{\psi}_c t}(\boldsymbol{\psi}_c) & \boldsymbol{\Delta}[\boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho})] \end{array} \right\} \middle| \boldsymbol{\varphi}_0 \right] \\ = \left\{ \begin{array}{cc} D(\boldsymbol{\phi}; \boldsymbol{\varphi})(\mathbf{V} + \mathbf{W}\mathbf{W}') & \mathbf{W}\mathbf{r}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathcal{M}_{dd}^H(\boldsymbol{\phi}; \boldsymbol{\varphi})\boldsymbol{\Delta}[\boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho})] \\ \boldsymbol{\Delta}[\boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho})]\mathcal{M}_{dd}^H(\boldsymbol{\phi}; \boldsymbol{\varphi})\mathbf{r}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathbf{W}' & \boldsymbol{\Delta}[\boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho})]\mathcal{M}_{dd}^H(\boldsymbol{\phi}; \boldsymbol{\varphi})\boldsymbol{\Delta}[\boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho})] \end{array} \right\}, \end{aligned} \tag{B49}$$

and

$$\begin{aligned} D(\boldsymbol{\phi};\boldsymbol{\varphi}) &= \mathbf{r}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\mathbf{r}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})] \\ &= \left[\frac{\psi^2_{im\infty}(\boldsymbol{\varrho})}{\psi_{ic\infty}(\boldsymbol{\varrho})}\mathcal{M}^H_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) + \frac{2\psi_{im\infty}(\boldsymbol{\varrho})}{\psi^{1/2}_{ic\infty}(\boldsymbol{\varrho})}\mathcal{M}^H_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) + \mathcal{M}^H_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi})\right]. \end{aligned}$$

We can then exploit the block structure of expression (B49) together with the partitioned inverse formula to obtain $\mathcal{A}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi}}$. Specifically, the upper left block of $\mathcal{A}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi}}$, $\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}$ say, will be given by the inverse of

$$\begin{aligned} &D(\boldsymbol{\phi};\boldsymbol{\varphi})(\mathbf{V}+\mathbf{W}\mathbf{W}') - \mathbf{W}\mathbf{r}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\boldsymbol{\Delta}[\psi_{ic}(\boldsymbol{\varrho})] \\ &\times\left\{\boldsymbol{\Delta}[\psi_{ic}(\boldsymbol{\varrho})]\mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\boldsymbol{\Delta}[\psi_{ic}(\boldsymbol{\varrho})]\right\}^{-1}\boldsymbol{\Delta}[\psi_{ic}(\boldsymbol{\varrho})]\mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\mathbf{r}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathbf{W}' = D(\boldsymbol{\phi};\boldsymbol{\varphi})\mathbf{V}. \quad \text{(B50)} \end{aligned}$$

Similarly, the bottom left block of $\mathcal{A}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi}}$ will be given by

$$\begin{aligned} \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c} &= -\left\{\boldsymbol{\Delta}[\psi_{ic}(\boldsymbol{\varrho})]\mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\boldsymbol{\Delta}[\psi_{ic}(\boldsymbol{\varrho})]\right\}^{-1}\boldsymbol{\Delta}[\psi_{ic}(\boldsymbol{\varrho})]\mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\mathbf{r}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathbf{W}'\mathbf{V}^{-1}D^{-1}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ &= -D^{-1}(\boldsymbol{\phi};\boldsymbol{\varphi})\mathbf{c}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathbf{W}'\mathbf{V}^{-1}, \quad \text{(B51)} \end{aligned}$$

where we have exploited that $\mathbf{c}(\boldsymbol{\psi}_i) = \boldsymbol{\Delta}^{-1}(\psi_{ic})\mathbf{r}(\boldsymbol{\psi}_i)$.

Finally, the bottom right block of $\mathcal{A}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi}}$ will be

$$\begin{aligned} \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_i} &= \left\{\boldsymbol{\Delta}[\psi_{ic}(\boldsymbol{\varrho})]\mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\boldsymbol{\Delta}[\psi_{ic}(\boldsymbol{\varrho})]\right\}^{-1} + \left\{\boldsymbol{\Delta}[\psi_{ic}(\boldsymbol{\varrho})]\mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\boldsymbol{\Delta}[\psi_{ic}(\boldsymbol{\varrho})]\right\}^{-1} \\ &\times\boldsymbol{\Delta}[\psi_{ic}(\boldsymbol{\varrho})]\mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\mathbf{r}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathbf{W}'\mathbf{V}^{-1}D^{-1}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ &\times\mathbf{W}\mathbf{r}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\boldsymbol{\Delta}[\psi_{ic}(\boldsymbol{\varrho})]\left\{\boldsymbol{\Delta}[\psi_{ic}(\boldsymbol{\varrho})]\mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\boldsymbol{\Delta}[\psi_{ic}(\boldsymbol{\varrho})]\right\}^{-1} \\ &= \boldsymbol{\Delta}^{-1}[\psi_{ic}(\boldsymbol{\varrho})]\mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\boldsymbol{\Delta}^{-1}[\psi_{ic}(\boldsymbol{\varrho})] + D^{-1}(\boldsymbol{\phi};\boldsymbol{\varphi})\mathbf{c}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathbf{c}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})](\mathbf{W}'\mathbf{V}^{-1}\mathbf{W}). \quad \text{(B52)} \end{aligned}$$

In turn, (C82) implies that the variance of the scores will be

$$\begin{aligned} \mathcal{B}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= V\{\mathbf{s}_{\boldsymbol{\psi}t}[\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho}),\boldsymbol{\varrho}]|\boldsymbol{\varphi}_0\} = \begin{Bmatrix} \mathcal{B}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \mathcal{B}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_i}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ \mathcal{B}'_{\boldsymbol{\psi}_c\boldsymbol{\psi}_i}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \mathcal{B}_{\boldsymbol{\psi}_i\boldsymbol{\psi}_i}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \end{Bmatrix} \\ &= E\left\{\begin{bmatrix} \mathbf{W}_{\boldsymbol{\psi}_c t}(\boldsymbol{\psi}_c)\mathbf{r}'(\boldsymbol{\psi}_i) \\ \boldsymbol{\Delta}(\psi_{ic}) \end{bmatrix}\mathcal{M}^O_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\begin{bmatrix} \mathbf{r}(\boldsymbol{\psi}_i)\mathbf{W}'_{\boldsymbol{\psi}_c t}(\boldsymbol{\psi}_c) & \boldsymbol{\Delta}(\psi_{ic}) \end{bmatrix}\middle|\boldsymbol{\varphi}\right\} \\ &= \begin{Bmatrix} N(\boldsymbol{\phi};\boldsymbol{\varphi})(\mathbf{V}+\mathbf{W}\mathbf{W}') & \mathbf{W}\mathbf{r}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathcal{M}^O_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\boldsymbol{\Delta}[\psi_{ic\infty}(\boldsymbol{\varrho})] \\ \boldsymbol{\Delta}[\psi_{ic\infty}(\boldsymbol{\varrho})]\mathcal{M}^O_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\mathbf{r}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathbf{W}' & \boldsymbol{\Delta}[\psi_{ic\infty}(\boldsymbol{\varrho})]\mathcal{M}^O_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\boldsymbol{\Delta}[\psi_{ic\infty}(\boldsymbol{\varrho})] \end{Bmatrix}, \quad \text{(B53)} \end{aligned}$$

where

$$\begin{aligned} N(\boldsymbol{\phi};\boldsymbol{\varphi}) &= \mathbf{r}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathcal{M}^O_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\mathbf{r}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})] \\ &= \left[\frac{\psi^2_{im\infty}(\boldsymbol{\varrho})}{\psi_{ic\infty}(\boldsymbol{\varrho})}\mathcal{M}^O_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) + \frac{2\psi_{im\infty}(\boldsymbol{\varrho})}{\psi^{1/2}_{ic\infty}(\boldsymbol{\varrho})}\mathcal{M}^O_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) + \mathcal{M}^O_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi})\right]. \end{aligned}$$

Given that the expression for $\mathcal{B}_{\boldsymbol{\psi}\boldsymbol{\psi}}$ in (B53) is entirely analogous to the expression for $\mathcal{A}_{\boldsymbol{\psi}\boldsymbol{\psi}}$ in (B49), except for the matrix $\mathcal{M}^O_{dd}$ replacing the matrix $\mathcal{M}^H_{dd}$, it turns out that $\mathcal{C}_{\boldsymbol{\psi}\boldsymbol{\psi}} =$

$\mathcal{A}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi}}\mathcal{B}_{\boldsymbol{\psi}\boldsymbol{\psi}}\mathcal{A}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi}}$ can be substantially simplified. Specifically,

$$\mathcal{C}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} = \begin{pmatrix} \mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_{c'}} \end{pmatrix} \begin{pmatrix} \mathcal{B}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{B}'_{\boldsymbol{\psi}_i\boldsymbol{\psi}_c} \\ \mathcal{B}_{\boldsymbol{\psi}_i\boldsymbol{\psi}_c} & \mathcal{B}_{\boldsymbol{\psi}_i\boldsymbol{\psi}_i} \end{pmatrix} \begin{pmatrix} \mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} \\ \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c} \end{pmatrix}$$
$$= \mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}\mathcal{B}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} + \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_{c'}}\mathcal{B}_{\boldsymbol{\psi}_i\boldsymbol{\psi}_c}\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} + \mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}\mathcal{B}'_{\boldsymbol{\psi}_i\boldsymbol{\psi}_c}\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c} + \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_{c'}}\mathcal{B}_{\boldsymbol{\psi}_i\boldsymbol{\psi}_i}\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c}.$$

Given the expressions for $\mathcal{B}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}$ in (B53) and the inverse of $\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}$ in (B50), the first term, i.e. $\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}\mathcal{B}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}$, will be

$$S(\boldsymbol{\varrho};\boldsymbol{\rho}_0)\mathbf{V}^{-1}(\mathbf{V}+\mathbf{W}\mathbf{W}')\mathbf{V}^{-1}.$$

In turn, the second and third terms, $\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_{c'}}\mathcal{B}_{\boldsymbol{\psi}_i\boldsymbol{\psi}_c}\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}$ and $\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}\mathcal{B}'_{\boldsymbol{\psi}_i\boldsymbol{\psi}_c}\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c}$, respectively, will be given by

$$-D^{-1}(\boldsymbol{\phi};\boldsymbol{\varphi})\mathbf{V}^{-1}\mathbf{W}\mathbf{c}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\boldsymbol{\Delta}[\psi_{ic\infty}(\boldsymbol{\varrho})]\mathcal{M}^O_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\mathbf{r}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathbf{W}'\mathbf{V}^{-1}D^{-1}(\boldsymbol{\phi};\boldsymbol{\varphi})$$
$$= -S(\boldsymbol{\phi};\boldsymbol{\varphi})\mathbf{V}^{-1}\mathbf{W}\mathbf{W}'\mathbf{V}^{-1}.$$

Finally, the fourth term, i.e. $\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_{c'}}\mathcal{B}_{\boldsymbol{\psi}_i\boldsymbol{\psi}_i}\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c}$, will be

$$D^{-1}(\boldsymbol{\phi};\boldsymbol{\varphi})\mathbf{V}^{-1}\mathbf{W}\mathbf{c}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\boldsymbol{\Delta}[\psi_{ic\infty}(\boldsymbol{\varrho})]\mathcal{M}^O_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})\boldsymbol{\Delta}[\psi_{ic}(\boldsymbol{\varrho})]\mathbf{c}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathbf{W}'\mathbf{V}^{-1}D^{-1}(\boldsymbol{\phi};\boldsymbol{\varphi})$$
$$= S(\boldsymbol{\phi};\boldsymbol{\varphi})\mathbf{V}^{-1}\mathbf{W}\mathbf{W}'\mathbf{V}^{-1}.$$

If we add up all these four terms together, we end up with the expression in the first part of the proposition.

Let us now move to the second part, which deals with the asymptotic covariance matrix of our consistent, closed-form estimators (16) and (17)**.** Proposition 12 implies that the only additional non-zero elements of the expected Jacobian of the moment conditions augmented with $\bar{\boldsymbol{\psi}}_i$ are

$$\mathcal{A}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c} = -\boldsymbol{\Delta}(\psi_{ic0})\mathcal{M}^H_{dd}(\boldsymbol{\psi},\mathbf{0};\boldsymbol{\varphi})\mathbf{r}(\boldsymbol{\psi}_{i0})\mathbf{W}',\ \mathcal{A}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i} = -\boldsymbol{\Delta}(\psi_{ic0})\mathcal{M}^H_{dd}(\boldsymbol{\vartheta},\mathbf{0};\boldsymbol{\varphi})\boldsymbol{\Delta}(\psi_{ic0}), \quad \text{(B54)}$$

where $vecd[\mathcal{M}^H_{dd}(\boldsymbol{\vartheta},\mathbf{0};\boldsymbol{\varphi})] = (1,2)'$. Thus, if we make use of the partitioned inverse formula once again, we get that the last diagonal element of the inverse of the expected Jacobian will be

$$\mathcal{A}^{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i} = \left[\mathcal{A}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i} - \begin{pmatrix} \mathcal{A}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_i} \\ \mathcal{A}'_{\boldsymbol{\psi}_c\boldsymbol{\psi}_i} & \mathcal{A}_{\boldsymbol{\psi}_i\boldsymbol{\psi}_i} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}\right]^{-1} = \mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i} = \boldsymbol{\Delta}^{-1}(\psi_{i0}).$$

As for the $\boldsymbol{\psi}\boldsymbol{\psi}$ block, it will trivially coincide with $\mathcal{A}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi}}$, while the $\boldsymbol{\psi}\bar{\boldsymbol{\psi}}_i$ block will be 0. Finally, the $\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}$ block will be given by

$$\mathcal{A}^{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}} = -\mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\mathcal{A}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}}\mathcal{A}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi}} = -\mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i} \begin{pmatrix} \mathcal{A}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{A}'_{\boldsymbol{\psi}_i\boldsymbol{\psi}_c} \\ \mathcal{A}_{\boldsymbol{\psi}_i\boldsymbol{\psi}_c} & \mathcal{A}_{\boldsymbol{\psi}_i\boldsymbol{\psi}_i} \end{pmatrix}^{-1}$$
$$= -\begin{pmatrix} \mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\mathcal{A}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c}\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\mathcal{A}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c}\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_{c'}} \end{pmatrix} = -\mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\mathcal{A}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c} \begin{pmatrix} \mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_{c'}} \end{pmatrix}.$$

As for the asymptotic covariances of the sample averages of the non-Gaussian scores for $\boldsymbol{\psi}$, $\mathbf{s}_{\boldsymbol{\psi}t}(\boldsymbol{\psi}_{c0}, \boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho}), \boldsymbol{\varrho})$, and the Gaussian scores for $\bar{\boldsymbol{\psi}}_i$, $\mathbf{s}_{\bar{\boldsymbol{\psi}}_i t}(\psi_{c0}, \boldsymbol{\psi}_{i0}, \mathbf{0})$, the same proposition implies that

$$\mathcal{B}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i} = \boldsymbol{\Delta}(\psi_{ic0})\mathcal{M}^{O}_{d\bar{d}}(\boldsymbol{\phi};\boldsymbol{\varphi})\boldsymbol{\Delta}(\psi_{ic0}), \tag{B55}$$

$$\mathcal{B}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c} = \boldsymbol{\Delta}(\psi_{ic0})\mathcal{M}^{O}_{d\bar{d}}(\boldsymbol{\phi};\boldsymbol{\varphi})\mathbf{r}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathbf{W}' \tag{B56}$$

and

$$\mathcal{B}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_i} = \boldsymbol{\Delta}(\psi_{ic0})\mathcal{M}^{O}_{d\bar{d}}(\boldsymbol{\phi};\boldsymbol{\varphi})\boldsymbol{\Delta}[\psi_{ic\infty}(\boldsymbol{\varrho})]. \tag{B57}$$

where the elements of $\mathcal{M}^{O}_{d\bar{d}}(\boldsymbol{\phi};\boldsymbol{\varphi})$ are defined in the proof of Proposition 12 and $vec[\mathcal{M}^{O}_{d\bar{d}}(\boldsymbol{\phi};\boldsymbol{\varphi})] = vec[\mathcal{M}^{O}_{dd}(\boldsymbol{\psi},\mathbf{0};\boldsymbol{\varphi})] = (1,\phi,\kappa-1)'$.

Therefore, the asymptotic variance of $\boldsymbol{\psi}_i(\hat{\boldsymbol{\psi}}_{cT})$ will be given by

$$\begin{aligned}
&\mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\Big(-\mathcal{A}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c}\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} \;\; -\mathcal{A}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c}\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_{c'}} \;\; \mathbf{I}\Big)\begin{pmatrix}\mathcal{B}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{B}'_{\boldsymbol{\psi}_i\boldsymbol{\psi}_c} & \mathcal{B}'_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c}\\ \mathcal{B}_{\boldsymbol{\psi}_i\boldsymbol{\psi}_c} & \mathcal{B}_{\boldsymbol{\psi}_i\boldsymbol{\psi}_i} & \mathcal{B}'_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_i}\\ \mathcal{B}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c} & \mathcal{B}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_i} & \mathcal{B}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\end{pmatrix}\begin{pmatrix}-\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}\mathcal{A}'_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c}\\ -\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c}\mathcal{A}'_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c}\\ \mathbf{I}\end{pmatrix}\mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\\
&= \mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\left[\mathcal{A}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c}\begin{pmatrix}\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_{c'}}\end{pmatrix}\begin{pmatrix}\mathcal{B}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{B}'_{\boldsymbol{\psi}_i\boldsymbol{\psi}_c}\\ \mathcal{B}_{\boldsymbol{\psi}_i\boldsymbol{\psi}_c} & \mathcal{B}_{\boldsymbol{\psi}_i\boldsymbol{\psi}_i}\end{pmatrix}\begin{pmatrix}\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}\\ \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c}\end{pmatrix}\mathcal{A}'_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c}\right]\mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\\
&\qquad -\mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\left[\mathcal{A}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c}\begin{pmatrix}\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_{c'}}\end{pmatrix}\begin{pmatrix}\mathcal{B}'_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c}\\ \mathcal{B}'_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_i}\end{pmatrix}\right]\mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\\
&-\mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\left[\mathcal{A}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c}\begin{pmatrix}\mathcal{B}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c} & \mathcal{B}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_i}\end{pmatrix}\begin{pmatrix}\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}\\ \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c}\end{pmatrix}\mathcal{A}'_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c}\right]\mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i} + \mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\mathcal{B}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}.
\end{aligned}$$

Let us look at each of these terms in turn. The first term will be given by

$$\begin{aligned}
&\mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\mathcal{A}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c}\mathcal{C}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}\mathcal{A}'_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c}\mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\\
&= [\boldsymbol{\Delta}(\psi_{ic0})\mathcal{M}^{H}_{dd}(\boldsymbol{\psi},\mathbf{0};\boldsymbol{\varphi})\boldsymbol{\Delta}(\psi_{ic0})]^{-1}\boldsymbol{\Delta}(\psi_{ic0})\mathcal{M}^{H}_{dd}(\boldsymbol{\psi},\mathbf{0};\boldsymbol{\varphi})\mathbf{r}(\boldsymbol{\psi}_{i0})\mathbf{W}'\mathbf{V}^{-1}S(\boldsymbol{\varrho},\boldsymbol{\varrho})\\
&\times\mathbf{W}\mathbf{r}'(\boldsymbol{\psi}_{i0})\mathcal{M}^{H}_{dd}(\boldsymbol{\psi},\mathbf{0};\boldsymbol{\varphi})\boldsymbol{\Delta}(\psi_{ic0})[\boldsymbol{\Delta}(\psi_{ic0})\mathcal{M}^{H}_{dd}(\boldsymbol{\psi},\mathbf{0};\boldsymbol{\varphi})\boldsymbol{\Delta}(\psi_{ic0})]^{-1}\\
&= S(\boldsymbol{\varrho},\boldsymbol{\varrho})\cdot\mathbf{c}(\boldsymbol{\psi}_{i0})\mathbf{c}'(\boldsymbol{\psi}_{i0})(\mathbf{W}'\mathbf{V}^{-1}\mathbf{W}).
\end{aligned}$$

To obtain the second term, as well as the transpose of the third one, we need

$$\begin{aligned}
&\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}\mathcal{B}'_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_c} + \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_{c'}}\mathcal{B}'_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\psi}_i} = D^{-1}(\boldsymbol{\varrho},\boldsymbol{\varrho})\mathbf{V}^{-1}\mathbf{W}\mathbf{r}'[\boldsymbol{\psi}_{\mathbf{i}\infty}(\boldsymbol{\varrho})]\mathcal{M}^{O\prime}_{d\bar{d}}(\boldsymbol{\phi};\boldsymbol{\varphi})\boldsymbol{\Delta}(\psi_{ic0})\\
&-D^{-1}(\boldsymbol{\varrho};\boldsymbol{\rho}_0)\mathbf{V}^{-1}\mathbf{W}\mathbf{c}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\boldsymbol{\Delta}[\psi_{ic\infty}(\boldsymbol{\varrho})]\mathcal{M}^{O\prime}_{d\bar{d}}(\boldsymbol{\phi};\boldsymbol{\varphi})\boldsymbol{\Delta}(\psi_{ic0}) = \mathbf{0}.
\end{aligned}$$

Finally,

$$\begin{aligned}
&\mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\mathcal{B}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i}\mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i} = [\boldsymbol{\Delta}(\psi_{ic0})\mathcal{M}^{H}_{dd}(\boldsymbol{\psi},\mathbf{0};\boldsymbol{\varphi})\boldsymbol{\Delta}(\psi_{ic0})]^{-1}\boldsymbol{\Delta}(\psi_{ic0})\mathcal{M}^{O}_{d\bar{d}}(\boldsymbol{\phi};\boldsymbol{\varphi})\\
&\times\boldsymbol{\Delta}(\psi_{ic0})[\boldsymbol{\Delta}(\psi_{ic0})\mathcal{M}^{H}_{dd}(\boldsymbol{\psi},\mathbf{0};\boldsymbol{\varphi})\boldsymbol{\Delta}(\psi_{ic0})]^{-1} = \begin{bmatrix}\psi_{ic0} & \phi\psi^{3/2}_{ic0}\\ \phi\psi^{3/2}_{ic0} & (\varkappa-1)\psi^{2}_{ic0}\end{bmatrix}.
\end{aligned}$$

The sum of the four terms confirms the second part of the proposition.

Let us now move to the last part of the proposition, in which $\boldsymbol{\varrho}$ is jointly estimated. Assuming no inequality constraints are binding, we can define its pseudo-true value $\boldsymbol{\varrho}_\infty$ from the equation

$$E\{\mathbf{s}_{\boldsymbol{\varrho}t}[\boldsymbol{\psi}_{c0}, \boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho}), \boldsymbol{\varrho}_\infty]|\boldsymbol{\varphi}_0\} = E\left[\left.\frac{\partial \ln f\{\epsilon_t[\boldsymbol{\psi}_{c0}, \psi_{im\infty}(\boldsymbol{\varrho}_\infty), \psi_{ic\infty}(\boldsymbol{\varrho}_\infty)]; \boldsymbol{\rho}_\infty\}}{\partial \boldsymbol{\varrho}}\right| \boldsymbol{\varphi}_0\right] = \mathbf{0}. \quad \text{(B58)}$$

A direct application of Proposition 12 implies that

$$\mathcal{A}_{\boldsymbol{\psi}\boldsymbol{\varrho}} = -\left[\begin{array}{c} \mathbf{W}\mathbf{r}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})] \\ \boldsymbol{\Delta}(\psi_{ic}) \end{array}\right] \mathcal{M}_{dr}^H(\boldsymbol{\phi}; \boldsymbol{\varphi}), \quad \text{(B59)}$$

and

$$\mathcal{B}_{\boldsymbol{\psi}\boldsymbol{\varrho}} = -\left[\begin{array}{c} \mathbf{W}_{\boldsymbol{\psi}_c t}(\boldsymbol{\psi}_c)\mathbf{r}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})] \\ \boldsymbol{\Delta}(\psi_{ic}) \end{array}\right] \mathcal{M}_{dr}^O(\boldsymbol{\varrho}; \boldsymbol{\rho}),$$

with the elements of $\mathcal{M}_{dr}^H(\boldsymbol{\phi}; \boldsymbol{\varphi})$ and $\mathcal{M}_{dr}^O(\boldsymbol{\phi}; \boldsymbol{\varphi})$ defined in the proof of that proposition.

Now, to invert $\mathcal{A}_{\boldsymbol{\phi}\boldsymbol{\phi}}$, we need to compute

$$\begin{aligned}
&\mathcal{A}_{\boldsymbol{\varrho}\boldsymbol{\varrho}} - \mathcal{A}'_{\boldsymbol{\psi}\boldsymbol{\varrho}}\mathcal{A}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi}}\mathcal{A}_{\boldsymbol{\psi}\boldsymbol{\varrho}} = \mathcal{A}_{\boldsymbol{\varrho}\boldsymbol{\varrho}} - \left(\begin{array}{cc} \mathcal{A}'_{\boldsymbol{\psi}_c\boldsymbol{\varrho}} & \mathcal{A}'_{\boldsymbol{\psi}_i\boldsymbol{\varrho}} \end{array}\right)\left(\begin{array}{cc} \mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_{c'}} \\ \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c} & \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_i} \end{array}\right)\left(\begin{array}{c} \mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\varrho}} \\ \mathcal{A}_{\boldsymbol{\psi}_i\boldsymbol{\varrho}} \end{array}\right) \\
&= \mathcal{A}_{\boldsymbol{\varrho}\boldsymbol{\varrho}} - (\mathcal{A}'_{\boldsymbol{\psi}_c\boldsymbol{\varrho}}\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}\mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\varrho}} + \mathcal{A}'_{\boldsymbol{\psi}_i\boldsymbol{\varrho}}\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c}\mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\varrho}} + \mathcal{A}'_{\boldsymbol{\psi}_c\boldsymbol{\varrho}}\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_{c'}}\mathcal{A}_{\boldsymbol{\psi}_i\boldsymbol{\varrho}} + \mathcal{A}'_{\boldsymbol{\psi}_i\boldsymbol{\varrho}}\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_i}\mathcal{A}_{\boldsymbol{\psi}_i\boldsymbol{\varrho}})
\end{aligned}$$

But

$$\begin{aligned}
\mathcal{A}'_{\boldsymbol{\psi}_c\boldsymbol{\varrho}}\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}\mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\varrho}} &= D^{-1}(\boldsymbol{\varrho}, \boldsymbol{\varrho})(\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})\mathcal{M}_{dr}^{H\prime}(\boldsymbol{\phi}; \boldsymbol{\varphi})\mathbf{r}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathbf{r}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathcal{M}_{dr}^H(\boldsymbol{\phi}; \boldsymbol{\varphi}), \\
\mathcal{A}'_{\boldsymbol{\psi}_i\boldsymbol{\varrho}}\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c}\mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\varrho}} &= -D^{-1}(\boldsymbol{\varrho}, \boldsymbol{\varrho})(\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})\mathcal{M}_{dr}^{H\prime}(\boldsymbol{\phi}; \boldsymbol{\varphi})\mathbf{r}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathbf{r}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathcal{M}_{dr}^H(\boldsymbol{\phi}; \boldsymbol{\varphi}) \\
&= \mathcal{A}'_{\boldsymbol{\psi}_c\boldsymbol{\varrho}}\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_{c'}}\mathcal{A}_{\boldsymbol{\psi}_i\boldsymbol{\varrho}}
\end{aligned}$$

and finally

$$\begin{aligned}
\mathcal{A}'_{\boldsymbol{\psi}_i\boldsymbol{\varrho}}\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_i}\mathcal{A}_{\boldsymbol{\psi}_i\boldsymbol{\varrho}} &= \mathcal{M}_{dr}^{H\prime}(\boldsymbol{\phi}; \boldsymbol{\varphi})\boldsymbol{\Delta}(\psi_{ic})\{\boldsymbol{\Delta}^{-1}[\psi_{ic}(\boldsymbol{\varrho})][\mathcal{M}_{dd}^H(\boldsymbol{\phi}; \boldsymbol{\varphi})]^{-1}\boldsymbol{\Delta}^{-1}[\psi_{ic}(\boldsymbol{\varrho})] \\
&\quad + D^{-1}(\boldsymbol{\varrho}; \boldsymbol{\rho}_0)\mathbf{c}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathbf{c}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})](\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})\}\boldsymbol{\Delta}(\psi_{ic})\mathcal{M}_{dr}^H(\boldsymbol{\phi}; \boldsymbol{\varphi}) \\
&= \mathcal{M}_{dr}^{H\prime}(\boldsymbol{\phi}; \boldsymbol{\varphi})[\mathcal{M}_{dd}^H(\boldsymbol{\phi}; \boldsymbol{\varphi})]^{-1}\mathcal{M}_{dr}^H(\boldsymbol{\phi}; \boldsymbol{\varphi}) \\
&\quad + D^{-1}(\boldsymbol{\varrho}, \boldsymbol{\varrho})(\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})\mathcal{M}_{dr}^{H\prime}(\boldsymbol{\phi}; \boldsymbol{\varphi})\mathbf{r}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathbf{r}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathcal{M}_{dr}^H(\boldsymbol{\phi}; \boldsymbol{\varphi}),
\end{aligned}$$

where we have exploited the expressions for $\mathcal{A}_{\boldsymbol{\psi}\boldsymbol{\varrho}}$ in (B59), $\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}$ in (B50), $\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c}$ in (B51) and $\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_i}$ in (B52).

Hence

$$\begin{aligned}
\vec{\mathcal{A}}^{\boldsymbol{\varrho}\boldsymbol{\varrho}} &= (\mathcal{A}_{\boldsymbol{\varrho}\boldsymbol{\varrho}} - \mathcal{A}'_{\boldsymbol{\psi}\boldsymbol{\varrho}}\mathcal{A}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi}}\mathcal{A}_{\boldsymbol{\psi}\boldsymbol{\varrho}})^{-1} = \{\mathcal{M}_{rr}^H(\boldsymbol{\phi}; \boldsymbol{\varphi}) - \mathcal{M}_{dr}^{H\prime}(\boldsymbol{\phi}; \boldsymbol{\varphi})[\mathcal{M}_{dd}^H(\boldsymbol{\phi}; \boldsymbol{\varphi})]^{-1}\mathcal{M}_{dr}^H(\boldsymbol{\phi}; \boldsymbol{\varphi})\}^{-1} \\
&= [\mathcal{M}^H(\boldsymbol{\phi}; \boldsymbol{\varphi})]^{rr},
\end{aligned}$$

which does not depend at all on the dynamic specification of the model.

Similarly, the $\boldsymbol{\psi\varrho}$ block will be given by the matrix

$$\vec{\mathcal{A}}^{\boldsymbol{\varrho}\boldsymbol{\psi}\prime} = -\mathcal{A}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1}\mathcal{A}_{\boldsymbol{\psi}\boldsymbol{\varrho}}(\mathcal{A}_{\boldsymbol{\varrho}\boldsymbol{\varrho}} - \mathcal{A}'_{\boldsymbol{\psi}\boldsymbol{\varrho}}\mathcal{A}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1}\mathcal{A}_{\boldsymbol{\psi}\boldsymbol{\varrho}})^{-1} = -\begin{pmatrix} \mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c\prime} \\ \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c} & \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_i} \end{pmatrix}\begin{pmatrix} \mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\varrho}} \\ \mathcal{A}_{\boldsymbol{\psi}_i\boldsymbol{\varrho}} \end{pmatrix}[\mathcal{M}^H(\boldsymbol{\phi};\boldsymbol{\varphi})]^{rr}$$

$$= \begin{pmatrix} \mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}\mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\varrho}} + \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c\prime}\mathcal{A}_{\boldsymbol{\psi}_i\boldsymbol{\varrho}} \\ \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c}\mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\varrho}} + \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_i}\mathcal{A}_{\boldsymbol{\psi}_i\boldsymbol{\varrho}} \end{pmatrix}[\mathcal{M}^H(\boldsymbol{\phi};\boldsymbol{\varphi})]^{rr}$$

$$= \begin{Bmatrix} \mathbf{0} \\ \boldsymbol{\Delta}^{-1}[\psi_{ic\infty}(\boldsymbol{\varrho})][\mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})]^{-1}\mathcal{M}^H_{dr}(\boldsymbol{\phi};\boldsymbol{\varphi})[\mathcal{M}^H(\boldsymbol{\phi};\boldsymbol{\varphi})]^{rr} \end{Bmatrix}$$

because

$$\mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c}\mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\varrho}} + \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c\prime}\mathcal{A}_{\boldsymbol{\psi}_i\boldsymbol{\varrho}} = D^{-1}(\boldsymbol{\varrho},\boldsymbol{\varrho})\mathbf{V}^{-1}\mathbf{W}\mathbf{r}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathcal{M}^H_{dr}(\boldsymbol{\phi};\boldsymbol{\varphi})$$
$$-D^{-1}(\boldsymbol{\varrho},\boldsymbol{\varrho})\mathbf{V}^{-1}\mathbf{W}\mathbf{r}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathcal{M}^H_{dr}(\boldsymbol{\phi};\boldsymbol{\varphi}) = \mathbf{0}$$

and

$$\mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c}\mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\varrho}} + \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_i}\mathcal{A}_{\boldsymbol{\psi}_i\boldsymbol{\varrho}} = -D^{-1}(\boldsymbol{\varrho},\boldsymbol{\varrho})\mathbf{c}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathbf{W}'\mathbf{V}^{-1}\mathbf{W}\mathbf{r}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathcal{M}^H_{dr}(\boldsymbol{\phi};\boldsymbol{\varphi})$$
$$+\left\{\boldsymbol{\Delta}^{-1}(\psi_{ic\infty})[\mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})]^{-1}\boldsymbol{\Delta}^{-1}(\psi_{ic\infty}) + \mathbf{c}[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})]\mathbf{c}'[\boldsymbol{\psi}_{i\infty}(\boldsymbol{\varrho})](\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})D^{-1}(\boldsymbol{\varrho},\boldsymbol{\varrho})\right\}$$
$$\times\boldsymbol{\Delta}(\psi_{ic})\mathcal{M}^H_{dr}(\boldsymbol{\phi};\boldsymbol{\varphi}) = \boldsymbol{\Delta}^{-1}(\psi_{ic\infty})[\mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})]^{-1}\mathcal{M}^H_{dr}(\boldsymbol{\phi};\boldsymbol{\varphi}).$$

Finally, the $\boldsymbol{\psi\psi}$ block will be given by the matrix

$$\vec{\mathcal{A}}^{\boldsymbol{\psi}\boldsymbol{\psi}} = \mathcal{A}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1} + \mathcal{A}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1}\mathcal{A}_{\boldsymbol{\psi}\boldsymbol{\varrho}}\mathcal{M}^{Hrr}(\boldsymbol{\phi};\boldsymbol{\varphi})\mathcal{A}'_{\boldsymbol{\psi}\boldsymbol{\varrho}}\mathcal{A}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1} = \begin{pmatrix} \mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c\prime} \\ \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c} & \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_i} \end{pmatrix}$$
$$+\begin{Bmatrix} \mathbf{0} \\ \boldsymbol{\Delta}^{-1}(\psi_{ic\infty})[\mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})]^{-1}\mathcal{M}^H_{dr}(\boldsymbol{\phi};\boldsymbol{\varphi})[\mathcal{M}^H(\boldsymbol{\phi};\boldsymbol{\varphi})]^{rr} \end{Bmatrix}$$
$$\times\left\{\begin{array}{cc} \mathbf{0} & \mathcal{M}^{H\prime}_{dr}(\boldsymbol{\phi};\boldsymbol{\varphi})[\mathcal{M}^H_{dd}(\boldsymbol{\phi};\boldsymbol{\varphi})]^{-1}\boldsymbol{\Delta}^{-1}(\psi_{ic\infty}) \end{array}\right\}.$$

As a result, the first row/column of the inverse of this augmented expected Hessian matrix $\mathcal{A}_{\boldsymbol{\phi}\boldsymbol{\phi}}$, $\vec{\mathcal{A}}^{\boldsymbol{\phi}\boldsymbol{\phi}}$ say, will be equal to

$$\begin{pmatrix} \vec{\mathcal{A}}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \vec{\mathcal{A}}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c\prime} & \vec{\mathcal{A}}^{\boldsymbol{\psi}_{\boldsymbol{\varrho}}\boldsymbol{\psi}_c\prime} \end{pmatrix} = \begin{pmatrix} \mathcal{A}^{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{A}^{\boldsymbol{\psi}_i\boldsymbol{\psi}_c\prime} & \mathbf{0} \end{pmatrix},$$

which coincides with the inverse of $\mathcal{A}_{\boldsymbol{\psi}\boldsymbol{\psi}}$ plus some 0's. But since the asymptotic variance of the pseudo ML estimators is given by $\mathcal{C}_{\boldsymbol{\phi}\boldsymbol{\phi}} = \mathcal{A}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}\mathcal{B}_{\boldsymbol{\phi}\boldsymbol{\phi}}\mathcal{A}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}$ and the $\mathcal{B}_{\boldsymbol{\psi}\boldsymbol{\psi}}$ block is unchanged, the expression for the asymptotic variance of the pseudo ML estimator of $\boldsymbol{\psi}_c$ in (D96) remains valid, except that it will be evaluated at $\boldsymbol{\varrho}_\infty$.

As for our consistent estimators, Proposition 12 implies that the additional terms of the expected Jacobian are 0 while

$$\mathcal{B}_{\boldsymbol{\varrho}\bar{\boldsymbol{\psi}}_i} = \boldsymbol{\Delta}^{-1}(\psi_{ic0})\mathcal{M}^O_{r\bar{d}}(\boldsymbol{\phi};\boldsymbol{\varphi}). \tag{B60}$$

If we follow the same steps as before we get that

$$\begin{pmatrix} \mathcal{A}_{\boldsymbol{\phi}\boldsymbol{\phi}} & \mathbf{0} \\ \mathcal{A}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\phi}} & \mathcal{A}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i} \end{pmatrix}^{-1} = \begin{pmatrix} \mathcal{A}^{-1}_{\boldsymbol{\phi}\boldsymbol{\phi}} & \mathbf{0} \\ \mathcal{A}^{\bar{\boldsymbol{\psi}}_i\boldsymbol{\phi}} & \mathcal{A}^{-1}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i} \end{pmatrix},$$

where the different components of $\mathcal{A}^{-1}_{\phi\phi}$ can be found above. Similarly, the $\bar{\psi}_i\phi$ block will be given by

$$\mathcal{A}^{\bar{\psi}_i\phi} = -\mathcal{A}^{-1}_{\bar{\psi}_i\bar{\psi}_i}\left(\begin{array}{ccc} \mathcal{A}_{\bar{\psi}_i\psi_c} & \mathbf{0} & \mathbf{0}\end{array}\right)\mathcal{A}^{-1}_{\phi\phi} = -\mathcal{A}^{-1}_{\bar{\psi}_i\bar{\psi}_i}\left(\begin{array}{ccc} \mathcal{A}_{\bar{\psi}_i\psi_c} & \mathbf{0} & \mathbf{0}\end{array}\right)\left(\begin{array}{ccc} \mathcal{A}^{\psi_c\psi_c} & \mathcal{A}^{\psi_i\psi_{c'}} & \mathbf{0} \\ \mathcal{A}^{\psi_i\psi_c} & \mathcal{A}^{\psi_i\psi_i} & \mathcal{A}^{\varrho\psi_{i'}} \\ \mathbf{0} & \mathcal{A}^{\varrho\psi_i} & \mathcal{A}^{\varrho\varrho}\end{array}\right)$$
$$= -\mathcal{A}^{-1}_{\bar{\psi}_i\bar{\psi}_i}\left(\begin{array}{cc} \mathcal{A}_{\bar{\psi}_i\psi_c}\mathcal{A}^{\psi_c\psi_c} & \mathcal{A}_{\bar{\psi}_i\psi_c}\mathcal{A}^{\psi_i\psi_{c'}} \quad \mathbf{0}\end{array}\right).$$

because $\vec{\mathcal{A}}^{\psi_\varrho\psi_{c'}}$ is 0. Thus, the asymptotic covariance matrix of $\hat{\boldsymbol{\psi}}_c$ and $\bar{\boldsymbol{\psi}}_i(\hat{\boldsymbol{\psi}}_{cT})$ will be the same whether or not we estimate $\boldsymbol{\varrho}$. □

When the true distribution is symmetric and the researcher imposes this restriction in estimation, the non-Gaussian PLME of ψ_{im} will also be consistent, in which case we only propose to replace ψ_{ic}. At the same time, ψ_{im} effectively becomes part of the consistent parameter set. As a result, the Jacobian of the Gaussian score for ψ_{ic} with respect to ψ_{im} will no longer be 0. In fact, it will coincide with the Jacobian of this score with respect to $\bar{\psi}_{im}$. However, the expected Jacobian continues to be 0, which means that the asymptotic variance of $\psi_{icT}(\hat{\boldsymbol{\psi}}_{cT})$ which appears in the (2,2) element of (24) remains valid. In this context, we can also show that the asymptotic variance of the non-Gaussian PMLE of ψ_{im} will be given by

$$\psi_{ic\infty}(\boldsymbol{\varrho})\frac{\mathcal{M}^O_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi})}{[\mathcal{M}^H_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi})]^2} + \psi^2_{im0}S(\boldsymbol{\phi};\boldsymbol{\varphi})(\mathbf{W}'_{\boldsymbol{\psi}_c}(\boldsymbol{\psi}_{c0};\boldsymbol{\varphi}_0)\mathbf{V}^{-1}_{\boldsymbol{\psi}_c}(\boldsymbol{\psi}_{c0};\boldsymbol{\varphi}_0)\mathbf{W}_{\boldsymbol{\psi}_c}(\boldsymbol{\psi}_{c0};\boldsymbol{\varphi}_0),$$

where we have exploited the fact that $\mathcal{M}^O_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) = \mathcal{M}^H_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) = \phi = 0$ under symmetry. Further, this asymptotic variance will continue to be valid when we simultaneously estimate $\boldsymbol{\varrho}$ because $\mathcal{M}^O_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) = \mathcal{M}^H_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) = 0$ too.

Proposition 7

We are going to exploit the results in Proposition 2 together with the fact that the parametrisations in Appendix A.5 guarantee that $|\boldsymbol{\Omega}^\circ(\boldsymbol{\varpi})| = 1$ and consequently, that $\mathbf{W}_{\boldsymbol{\vartheta}_c}(\boldsymbol{\vartheta}_{c0};\boldsymbol{\varphi}_0) = \mathbf{0}$. The only new elements we need are the Jacobian matrices:

$$\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\vartheta}_c)}{\partial(\mathbf{a}',\mathbf{b}',\boldsymbol{\varpi}'_L,\boldsymbol{\varpi}'_D)} = \left(\begin{array}{cccc} \mathbf{I}_N & \mathbf{I}_N r_{Mt} & \mathbf{0} & \mathbf{0}\end{array}\right),$$
$$\frac{\partial vec[\Sigma^\circ_t(\boldsymbol{\vartheta}_{c0})]}{\partial(\mathbf{a}',\mathbf{b}',\boldsymbol{\varpi}'_L,\boldsymbol{\varpi}'_D)} = \left[\begin{array}{cccc} \mathbf{0} & \mathbf{0} & (\mathbf{I}_N + \mathbf{K}_{NN})(\boldsymbol{\Omega}^\circ_L\boldsymbol{\Omega}^\circ_D \otimes \mathbf{I}_N)\mathbf{S}'_N & (\boldsymbol{\Omega}^\circ_L \otimes \boldsymbol{\Omega}^\circ_L)\mathbf{E}'_N\frac{\partial vecd(\boldsymbol{\Omega}^\circ_D)}{\partial\boldsymbol{\varpi}'_D}\end{array}\right],$$

with $\partial vecd(\boldsymbol{\Omega}^\circ_D)/\partial\boldsymbol{\varpi}'_D$ in (A19).

Given the block diagonality of the Jacobian between the conditional mean parameters $\boldsymbol{\gamma} = (\mathbf{a}',\mathbf{b}')'$ and the conditional variance parameters $\boldsymbol{\varpi}$, it is clear that both $\mathcal{A}_{\boldsymbol{\vartheta}_c\boldsymbol{\vartheta}_c}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ and $\mathcal{B}_{\boldsymbol{\vartheta}_c\boldsymbol{\vartheta}_c}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ will also be block-diagonal, with

$$\mathcal{A}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) = \frac{\mathrm{M}^H_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)}{\vartheta_{i\infty}}\left(\begin{array}{cc} 1 & \mu_M \\ \mu_M & \sigma^2_M + \mu^2_M\end{array}\right)\otimes\boldsymbol{\Omega}^{\circ-1}$$

and

$$\mathcal{B}_{\boldsymbol{\gamma\gamma}}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)=\frac{\mathrm{M}_{ll}^{O}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)}{\vartheta_{i\infty}}\left(\begin{array}{cc}1 & \mu_M\\ \mu_M & \sigma_M^2+\mu_M^2\end{array}\right)\otimes\boldsymbol{\Omega}^{\circ-1},$$

where $\mu_M=E(r_{Mt})$ and $\sigma_M^2=V(r_{Mt})$, so that

$$\mathcal{C}_{\boldsymbol{\gamma\gamma}}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)=\frac{\vartheta_{i\infty}\mathrm{M}_{ll}^{O}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)}{[\mathrm{M}_{ll}^{H}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)]^2}\left(\begin{array}{cc}(1+\mu_M^2/\sigma_M^2) & -\mu_M/\sigma_M^2\\ -\mu_M/\sigma_M^2 & 1/\sigma_M^2\end{array}\right)\otimes\boldsymbol{\Omega}^{\circ}.$$

Since this expression applies to the Gaussian PMLE estimator too, we have that the efficiency ratio for the conditional mean parameters is given by

$$\frac{\mathrm{M}_{ll}^{O}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)}{\left[\mathrm{M}_{ll}^{H}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\right]^2}\cdot\frac{\vartheta_{i\infty}}{\vartheta_{i0}}$$

which agrees with expression (16) in Amengual and Sentana (2010).

Let us now look at the conditional variance parameters. It follows from (B33) and (B35) that both $\mathcal{B}_{\boldsymbol{\varpi\varpi}}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ and $\mathcal{A}_{\boldsymbol{\varpi\varpi}}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ require the computation of the following two terms:

$$\begin{aligned}&\left[\begin{array}{c}\mathbf{S}_N(\boldsymbol{\Omega}_D^{\circ}\boldsymbol{\Omega}_L'\otimes\mathbf{I}_N)(\mathbf{I}_N+\mathbf{K}_{NN})\\ \frac{\partial vecd'(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D}\mathbf{E}_N(\boldsymbol{\Omega}_L'\otimes\boldsymbol{\Omega}_L')\end{array}\right](\boldsymbol{\Omega}^{\circ-1}\otimes\boldsymbol{\Omega}^{\circ-1})\\ &\times\left[\begin{array}{cc}(\mathbf{I}_N+\mathbf{K}_{NN})(\boldsymbol{\Omega}_L\boldsymbol{\Omega}_D^{\circ}\otimes\mathbf{I}_N)\mathbf{S}_N' & (\boldsymbol{\Omega}_L\otimes\boldsymbol{\Omega}_L)\mathbf{E}_N'\frac{\partial vecd(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D'}\end{array}\right]\end{aligned}\tag{B61}$$

and

$$\begin{aligned}&\left[\begin{array}{c}\mathbf{S}_N(\boldsymbol{\Omega}_D^{\circ}\boldsymbol{\Omega}_L'\otimes\mathbf{I}_N)(\mathbf{I}_N+\mathbf{K}_{NN})\\ \frac{\partial vecd'(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D}\mathbf{E}_N(\boldsymbol{\Omega}_L'\otimes\boldsymbol{\Omega}_L')\end{array}\right]vec(\boldsymbol{\Omega}^{\circ-1})vec'(\boldsymbol{\Omega}^{\circ-1})\\ &\times\left[\begin{array}{cc}(\mathbf{I}_N+\mathbf{K}_{NN})(\boldsymbol{\Omega}_L\boldsymbol{\Omega}_D\otimes\mathbf{I}_N)\mathbf{S}_N' & (\boldsymbol{\Omega}_L\otimes\boldsymbol{\Omega}_L)\mathbf{E}_N'\frac{\partial vecd(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D'}\end{array}\right].\end{aligned}\tag{B62}$$

However, the rank-1 matrix (B62) is identically zero. Specifically, $\boldsymbol{\Omega}^{\circ-1}=\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1}$, so

$$\begin{aligned}&vec'(\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1})(\mathbf{I}_N+\mathbf{K}_{NN})(\boldsymbol{\Omega}_L\boldsymbol{\Omega}_D^{\circ}\otimes\mathbf{I}_N)\mathbf{S}_N'\\ &=vec'(\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1})[(\boldsymbol{\Omega}_L\boldsymbol{\Omega}_D^{\circ}\otimes\mathbf{I}_N)+(\mathbf{I}_N\otimes\boldsymbol{\Omega}_L\boldsymbol{\Omega}_D^{\circ})\mathbf{K}_{NN}]\mathbf{S}_N'\\ &=[vec'(\boldsymbol{\Omega}_L^{-1\prime})+vec'(\boldsymbol{\Omega}_L^{-1})\mathbf{K}_{NN}]\mathbf{S}_N'=2vec'(\boldsymbol{\Omega}_L^{-1\prime})\mathbf{S}_N'=2vecl'(\boldsymbol{\Omega}_L^{-1\prime})\end{aligned}$$

by virtue of theorem 6.7 of Magnus (1988). But $\boldsymbol{\Omega}_L^{-1\prime}$ is unit upper triangular so $vecl(\boldsymbol{\Omega}_L^{-1\prime})=\mathbf{0}$.

Similarly, we have

$$vec'(\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1})(\boldsymbol{\Omega}_L\otimes\boldsymbol{\Omega}_L)\mathbf{E}_N'\frac{\partial vecd(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D'}=vec'(\boldsymbol{\Omega}_D^{\circ-1})\mathbf{E}_N'\frac{\partial vecd(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D'}=vecd'(\boldsymbol{\Omega}_D^{\circ-1})\frac{\partial vecd(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D'}$$

by virtue of theorem 7.3 in Magnus (1988). But

$$\begin{aligned}vecd'(\boldsymbol{\Omega}_D^{\circ-1})\frac{\partial vecd(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D'}&=vecd'([\exp(-\varpi_{D1}),\ldots,\exp(-\varpi_{DN-1}),\exp\left(\textstyle\sum_{j=1}^{N-1}\varpi_j\right)]\\ &\times\left[\begin{array}{c}diag[\exp(\varpi_{D1}),\ldots,\exp(\varpi_{DN-1})]\\ -\exp\left(-\sum_{j=1}^{N-1}\varpi_j\right)\ell_{N-1}'\end{array}\right]=\mathbf{0}.\end{aligned}$$

Therefore, the asymptotic covariance matrix of the pseudo ML estimators of $\boldsymbol{\varpi}$ will be given by (B61) times (28).

In turn, the $\boldsymbol{\varpi}_L\boldsymbol{\varpi}_L$ block of (B61) will be proportional to

$$\begin{aligned}
&\mathbf{S}_N(\boldsymbol{\Omega}_D^{\circ}\boldsymbol{\Omega}_L'\otimes\mathbf{I}_N)(\mathbf{I}_N+\mathbf{K}_{NN})(\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1}\otimes\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1})(\mathbf{I}_N+\mathbf{K}_{NN})(\boldsymbol{\Omega}_L\boldsymbol{\Omega}_D^{\circ}\otimes\mathbf{I}_N)\mathbf{S}_N'\\
&=\mathbf{S}_N[(\boldsymbol{\Omega}_D^{\circ}\boldsymbol{\Omega}_L'\otimes\mathbf{I}_N)+\mathbf{K}_{NN}(\mathbf{I}_N\otimes\boldsymbol{\Omega}_D^{\circ}\boldsymbol{\Omega}_L')](\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1}\otimes\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1})\\
&\quad\times[(\boldsymbol{\Omega}_L\boldsymbol{\Omega}_D^{\circ}\otimes\mathbf{I}_N)+(\mathbf{I}_N\otimes\boldsymbol{\Omega}_L\boldsymbol{\Omega}_D^{\circ})\mathbf{K}_{NN}]\mathbf{S}_N'\\
&=\mathbf{S}_N(\boldsymbol{\Omega}_D^{\circ}\boldsymbol{\Omega}_L'\otimes\mathbf{I}_N)(\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1}\otimes\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1})(\boldsymbol{\Omega}_L\boldsymbol{\Omega}_D^{\circ}\otimes\mathbf{I}_N)\mathbf{S}_N'\\
&+\mathbf{S}_N(\boldsymbol{\Omega}_D^{\circ}\boldsymbol{\Omega}_L'\otimes\mathbf{I}_N)(\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1}\otimes\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1})(\mathbf{I}_N\otimes\boldsymbol{\Omega}_L\boldsymbol{\Omega}_D^{\circ})\mathbf{K}_{NN}\mathbf{S}_N'\\
&+\mathbf{S}_N\mathbf{K}_{NN}(\mathbf{I}_N\otimes\boldsymbol{\Omega}_D^{\circ}\boldsymbol{\Omega}_L')(\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1}\otimes\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1})(\boldsymbol{\Omega}_L\boldsymbol{\Omega}_D^{\circ}\otimes\mathbf{I}_N)\mathbf{S}_N'\\
&+\mathbf{S}_N\mathbf{K}_{NN}(\mathbf{I}_N\otimes\boldsymbol{\Omega}_D^{\circ}\boldsymbol{\Omega}_L')(\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1}\otimes\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1})(\mathbf{I}_N\otimes\boldsymbol{\Omega}_L\boldsymbol{\Omega}_D^{\circ})\mathbf{K}_{NN}\mathbf{S}_N'\\
&=\mathbf{S}_N(\boldsymbol{\Omega}_D^{\circ}\otimes\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1})\mathbf{S}_N'+\mathbf{S}_N(\boldsymbol{\Omega}_L^{-1}\otimes\boldsymbol{\Omega}_L^{-1\prime})\mathbf{K}_{NN}\mathbf{S}_N'\\
&+\mathbf{S}_N\mathbf{K}_{NN}(\boldsymbol{\Omega}_L^{-1\prime}\otimes\boldsymbol{\Omega}_L^{-1})\mathbf{S}_N'+\mathbf{S}_N\mathbf{K}_{NN}(\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1}\otimes\boldsymbol{\Omega}_D^{\circ})\mathbf{K}_{NN}\mathbf{S}_N'\\
&=2\mathbf{S}_N\{(\boldsymbol{\Omega}_D^{\circ}\otimes\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1})+(\boldsymbol{\Omega}_L^{-1}\otimes\boldsymbol{\Omega}_L^{-1\prime})\mathbf{K}_{NN}\}\mathbf{S}_N'
\end{aligned}$$

by virtue of theorems 3.1 and 3.5 in Magnus (1988). Note that premultiplying by $\mathbf{S}_N$ effectively selects the rows corresponding to the elements in the strict lower triangle of $\boldsymbol{\Omega}_L$ while postmultiplying by $\mathbf{S}_N'$ does the same for the columns. But since $\boldsymbol{\Omega}_D^{\circ}\otimes\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1}$ is a block diagonal matrix with blocks $\omega_{jj}^{\circ}\boldsymbol{\Omega}^{\circ-1}$ and $(\boldsymbol{\Omega}_L^{-1}\otimes\boldsymbol{\Omega}_L^{-1\prime})\mathbf{K}_{NN}$ is symmetric, we can tediously prove that the matrix above is block diagonal with respect to $\boldsymbol{\varpi}_{L1},\boldsymbol{\varpi}_{L2}\ldots,\boldsymbol{\varpi}_{LN-1}$, which implies that the estimators of the elements in different columns of $\boldsymbol{\Omega}_L$ are asymptotically independent.

Similarly, the $\boldsymbol{\varpi}_D\boldsymbol{\varpi}_D$ block will be proportional to

$$\begin{aligned}
&\frac{\partial vecd'(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D}\mathbf{E}_N(\boldsymbol{\Omega}_L'\otimes\boldsymbol{\Omega}_L')(\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1}\otimes\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1})(\boldsymbol{\Omega}_L\otimes\boldsymbol{\Omega}_L)\mathbf{E}_N'\frac{\partial vecd(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D'}\\
=\;&\frac{\partial vecd'(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D}\mathbf{E}_N(\boldsymbol{\Omega}_D^{\circ-1}\otimes\boldsymbol{\Omega}_D^{\circ-1})\mathbf{E}_N'\frac{\partial vecd(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D'}=\frac{\partial vecd'(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D}(\boldsymbol{\Omega}_D^{\circ-1}\odot\boldsymbol{\Omega}_D^{\circ-1})\frac{\partial vecd(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D'}
\end{aligned}$$

by virtue of theorem 7.7 of Magnus (1988)), where $\odot$ denotes the element by element Hadamard product. But since $\boldsymbol{\Omega}_D^{\circ}$ is diagonal, $\boldsymbol{\Omega}_D^{\circ-1}\odot\boldsymbol{\Omega}_D^{\circ-1}=\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_D^{\circ-1}$, so

$$\frac{\partial vecd'(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D}(\boldsymbol{\Omega}_D^{\circ-1}\odot\boldsymbol{\Omega}_D^{\circ-1})\frac{\partial vecd(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D'}=\begin{pmatrix}\mathbf{I}_{N-1} & -\ell_{N-1}\end{pmatrix}\begin{pmatrix}\mathbf{I}_{N-1}\\ -\ell_{N-1}'\end{pmatrix}=\left(\mathbf{I}_{N-1}+\ell_{N-1}\ell_{N-1}'\right)$$

because

$$\begin{aligned}
&diag[\exp(-\varpi_{D1}),\ldots,\exp(-\varpi_{DN-1}),\exp\left(\textstyle\sum_{j=1}^{N-1}\varpi_j\right)]\begin{bmatrix}diag[\exp(\varpi_{D1}),\ldots,\exp(\varpi_{DN-1})]\\ -\exp\left(-\sum_{j=1}^{N-1}\varpi_j\right)\ell_{N-1}'\end{bmatrix}\\
&=\begin{pmatrix}\mathbf{I}_{N-1}\\ -\ell_{N-1}'\end{pmatrix}.
\end{aligned}$$

Finally, the $\boldsymbol{\varpi}_L\boldsymbol{\varpi}_D$ block will be proportional to

$$\begin{aligned}
& \mathbf{S}_N(\boldsymbol{\Omega}_D^{\circ}\boldsymbol{\Omega}_L'\otimes\mathbf{I}_N)(\mathbf{I}_N+\mathbf{K}_{NN})(\boldsymbol{\Omega}^{\circ-1}\otimes\boldsymbol{\Omega}^{\circ-1})(\boldsymbol{\Omega}_L\otimes\boldsymbol{\Omega}_L)\mathbf{E}_N'\frac{\partial vecd(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D'} \\
& \mathbf{S}_N[(\boldsymbol{\Omega}_D^{\circ}\boldsymbol{\Omega}_L'\otimes\mathbf{I}_N)+\mathbf{K}_{NN}(\mathbf{I}_N\otimes\boldsymbol{\Omega}_D^{\circ}\boldsymbol{\Omega}_L')](\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1}\otimes\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\boldsymbol{\Omega}_L^{-1}) \\
& \times(\boldsymbol{\Omega}_L\otimes\boldsymbol{\Omega}_L)\mathbf{E}_N'\frac{\partial vecd(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D'} \\
= \;& \mathbf{S}_N(\mathbf{I}_N\otimes\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1})\mathbf{E}_N'\frac{\partial vecd(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D'}+\mathbf{S}_N\mathbf{K}_{NN}(\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1}\otimes\mathbf{I}_N)\mathbf{E}_N'\frac{\partial vecd(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D'} \\
= \;& 2\mathbf{S}_N(\mathbf{I}_N\otimes\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1})\mathbf{E}_N'\frac{\partial vecd(\boldsymbol{\Omega}_D)}{\partial\boldsymbol{\varpi}_D'}
\end{aligned}$$

by virtue of theorem 7.4 in Magnus (1988). Once again, premultiplying by $\mathbf{S}_N$ selects the rows corresponding to the elements in the strict lower triangle of $\boldsymbol{\Omega}_L$ while postmultiplying by $\mathbf{E}_N'$ does the same for the columns corresponding to its diagonal elements. But since $\boldsymbol{\Omega}_L^{-1\prime}$ is upper triangular and $\boldsymbol{\Omega}_D^{\circ-1}$ diagonal, which in turn implies that $(\mathbf{I}_N\otimes\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1})$ is a block diagonal matrix with identical upper triangular diagonal blocks, it is possible to tediously prove that $\mathbf{S}_N(\mathbf{I}_N\otimes\boldsymbol{\Omega}_L^{-1\prime}\boldsymbol{\Omega}_D^{\circ-1})\mathbf{E}_N'$ will be identically $\mathbf{0}$. As a result, the estimators of $\boldsymbol{\varpi}_L$ and $\boldsymbol{\varpi}_D$ will be asymptotically orthogonal too.

Given the diagonality of the Jacobian matrices, the asymptotic variance of our consistent estimator of ϑ_i will coincide with the asymptotic variance of its Gaussian version, which is given by expression (29) because of (B36) and (B37) coupled with $\mathcal{A}_{\boldsymbol{\vartheta}_c\bar{\vartheta}_i}(\boldsymbol{\vartheta}_0,\mathbf{0};\boldsymbol{\varphi}_0)=\mathcal{B}_{\boldsymbol{\vartheta}_c\bar{\vartheta}_i}(\phi_\infty;\boldsymbol{\varphi}_0)=\mathbf{0}$.

Finally, the estimation of $\boldsymbol{\eta}$ is irrelevant because both the $\mathcal{A}$ and $\mathcal{B}$ matrices are block diagonal between $\boldsymbol{\vartheta}_c=(\mathbf{a}',\mathbf{b}',\boldsymbol{\varpi}_L',\boldsymbol{\varpi}_D')'$ and $(\vartheta_i,\boldsymbol{\eta}')'$ since $\mathbf{W}_{\boldsymbol{\vartheta}_c}(\boldsymbol{\vartheta}_{c0};\boldsymbol{\varphi}_0)=\mathbf{0}$ in this case. □

C The general case of non-spherical pseudo likelihoods

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

Let $f(\boldsymbol{\varepsilon}^*;\boldsymbol{\varrho})$ denote the assumed conditional density of $\boldsymbol{\varepsilon}_t^*$ given I_{t-1} and some shape parameters $\boldsymbol{\varrho}$. Let also $\boldsymbol{\phi}=(\boldsymbol{\theta}',\boldsymbol{\varrho})'$ denote the $p+q$ parameters of interest, which once again we assume variation free. Ignoring initial conditions, the log-likelihood function of a sample of size T for those values of $\boldsymbol{\theta}$ for which $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ has full rank will take the form $L_T(\boldsymbol{\phi})=\sum_{t=1}^T l_t(\boldsymbol{\phi})$, where $l_t(\boldsymbol{\phi})=d_t(\boldsymbol{\theta})+\ln f\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}),\boldsymbol{\varrho}\right]$, $d_t(\boldsymbol{\theta})=\ln|\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})|$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})=\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$, and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})=\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta})$.

The most common choices of square root matrices are the Cholesky decomposition, which leads to a lower triangular matrix for a given ordering of $\mathbf{y}_t$, or the spectral decomposition, which yields a symmetric matrix. The choice of square root matrix is non-trivial because $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ affects the value of the log-likelihood function and its score in multivariate non-spherical contexts. In

what follows, we rely mostly on the Cholesky decomposition because it is much faster to compute than the sprectral one, especially when $\mathbf{\Sigma}_t(\boldsymbol{\theta})$ is time-varying. Nevertheless, we also discuss some modifications required for the spectral decomposition later on.

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

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

But since

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

and

$$\begin{aligned} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} &= -\mathbf{\Sigma}_t^{-1/2}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} - [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{\Sigma}_t^{-1/2}(\boldsymbol{\theta})] \frac{\partial vec[\mathbf{\Sigma}_t^{1/2}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'} \\ &= -\{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \mathbf{Z}_{st}'(\boldsymbol{\theta})\}, \end{aligned} \tag{C63}$$

where

$$\left.\begin{array}{l} \mathbf{Z}_{lt}(\boldsymbol{\theta}) = \partial \boldsymbol{\mu}_t'(\boldsymbol{\theta})/\partial\boldsymbol{\theta} \cdot \mathbf{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}) \\ \mathbf{Z}_{st}(\boldsymbol{\theta}) = \partial vec'[\mathbf{\Sigma}_t^{1/2}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta} \cdot [\mathbf{I}_N \otimes \mathbf{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})] \end{array}\right\}, \tag{C64}$$

it follows that

$$\begin{aligned} \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) &= [\mathbf{Z}_{lt}(\boldsymbol{\theta}), \mathbf{Z}_{st}(\boldsymbol{\theta})] \begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{e}_{st}(\boldsymbol{\phi}) \end{bmatrix} = \mathbf{Z}_{dt}(\boldsymbol{\theta}) \mathbf{e}_{dt}(\boldsymbol{\phi}), \\ \mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi}) &= \partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial\boldsymbol{\varrho} = \mathbf{e}_{rt}(\boldsymbol{\phi}), \end{aligned} \tag{C65}$$

with

$$\mathbf{e}_{dt}(\boldsymbol{\phi}) = \begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{e}_{st}(\boldsymbol{\phi}) \end{bmatrix} = \begin{bmatrix} -\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*, \\ -vec\{\mathbf{I}_N + \partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^* \cdot \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\} \end{bmatrix}. \tag{C66}$$

Similarly, let $\mathbf{h}_t(\boldsymbol{\phi})$ denote the Hessian function $\partial \mathbf{s}_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}' = \partial^2 l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'$. Assuming twice differentiability of the different functions involved, expression (C63) implies that

$$\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} = -\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \mathbf{Z}_{st}'(\boldsymbol{\theta})\} \tag{C67}$$

because

$$d\mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho}) = -d\{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\}. \tag{C68}$$

In turn,

$$\begin{aligned} d\mathbf{e}_{st}(\boldsymbol{\theta}, \boldsymbol{\varrho}) &= -dvec\left[\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \cdot \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right] \\ = -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] d\left\{\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right\} &- \left\{\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right\} d\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \end{aligned} \tag{C69}$$

implies that

$$\frac{\partial \mathbf{e}_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} = \frac{\partial \mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} = -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes \mathbf{I}_N]\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}}\frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} - \left\{\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right\}\frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$$
$$\left\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes \mathbf{I}_N]\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} + \left[\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right]\right\}\{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta})\otimes \mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\}. \quad \text{(C70)}$$

Finally, (C68) and (C69) trivially imply that

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

Using these results, we can easily obtained the required expressions for

$$\begin{aligned}\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\phi}) &= \mathbf{Z}_{lt}(\boldsymbol{\theta})\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} + \mathbf{Z}_{st}(\boldsymbol{\theta})\frac{\partial \mathbf{e}_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'}\\ &\quad + \left[\mathbf{e}_{lt}'(\boldsymbol{\phi})\otimes \mathbf{I}_p\right]\frac{\partial vec[\mathbf{Z}_{lt}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'} + \left[\mathbf{e}_{st}'(\boldsymbol{\phi})\otimes \mathbf{I}_p\right]\frac{\partial vec[\mathbf{Z}_{st}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'}, \quad \text{(C71)}\\ \mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\varrho}t}(\boldsymbol{\phi}) &= \mathbf{Z}_{lt}(\boldsymbol{\theta})\partial \mathbf{e}_{lt}(\boldsymbol{\phi})/\partial \boldsymbol{\varrho}' + \mathbf{Z}_{st}(\boldsymbol{\theta})\partial \mathbf{e}_{st}(\boldsymbol{\phi})/\partial \boldsymbol{\varrho}', \quad \text{(C72)}\\ \mathbf{h}_{\boldsymbol{\varrho}\boldsymbol{\varrho}t}(\boldsymbol{\phi}) &= \partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial \boldsymbol{\varrho}\partial \boldsymbol{\varrho}'.\end{aligned}$$

Importantly, while $\mathbf{Z}_{lt}(\boldsymbol{\theta})$, $\mathbf{Z}_{st}(\boldsymbol{\theta})$, $\partial vec[\mathbf{Z}_{lt}(\boldsymbol{\theta})]/\partial \boldsymbol{\theta}'$ and $\partial vec[\mathbf{Z}_{st}(\boldsymbol{\theta})]/\partial \boldsymbol{\theta}'$ depend on the dynamic model specification, the first and second derivatives of $\ln f(\boldsymbol{\varepsilon}^*;\boldsymbol{\varrho})$ depend on the specific distribution assumed for estimation purposes.

For the standard (i.e. lower triangular) Cholesky decomposition of $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, we will have that

$$dvec(\boldsymbol{\Sigma}_t) = [(\boldsymbol{\Sigma}_t^{1/2}\otimes \mathbf{I}_N) + (\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{1/2})\mathbf{K}_{NN}]dvec(\boldsymbol{\Sigma}_t^{1/2}).$$

Unfortunately, this transformation is singular, which means that we must find an analogous transformation between the corresponding $dvech's$. In this sense, we can write the previous expression as

$$dvech(\boldsymbol{\Sigma}_t) = [\mathbf{L}_N(\boldsymbol{\Sigma}_t^{1/2}\otimes \mathbf{I}_N)\mathbf{L}_N' + \mathbf{L}_N(\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{1/2})\mathbf{K}_{NN}\mathbf{L}_N']dvech(\boldsymbol{\Sigma}_t^{1/2}), \quad \text{(C73)}$$

where $\mathbf{L}_N$ is the elimination matrix (see Magnus, 1988). We can then use the results in chapter 5 of Magnus (1988) to show that the above mapping will be lower triangular of full rank as long as $\boldsymbol{\Sigma}_t^{1/2}$ has full rank, which means that we can readily obtain the Jacobian matrix of $vech(\boldsymbol{\Sigma}_t^{1/2})$ from the Jacobian matrix of $vech(\boldsymbol{\Sigma}_t)$.

In the case of the symmetric square root matrix, the analogous transformation would be

$$dvech(\boldsymbol{\Sigma}_t) = [\mathbf{D}_N^+(\boldsymbol{\Sigma}_t^{1/2}\otimes \mathbf{I}_N)\mathbf{D}_N + \mathbf{D}_N^+(\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{1/2})\mathbf{D}_N]dvech(\boldsymbol{\Sigma}_t^{1/2}),$$

where $\mathbf{D}_N^+ = (\mathbf{D}_N'\mathbf{D}_N)^{-1}\mathbf{D}_N'$ is the Moore-Penrose inverse of the duplication matrix (see Magnus and Neudecker, 1988).

From a numerical point of view, the calculation of both $\mathbf{L}_N(\boldsymbol{\Sigma}_t^{1/2} \otimes \mathbf{I}_N)\mathbf{L}_N'$ and $\mathbf{L}_N(\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{1/2})\mathbf{K}_{NN}\mathbf{L}_N'$ is straightforward. Specifically, given that $\mathbf{L}_N vec(\mathbf{A}) = vech(\mathbf{A})$ for any square matrix $\mathbf{A}$, the effect of premultiplying by the $\frac{1}{2}N(N+1)\times N^2$ matrix $\mathbf{L}_N$ is to eliminate rows N+1, 2N+1 and 2N+2, 3N+1, 3N+2 and 3N+3, etc. Similarly, given that $\mathbf{L}_N\mathbf{K}_{NN}vec(\mathbf{A}) = vech(\mathbf{A}')$, the effect of postmultiplying by $\mathbf{K}_{NN}\mathbf{L}_N'$ is to delete all columns but those in positions 1, N+1, 2N+1,...,N+2, 2N+2,..., N+3, 2N+3,..., N^2.

Let $\mathbf{F}_t$ denote the transpose of the inverse of $\mathbf{L}_N(\boldsymbol{\Sigma}_t^{1/2} \otimes \mathbf{I}_N)\mathbf{L}_N' + \mathbf{L}_N(\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{1/2})\mathbf{K}_{NN}\mathbf{L}_N'$, which will be upper triangular. The fastest way to compute

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

is as follows:

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

2. Then we postmultiply the resulting matrix by $\mathbf{F}_t$

3. Next, we construct the matrix

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

   by eliminating the first row from the second block, the first two rows from the third block, ..., and all the rows but the last one from the last block

4. Finally, we premultiply the resulting matrix by $\partial vech'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta} \cdot \mathbf{F}_t$.

C.2 Asymptotic distribution

C.2.1 Under correct specification

Proposition 10 *If $\boldsymbol{\varepsilon}_t^*|;\boldsymbol{\phi}$ is i.i.d. $D(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\varrho})$ with density $f(\boldsymbol{\varepsilon}^*, \boldsymbol{\varrho})$, then*

$$\begin{aligned} \mathcal{I}_t(\boldsymbol{\phi}) &= \mathbf{Z}_t(\boldsymbol{\theta})\mathcal{M}(\boldsymbol{\varrho})\mathbf{Z}_t'(\boldsymbol{\theta}), \\ \mathbf{Z}_t(\boldsymbol{\theta}) &= \begin{pmatrix} \mathbf{Z}_{dt}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{pmatrix}, \end{aligned}$$

and

$$\mathcal{M}(\boldsymbol{\varrho}) = \begin{bmatrix} \mathcal{M}_{dd}(\boldsymbol{\varrho}) & \mathcal{M}_{dr}(\boldsymbol{\varrho}) \\ \mathcal{M}_{dr}'(\boldsymbol{\varrho}) & \mathcal{M}_{rr}(\boldsymbol{\varrho}) \end{bmatrix} = \begin{bmatrix} \mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{lr}(\boldsymbol{\varrho}) \\ \mathcal{M}_{ls}'(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho}) & \mathcal{M}_{sr}(\boldsymbol{\varrho}) \\ \mathcal{M}_{lr}'(\boldsymbol{\varrho}) & \mathcal{M}_{sr}'(\boldsymbol{\varrho}) & \mathcal{M}_{rr}(\boldsymbol{\varrho}) \end{bmatrix},$$

with

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

and

$$\mathcal{M}_{rr}(\boldsymbol{\varrho}) = V[\mathbf{e}_{rt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = -E\left[\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'|\boldsymbol{\phi}\right].$$

C.2.2 Under misspecification

Proposition 11 *If (12) holds, and $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\varphi}_0$ is i.i.d. $(\mathbf{0},\mathbf{I}_N)$, where $\boldsymbol{\varphi}$ includes $\boldsymbol{\psi}$ and the true shape parameters $\boldsymbol{\rho}$, but the distribution assumed for estimation purposes does not necessarily nest the true density, then the pseudo-true value of the feasible parametric ML estimator of $\boldsymbol{\phi} = (\boldsymbol{\psi}_c', \boldsymbol{\psi}_{im}', \boldsymbol{\psi}_{ic}', \boldsymbol{\varrho})'$, $\boldsymbol{\phi}_\infty$, is such that $\boldsymbol{\psi}_{c\infty}$ is equal to the true value $\boldsymbol{\psi}_{c0}$.*

Proof. We can directly work in terms of the $\boldsymbol{\psi}$ parameters thanks to our assumptions on the mapping $\mathbf{r}_g(.)$. Let us initially keep $\boldsymbol{\varrho}$ fixed to some admissible value. The parametric score vector for the remaining parameters will then be given by (C65), with $\mathbf{Z}_{\boldsymbol{\psi}_{ic}lt}(\boldsymbol{\psi}) = \mathbf{0}$ and $\mathbf{Z}_{\boldsymbol{\psi}_{im}st}(\boldsymbol{\psi}) = \mathbf{0}$.

Since we are systematically working with lower triangular square root decompositions, we can write

$$\begin{aligned}
\mathbf{Z}_{\boldsymbol{\psi}_c st}(\boldsymbol{\psi}) &= \partial vech'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)]/\partial\boldsymbol{\psi}_c\cdot\mathbf{L}_N[\boldsymbol{\Psi}_{ic}^{1/2}\otimes\boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\boldsymbol{\psi}_c)\boldsymbol{\Psi}_{ic}^{-1/2\prime}], \\
\mathbf{Z}_{\boldsymbol{\psi}_{ic}s}(\boldsymbol{\psi}) &= \partial vech'(\boldsymbol{\Psi}_{ic}^{1/2})/\partial\boldsymbol{\psi}_{ic}\cdot\mathbf{L}_N[\mathbf{I}_N\otimes\boldsymbol{\Psi}_{ic}^{-1/2\prime}].
\end{aligned}$$

Given that $\boldsymbol{\Psi}_{ic}^{1/2\prime}$ is upper triangular, $\boldsymbol{\Psi}_{ic}^{-1/2}\boldsymbol{\Sigma}_t^{\diamond -1/2}(\boldsymbol{\psi}_c)$ is lower triangular and $\mathbf{I}_N$ is diagonal, Theorem 5.7.i in Magnus (1988) implies that

$$\begin{aligned}
[\boldsymbol{\Psi}_{ic}^{1/2\prime}\otimes\boldsymbol{\Psi}_{ic}^{-1/2}\boldsymbol{\Sigma}_t^{\diamond -1/2}(\boldsymbol{\psi}_c)]\mathbf{L}_N' &= \mathbf{L}_N'\mathbf{L}_N[\boldsymbol{\Psi}_{ic}^{1/2\prime}\otimes\boldsymbol{\Psi}_{ic}^{-1/2}\boldsymbol{\Sigma}_t^{\diamond -1/2}(\boldsymbol{\psi}_c)]\mathbf{L}_N', \\
(\mathbf{I}_N\otimes\boldsymbol{\Psi}_{ic}^{-1/2})\mathbf{L}_N' &= \mathbf{L}_N'\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Psi}_{ic}^{-1/2})\mathbf{L}_N',
\end{aligned}$$

whence

$$\begin{aligned}
\mathbf{Z}_{\boldsymbol{\psi}_c st}(\boldsymbol{\psi}) &= \frac{\partial vech'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)]}{\partial\boldsymbol{\psi}_c}\mathbf{L}_N[\boldsymbol{\Psi}_{ic}^{1/2}\otimes\boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\boldsymbol{\psi}_c)\boldsymbol{\Psi}_{ic}^{-1/2\prime}]\mathbf{L}_N'\mathbf{L}_N, \\
\mathbf{Z}_{\boldsymbol{\psi}_{ic}s}(\boldsymbol{\psi}) &= \frac{\partial vech'(\boldsymbol{\Psi}_{ic}^{1/2})}{\partial\boldsymbol{\psi}_{ic}}\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Psi}_{ic}^{-1/2\prime})\mathbf{L}_N'\mathbf{L}_N.
\end{aligned}$$

As a result,

$$\begin{aligned}
\mathbf{s}_{\boldsymbol{\psi}_{ic}t}(\boldsymbol{\psi},\boldsymbol{\varrho}) &= -\frac{\partial vech'(\boldsymbol{\Psi}_{ic}^{1/2})}{\partial\boldsymbol{\psi}_{ic}}\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Psi}_{ic}^{-1/2\prime})\mathbf{L}_N' vech\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi})\right\} \\
\mathbf{s}_{\boldsymbol{\psi}_{im}t}(\boldsymbol{\psi},\boldsymbol{\varrho}) &= -\boldsymbol{\Psi}_{ic}^{-1/2\prime}\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}
\end{aligned}$$

and

$$\mathbf{s}_{\boldsymbol{\psi}_c t}(\boldsymbol{\psi},\boldsymbol{\varrho}) = \left\{ \frac{\partial \boldsymbol{\mu}_t^{\diamond\prime}(\boldsymbol{\psi}_c)}{\partial \boldsymbol{\psi}_c} + \frac{\partial vec'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)]}{\partial \boldsymbol{\psi}_c}(\boldsymbol{\psi}_{im} \otimes \mathbf{I}_N) \right\} \boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\boldsymbol{\psi}_c)\mathbf{s}_{\boldsymbol{\psi}_{im} t}(\boldsymbol{\psi},\boldsymbol{\varrho})$$

$$-\frac{\partial vec'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_c)]}{\partial \boldsymbol{\psi}_c} \cdot \mathbf{L}_N[\boldsymbol{\Psi}_{ic}^{1/2} \otimes \boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\boldsymbol{\psi}_c)\boldsymbol{\Psi}_{ic}^{-1/2\prime}]\mathbf{L}_N' vech \left\{ \mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}) \right\}$$

since $vech(\mathbf{A}) = \mathbf{L}_N vec(\mathbf{A})$ for any $N \times N$ square matrix $\mathbf{A}$ regardless of its structure.

Let $\boldsymbol{\psi}_{im\infty}(\boldsymbol{\varrho})$ and $\boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho})$ denote the solution to the implicit system of $N + N(N+1)/2$ equations (B38), which we assume is such that $\boldsymbol{\Psi}_{ic\infty}(\boldsymbol{\varrho})$ is p.d. Given the expression for $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi})$ in (B39), we can immediately see that $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_{c0}, \boldsymbol{\psi}_{im}, \boldsymbol{\psi}_{ic})$ will be $i.i.d.[\boldsymbol{\Psi}_{ic}^{-1/2}(\boldsymbol{\psi}_{im0} - \boldsymbol{\psi}_{im}), \boldsymbol{\Psi}_{ic}^{-1/2}\boldsymbol{\Psi}_{ic0}\boldsymbol{\Psi}_{ic}^{-1/2\prime}]$ conditional on I_{t-1}. This, together with the full rank of $\boldsymbol{\Psi}_{ic}^{-1/2\prime}$ implies that

$$E\left[\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*[\boldsymbol{\psi}_{c0}, \boldsymbol{\psi}_{im\infty}(\boldsymbol{\varrho}), \boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho})];\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \middle| I_{t-1};\boldsymbol{\varphi}_0 \right] = \mathbf{0}.$$

In addition, we know from Theorem 5.6 in Magnus (1988) that the matrix

$$\mathbf{L}_N(\mathbf{I}_N \otimes \boldsymbol{\Psi}_{ic}^{-1/2\prime})\mathbf{L}_N'$$

will be upper triangular of full rank. Similarly, given that we have defined $\boldsymbol{\psi}_{ic} = vech(\boldsymbol{\Psi}_{ic})$, the matrix $\partial vech'(\boldsymbol{\Psi}_{ic}^{1/2})/\partial \boldsymbol{\psi}_{ic}$ would also be of full rank in view of the discussion that follows expression (C73).

As a result, we will also have that

$$vech\left\{ E\left[\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*[\boldsymbol{\psi}_{c0}, \boldsymbol{\psi}_{im\infty}(\boldsymbol{\varrho}), \boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho})];\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \boldsymbol{\varepsilon}_t^{*\prime}[\boldsymbol{\psi}_{c0}, \boldsymbol{\psi}_{im\infty}(\boldsymbol{\varrho}), \boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho})] \middle| I_{t-1};\boldsymbol{\varphi}_0 \right] \right\} = \mathbf{0}.$$

Consequently,

$$E\{\mathbf{s}_{\boldsymbol{\psi} t}[\boldsymbol{\psi}_{c0}, \boldsymbol{\psi}_{im\infty}(\boldsymbol{\varrho}), \boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho}), \boldsymbol{\varrho}] | I_{t-1};\boldsymbol{\varphi}_0\} = \mathbf{0}, \tag{C74}$$

which confirms that $\boldsymbol{\psi}_{c0}$, $\boldsymbol{\psi}_{im\infty}(\boldsymbol{\varrho})$ and $\boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho})$ will be the pseudo-true values corresponding to a restricted PML estimator that keeps $\boldsymbol{\varrho}$ fixed.

If we define $\boldsymbol{\varrho}_\infty$ as the solution to the q equations

$$E\{\mathbf{s}_{\boldsymbol{\varrho} t}[\boldsymbol{\psi}_{c0}, \boldsymbol{\psi}_{im\infty}(\boldsymbol{\varrho}), \boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho}), \boldsymbol{\varrho}] | \boldsymbol{\varphi}_0\} = \mathbf{0},$$

which we assume lies in the interior of the admissible parameter space, then it is clear that $\boldsymbol{\psi}_{c0}$, $\boldsymbol{\psi}_{im\infty} = \boldsymbol{\psi}_{im\infty}(\boldsymbol{\varrho}_\infty)$, $\boldsymbol{\psi}_{ic\infty} = \boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho}_\infty)$ and $\boldsymbol{\varrho}_\infty$ will be the pseudo-true values of the parameters corresponding to an unrestricted PMLE that also estimates $\boldsymbol{\varrho}$. □

If we further assume that the true conditional mean of $\mathbf{y}_t$ is $\mathbf{0}$, and this restriction is imposed in estimation, then $\boldsymbol{\psi}_{im}$ becomes unnecessary, thereby generalising the second part of Theorem 1 in Newey and Steigerwald (1997).

The next result, which extends propositions 2 and 4, contains the ingredients necessary to compute the joint asymptotic covariance matrix of the consistent estimators $\boldsymbol{\psi}_{im}(\hat{\boldsymbol{\psi}}_{cT})$ and $\boldsymbol{\psi}_{ic}(\hat{\boldsymbol{\psi}}_{cT})$ defined in (16) and (17), respectively, and $\hat{\boldsymbol{\phi}}_T$:

Proposition 12 *If (12) holds, and $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\varphi}_0$ is i.i.d. $(\mathbf{0},\mathbf{I}_N)$, where $\boldsymbol{\varphi}$ includes $\boldsymbol{\psi}$ and the true shape parameters, but the distribution assumed for estimation purposes does not necessarily nest the true density, then:*

1.

$$\mathcal{A}=\begin{pmatrix} \mathcal{A}_{\boldsymbol{\phi}\boldsymbol{\phi}} & \mathbf{0} \\ \mathcal{A}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\phi}} & \mathcal{A}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i} \end{pmatrix}=\begin{pmatrix} \mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_{im}} & \mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_{ic}} & \mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\varrho}} & \mathbf{0} & \mathbf{0} \\ \mathcal{A}'_{\boldsymbol{\psi}_c\boldsymbol{\psi}_{im}} & \mathcal{A}_{\boldsymbol{\psi}_{im}\boldsymbol{\psi}_{im}} & \mathcal{A}_{\boldsymbol{\psi}_{im}\boldsymbol{\psi}_{ic}} & \mathcal{A}_{\boldsymbol{\psi}_{ic}\boldsymbol{\varrho}} & \mathbf{0} & \mathbf{0} \\ \mathcal{A}'_{\boldsymbol{\psi}_c\boldsymbol{\psi}_{ic}} & \mathcal{A}'_{\boldsymbol{\psi}_{im}\boldsymbol{\psi}_{ic}} & \mathcal{A}_{\boldsymbol{\psi}_{ic}\boldsymbol{\psi}_{ic}} & \mathcal{A}_{\boldsymbol{\psi}_{im}\boldsymbol{\varrho}} & \mathbf{0} & \mathbf{0} \\ \mathcal{A}'_{\boldsymbol{\psi}_c\boldsymbol{\varrho}} & \mathcal{A}'_{\boldsymbol{\psi}_{im}\boldsymbol{\varrho}} & \mathcal{A}'_{\boldsymbol{\psi}_{ic}\boldsymbol{\varrho}} & \mathcal{A}_{\boldsymbol{\varrho}\boldsymbol{\varrho}} & \mathbf{0} & \mathbf{0} \\ \mathcal{A}_{\bar{\boldsymbol{\psi}}_{im}\boldsymbol{\psi}_c} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathcal{A}_{\bar{\boldsymbol{\psi}}_{im}\bar{\boldsymbol{\psi}}_{im}} & \mathbf{0} \\ \mathcal{A}_{\bar{\boldsymbol{\psi}}_{ic}\boldsymbol{\psi}_c} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathcal{A}_{\bar{\boldsymbol{\psi}}_{ic}\bar{\boldsymbol{\psi}}_{ic}} \end{pmatrix},$$

$$\mathcal{B}=\begin{pmatrix} \mathcal{B}_{\boldsymbol{\phi}\boldsymbol{\phi}} & \mathcal{B}_{\boldsymbol{\phi}\bar{\boldsymbol{\psi}}_i} \\ \mathcal{B}'_{\boldsymbol{\phi}\bar{\boldsymbol{\psi}}_i} & \mathcal{B}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i} \end{pmatrix}=\begin{pmatrix} \mathcal{B}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{B}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_{im}} & \mathcal{B}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_{ic}} & \mathcal{B}_{\boldsymbol{\psi}_c\boldsymbol{\eta}} & \mathcal{B}_{\boldsymbol{\psi}_c\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}_{\boldsymbol{\psi}_c\bar{\boldsymbol{\psi}}_{ic}} \\ \mathcal{B}'_{\boldsymbol{\psi}_c\boldsymbol{\psi}_{im}} & \mathcal{B}_{\boldsymbol{\psi}_{im}\boldsymbol{\psi}_{im}} & \mathcal{B}_{\boldsymbol{\psi}_{im}\boldsymbol{\psi}_{ic}} & \mathcal{B}_{\boldsymbol{\psi}_{im}\boldsymbol{\eta}} & \mathcal{B}_{\boldsymbol{\psi}_{im}\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}_{\boldsymbol{\psi}_{im}\bar{\boldsymbol{\psi}}_{ic}} \\ \mathcal{B}'_{\boldsymbol{\psi}_c\boldsymbol{\psi}_{ic}} & \mathcal{B}'_{\boldsymbol{\psi}_{im}\boldsymbol{\psi}_{ic}} & \mathcal{B}_{\boldsymbol{\psi}_{ic}\boldsymbol{\psi}_{ic}} & \mathcal{B}_{\boldsymbol{\psi}_{ic}\boldsymbol{\eta}} & \mathcal{B}_{\boldsymbol{\psi}_{ic}\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}_{\boldsymbol{\psi}_{ic}\bar{\boldsymbol{\psi}}_{ic}} \\ \mathcal{B}'_{\boldsymbol{\psi}_c\boldsymbol{\eta}} & \mathcal{B}'_{\boldsymbol{\psi}_{im}\boldsymbol{\eta}} & \mathcal{B}'_{\boldsymbol{\psi}_{ic}\boldsymbol{\eta}} & \mathcal{B}_{\boldsymbol{\eta}\boldsymbol{\eta}} & \mathcal{B}_{\boldsymbol{\eta}\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}_{\boldsymbol{\eta}\bar{\boldsymbol{\psi}}_{ic}} \\ \mathcal{B}'_{\boldsymbol{\psi}_c\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}'_{\boldsymbol{\psi}_{im}\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}'_{\boldsymbol{\psi}_{ic}\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}'_{\boldsymbol{\eta}\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}_{\bar{\boldsymbol{\psi}}_{im}\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}_{\bar{\boldsymbol{\psi}}_{im}\bar{\boldsymbol{\psi}}_{ic}} \\ \mathcal{B}'_{\boldsymbol{\psi}_c\bar{\boldsymbol{\psi}}_{ic}} & \mathcal{B}'_{\boldsymbol{\psi}_{ic}\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}'_{\boldsymbol{\psi}_{ic}\bar{\boldsymbol{\psi}}_{ic}} & \mathcal{B}'_{\boldsymbol{\eta}\bar{\boldsymbol{\psi}}_{ic}} & \mathcal{B}'_{\bar{\boldsymbol{\psi}}_{im}\bar{\boldsymbol{\psi}}_{ic}} & \mathcal{B}_{\bar{\boldsymbol{\psi}}_{ic}\bar{\boldsymbol{\psi}}_{ic}} \end{pmatrix},$$

with detailed expressions for all the elements in the proof.

2. If in addition (14) holds, then both $\mathcal{A}$ and $\mathcal{B}$ become block diagonal between $\boldsymbol{\psi}_c$ and $(\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic},\boldsymbol{\varrho},\bar{\boldsymbol{\psi}}_{im},\bar{\boldsymbol{\psi}}_{ic})$.

Proof.

To obtain the asymptotic distribution of the unrestricted pseudo ML estimators $\hat{\boldsymbol{\psi}}_T$ and $\hat{\boldsymbol{\varrho}}_T$, we need the asymptotic covariance matrix of the average scores as well as the expected value of the average Hessian matrix evaluated at the pseudo true values $\boldsymbol{\phi}'_\infty=(\boldsymbol{\psi}'_{c0},\boldsymbol{\psi}'_{im\infty},\boldsymbol{\psi}'_{ic\infty},\boldsymbol{\varrho}'_\infty)$. Given that $\mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi}_\infty)$ only depends on $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im\infty},\boldsymbol{\psi}_{ix\infty})$, which is *i.i.d.* over time, it follows that

$$E[\mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0]=\mathbf{0}, \tag{C75}$$

which in conjunction with (B30) proves the martingale difference nature of the spherical score evaluated at the pseudo-true values. As a result, we only need the contemporaneous covariance matrix of the component of the score corresponding to the t^{th} observation, which in turn depends on the contemporaneous covariance matrix of $\mathbf{e}_{dt}(\boldsymbol{\phi}_\infty)$ and $\mathbf{e}_{rt}(\boldsymbol{\phi}_\infty)$. Given the expression for

$\mathbf{e}_{dt}(\boldsymbol{\phi}_\infty)$ in (C66), it immediately follows that

$$E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{lt}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0]=E\left\{\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^{*\prime}}\middle|\boldsymbol{\varphi}_0\right\}=\mathcal{M}_{ll}^O(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0), \quad \text{(C76)}$$

$$E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{st}(\boldsymbol{\phi}_\infty)] = E\left\{\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\right.$$

$$\left.\times vec'\left\{\mathbf{I}_N+\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_\infty)\right\}\middle|\boldsymbol{\varphi}_0\right\}=\mathcal{M}_{ls}^O(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0), \quad \text{(C77)}$$

$$E[\mathbf{e}_{st}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{st}(\boldsymbol{\phi}_\infty)] = E\left\{vec\left\{\mathbf{I}_N+\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_\infty)\right\}\right.$$

$$\left.\times vec'\left\{\mathbf{I}_N+\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi})\right\}\middle|\boldsymbol{\varphi}_0\right\}=\mathcal{M}_{ss}^O(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0). \quad \text{(C78)}$$

Similarly,

$$E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{rt}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0]=E\left\{-\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*\left(\boldsymbol{\psi}_\infty\right);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varrho}'}\middle|\boldsymbol{\varphi}_0\right\}=\mathcal{M}_{lr}^O(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \quad \text{(C79)}$$

$$E[\mathbf{e}_{st}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{rt}(\boldsymbol{\phi}_\infty)] = E\left\{-vec\left\{\mathbf{I}_N+\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\cdot\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*\left(\boldsymbol{\psi}_\infty\right);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varrho}'}\right\}\right.$$

$$\left.\times[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_\infty)\otimes\mathbf{I}_N])\middle|\boldsymbol{\varphi}_0\right\}=\mathcal{M}_{sr}^O(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \quad \text{(C80)}$$

and

$$E[\mathbf{e}_{rt}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{rt}(\boldsymbol{\phi}_\infty)] = E\left\{\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*\left(\boldsymbol{\psi}_\infty\right);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varrho}}\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*\left(\boldsymbol{\psi}_\infty\right);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varrho}'}|\boldsymbol{\varphi}_0\right\}=\mathcal{M}_{rr}^O(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0). \quad \text{(C81)}$$

Hence, we will have that $\mathcal{B}_{\boldsymbol{\phi}\boldsymbol{\phi}}=E[\mathcal{B}_{\boldsymbol{\phi}\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)]$, where

$$\mathcal{B}_{\boldsymbol{\phi}\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)=V[\mathbf{s}_t(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0]=\mathbf{Z}_t(\boldsymbol{\psi}_\infty)\mathcal{M}^O(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}_t(\boldsymbol{\psi}_\infty), \quad \text{(C82)}$$

and $\mathcal{M}^O(\boldsymbol{\phi};\boldsymbol{\varphi})=V[\mathbf{e}_t(\boldsymbol{\phi})|\boldsymbol{\varphi}]$.

Tedious algebra shows that $\mathcal{A}_{\boldsymbol{\phi}\boldsymbol{\phi}}=E[\mathcal{A}_t(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)]$, where

$$\mathcal{A}_t(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)=-E[\mathbf{h}_t(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0]=\mathbf{Z}_t(\boldsymbol{\psi}_\infty)\mathcal{M}^H(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}_t(\boldsymbol{\psi}_\infty), \quad \text{(C83)}$$

and $\mathcal{M}^H(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ contains the following elements

$$\mathcal{M}_{ll}^H(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)=E\left\{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\middle|\boldsymbol{\varphi}_0\right\}, \quad \text{(C84)}$$

$$\mathcal{M}_{ls}^H(\boldsymbol{\phi};\boldsymbol{\varphi})=E\left\{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\cdot[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi})\otimes\mathbf{I}_N])\middle|\boldsymbol{\varphi}\right\}, \quad \text{(C85)}$$

$$\mathcal{M}_{ss}^H(\boldsymbol{\phi};\boldsymbol{\varphi})=E\left\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi})\otimes\mathbf{I}_N]\cdot\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\cdot[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi})\otimes\mathbf{I}_N]|\boldsymbol{\varphi}\right\}-\mathbf{K}_{NN} \quad \text{(C86)}$$

$$\mathcal{M}_{lr}^H(\boldsymbol{\phi};\boldsymbol{\varphi})=-E\left[\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\varphi}\right], \quad \text{(C87)}$$

$$\mathcal{M}_{sr}^H(\boldsymbol{\phi};\boldsymbol{\varphi})=-E\left[[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi})\otimes\mathbf{I}_N]\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\varphi}\right\}, \quad \text{(C88)}$$

and

$$\mathcal{M}_{rr}^H(\boldsymbol{\phi};\boldsymbol{\varphi})=-E\left\{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi});\boldsymbol{\varrho}]/\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'|\boldsymbol{\varphi}\right\}. \quad \text{(C89)}$$

Let us now turn to our consistent estimators of $\boldsymbol{\psi}_{ic}$ and $\boldsymbol{\psi}_{im}$. The fact that the Gaussian pseudo score for these parameters are influence functions that only depend on $\boldsymbol{\psi}_c$ and $\bar{\boldsymbol{\psi}}_i$ trivially implies that

$$\frac{\partial \mathbf{s}_{\boldsymbol{\psi}_i t}(\boldsymbol{\psi}_c, \bar{\boldsymbol{\psi}}_i; \mathbf{0})}{\partial \bar{\boldsymbol{\psi}}_i'} = \mathbf{0} \text{ and } \frac{\partial \mathbf{s}_{\boldsymbol{\psi}_i t}(\boldsymbol{\psi}_c, \bar{\boldsymbol{\psi}}_i; \mathbf{0})}{\partial \boldsymbol{\varrho}'} = \mathbf{0}.$$

For analogous reasons,

$$\frac{\partial \mathbf{s}_{\boldsymbol{\psi}_c t}(\boldsymbol{\psi}_c, \boldsymbol{\psi}_i, \boldsymbol{\varrho})}{\partial \bar{\boldsymbol{\psi}}_i'} = \mathbf{0}, \ \frac{\partial \mathbf{s}_{\boldsymbol{\psi}_i t}(\boldsymbol{\psi}_c, \boldsymbol{\psi}_i, \boldsymbol{\varrho})}{\partial \bar{\boldsymbol{\psi}}_i'} = \mathbf{0}, \frac{\partial \mathbf{s}_{\boldsymbol{\varrho} t}(\boldsymbol{\psi}_c, \boldsymbol{\psi}_i, \boldsymbol{\varrho})}{\partial \bar{\boldsymbol{\psi}}_i'} = \mathbf{0},$$

We will also have that

$$\frac{\partial \mathbf{s}'_{\boldsymbol{\psi}_i t}(\boldsymbol{\psi}_c, \bar{\boldsymbol{\psi}}_i; \mathbf{0})}{\partial \boldsymbol{\psi}_c} = \mathbf{h}'_{\boldsymbol{\psi}_c \boldsymbol{\psi}_i t}(\boldsymbol{\psi}, \mathbf{0})$$

and

$$\frac{\partial \mathbf{s}'_{\boldsymbol{\psi}_i t}(\boldsymbol{\psi}_c, \bar{\boldsymbol{\psi}}_i; \mathbf{0})}{\partial \bar{\boldsymbol{\psi}}_i} = \mathbf{h}'_{\boldsymbol{\psi}_i \boldsymbol{\psi}_i t}(\boldsymbol{\psi}, \mathbf{0}).$$

But since we are evaluating these expressions at consistent estimators of $\boldsymbol{\psi}$, we will have that $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_0) = \boldsymbol{\varepsilon}_t^*$, whence we can obtain the remaining elements of $\mathcal{A}$. In particular, given that (B39) implies that for a fixed value of $\boldsymbol{\psi}_c$ we could understand the Gaussian log-likelihood function of $\mathbf{y}_t$ as a Gaussian log-likelihood for the pseudo-standardised residuals $\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\psi}_c)$ with mean $\boldsymbol{\psi}_{im}$ and covariance matrix $\boldsymbol{\Psi}_{ic}$, it immediately follows that $\mathcal{A}_{\bar{\boldsymbol{\psi}}_{im} \bar{\boldsymbol{\psi}}_{ic}} = \mathbf{0}$.

Next, we need to find out the asymptotic covariance matrix of the sample averages of $\mathbf{s}_{\boldsymbol{\psi}_{ic} t}(\boldsymbol{\psi}_0; \mathbf{0})$ and $\mathbf{s}_{\boldsymbol{\psi}_{im} t}(\boldsymbol{\psi}_0; \mathbf{0})$, as well as their asymptotic covariances with the sample averages of $\mathbf{s}_{\boldsymbol{\psi} t}(\boldsymbol{\phi}_\infty)$ and $\mathbf{s}_{\boldsymbol{\varrho} t}(\boldsymbol{\phi}_\infty)$, which coincide with contemporaneous variance and covariances of these influence functions because they are martingale difference sequences. In turn, they depend on the covariance matrix of $\mathbf{e}_{dt}(\boldsymbol{\psi}_0, \mathbf{0})$, which is given by (A17), as well as on the covariances of this vector with $\mathbf{e}_{dt}(\boldsymbol{\phi}_\infty)$ and $\mathbf{e}_{rt}(\boldsymbol{\phi}_\infty)$. Specifically, the required additional elements are

$$E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{lt}(\boldsymbol{\psi}_0, \mathbf{0})] = E\left\{ \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty); \boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*} \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_0) \middle| \boldsymbol{\varphi}_0 \right\} = \mathcal{M}_{l\bar{l}}^O(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0), \tag{C90}$$

$$E[\mathbf{e}_{st}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{lt}(\boldsymbol{\psi}_0, \mathbf{0})] = E\left\{ vec\left\{ \mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty); \boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*} \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_\infty) \right\} \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_0) \middle| \boldsymbol{\varphi}_0 \right\} = \mathcal{M}_{s\bar{l}}^O(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0), \tag{C91}$$

$$E[\mathbf{e}_{rt}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{lt}(\boldsymbol{\psi}_0, \mathbf{0})] = E\left\{ \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty); \boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varrho}'} \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_0) | \boldsymbol{\varphi}_0 \right\} = \mathcal{M}_{r\bar{l}}^O(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0), \tag{C92}$$

and

$$E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{st}(\boldsymbol{\psi}_0,\mathbf{0})] = E\left\{\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*} vec'\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_0) - \mathbf{I}_N\right]\middle|\boldsymbol{\varphi}_0\right\} = \mathcal{M}^O_{l\bar{s}}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0), \tag{C93}$$

$$\begin{aligned} E[\mathbf{e}_{st}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{st}(\boldsymbol{\psi}_0,\mathbf{0})] &= E\left\{vec\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_\infty)\right\}\right. \\ &\left.\times vec'\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_0) - \mathbf{I}_N\right]\middle|\boldsymbol{\varphi}\right\} = \mathcal{M}^O_{s\bar{s}}(\boldsymbol{\phi};\boldsymbol{\varphi}), \end{aligned} \tag{C94}$$

$$E[\mathbf{e}_{rt}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{ct}(\boldsymbol{\psi}_0,\mathbf{0})] = E\left\{\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varrho}'} vec'\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_0) - \mathbf{I}_N\right]\middle|\boldsymbol{\varphi}\right\} = \mathcal{M}^O_{r\bar{s}}(\boldsymbol{\phi};\boldsymbol{\varphi}). \tag{C95}$$

Finally, we can tediously show that the conditions for block-diagonality of the expected value of the Hessian and the covariance matrix of the score are that $E[\mathbf{Z}_{\boldsymbol{\psi}_c lt}(\boldsymbol{\psi}_\infty)|\boldsymbol{\varphi}_0]$ and $E[\mathbf{Z}_{\boldsymbol{\psi}_c st}(\boldsymbol{\psi}_\infty)|\boldsymbol{\varphi}_0]$ are both 0. But given that

$$\begin{aligned} \mathbf{Z}_{\psi_c lt}(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic}) &= \left[\partial \boldsymbol{\mu}_t^{\diamond\prime}(\boldsymbol{\psi}_{c0})/\partial \boldsymbol{\psi}_c \cdot \boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\boldsymbol{\psi}_{c0})\right]\boldsymbol{\Psi}_{ic}^{-1/2\prime} \\ &\quad + \left\{\partial vec'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_{c0})]/\partial \boldsymbol{\psi}_c \cdot [\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\boldsymbol{\psi}_{c0})]\right\}(\boldsymbol{\psi}_{im} \otimes \boldsymbol{\Psi}_{ic}^{-1/2\prime}), \\ \mathbf{Z}_{\psi_c st}(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic}) &= \left\{\partial vec'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\psi}_{c0})]/\partial \boldsymbol{\psi}_c \cdot [\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\boldsymbol{\psi}_{c0})]\right\}(\boldsymbol{\Psi}_{ic}^{1/2} \otimes \boldsymbol{\Psi}_{ic}^{-1/2\prime}), \end{aligned}$$

those conditions will be satisfied if (14) holds in view of the full rank of $\boldsymbol{\Psi}_{ic}$. □

D Relationship to Fan, Qi and Xiu (2014)

Fan, Qi and Xiu (2014) considered a special case of model (2) in which τ is set to its true value of 0. This means that there is one parameter less to estimate. We can exploit many of the results in the proof of Proposition 6 to study this model. Somewhat surprisingly, tedious algebraic manipulations show that the first part of this proposition remains valid provided that we set $\psi_{im\infty}(\boldsymbol{\varrho}) = 0$. In other words, the asymptotic covariance matrix of the pseudo ML estimator of $\boldsymbol{\psi}_c$ for fixed value of $\boldsymbol{\varrho}$ will be given by

$$\frac{\mathcal{M}^O_{ss}(\boldsymbol{\varrho};\boldsymbol{\varphi}_0)}{[\mathcal{M}^H_{ss}(\boldsymbol{\varrho};\boldsymbol{\varphi}_0)]^2} \cdot \mathbf{V}^{-1}. \tag{D96}$$

As for our proposed consistent closed-form estimator, we can also show that the asymptotic variance of $\psi_{ic}(\hat{\boldsymbol{\psi}}_{cT})$ will be given by

$$(\varkappa - 1)\psi_{ic0}^2 + 4\psi_{ic0}^2 \frac{\mathcal{M}^O_{ss}(\boldsymbol{\varrho};\boldsymbol{\varphi}_0)}{[\mathcal{M}^H_{ss}(\boldsymbol{\varrho};\boldsymbol{\varphi}_0)]^2}\mathbf{W}'\mathbf{V}^{-1}\mathbf{W}, \tag{D97}$$

which coincides with the (2,2) element of expression (24) with $\psi_{im\infty}(\boldsymbol{\varrho}) = 0$.

Obviously, the same relationship applies to the Gaussian PMLEs, so our estimator of ψ_{ic} will be more efficient than its Gaussian PMLE counterpart when the Pseudo ML estimator of $\boldsymbol{\psi}_c$ will be more efficient than its Gaussian PMLE counterpart.

Finally, we can also prove that the third part of Proposition 6 also holds, so that the only change that simultaneously estimating the shape parameters $\boldsymbol{\varrho}$ implies is that all the expressions must be evaluated at $\boldsymbol{\varrho}_\infty$.

Let us now compare the asymptotic distributions previously obtained for the case of $\boldsymbol{\varrho}$ fixed with the asymptotic distribution reported by Fan, Qi and Xiu (2014) for their estimation method, bearing in mind that the mapping between our notation and theirs is as follows: $\boldsymbol{\psi}_c = \boldsymbol{\gamma}$, $\psi_{ic} = \sigma^2$, $\lambda_\infty(\bar{\boldsymbol{\eta}}) = \eta_f^2$, $\sigma_t^{\diamond 2} = v_t$ and $\varepsilon_t^* = \varepsilon_t$.

A crucial ingredient of their results is the vector

$$\mathbf{k}(\boldsymbol{\psi}) = \begin{pmatrix} \sigma_t^{\diamond -1}(\boldsymbol{\psi}_{c0})\partial\sigma_t^{\diamond}(\boldsymbol{\psi}_{c0})/\partial\boldsymbol{\psi}_c \\ \psi_{ic}^{-1/2} \end{pmatrix} = \begin{pmatrix} .5\sigma_t^{\diamond -2}(\boldsymbol{\psi}_{c0})\partial\sigma_t^{\diamond 2}(\boldsymbol{\psi}_{c0})/\partial\boldsymbol{\psi}_c \\ \psi_{ic}^{-1/2} \end{pmatrix}$$

and the matrix

$$E[\mathbf{k}(\boldsymbol{\psi})\mathbf{k}'(\boldsymbol{\psi})|\boldsymbol{\varphi}] = \begin{pmatrix} \mathbf{V} + \mathbf{W}\mathbf{W}' & \psi_{ic}^{-1/2}\mathbf{W} \\ \psi_{ic}^{-1/2}\mathbf{W}' & \psi_{ic}^{-1} \end{pmatrix},$$

whose inverse is

$$\begin{pmatrix} \mathbf{V}^{-1} & -\psi_{ic}^{1/2}\mathbf{V}^{-1}\mathbf{W} \\ -\psi_{ic}^{1/2}\mathbf{W}'\mathbf{V}^{-1} & \psi_{ic}(1 + \mathbf{W}'\mathbf{V}^{-1}\mathbf{W}) \end{pmatrix}.$$

The difference between this matrix and the corresponding matrix in the proof of Proposition 6 is due to the fact that Fan, Qi and Xiu (2014) are interested in the asymptotic distribution of the estimator of $\psi_{ic}^{1/2}$, which the delta method implies is related to the asymptotic distribution of the estimator of ψ_{ic} through the quantity $-.5\psi_{ic}^{-1/2}$.

Theorem 2 in Fan, Qi and Xiu (2014) states that the asymptotic distribution of their three step estimator of $\boldsymbol{\psi}_c$ is given by $\mathbf{V}^{-1}$ times the following scalar

$$\frac{E[\{h_c[\varepsilon_t^*, \lambda_\infty^{1/2}(\boldsymbol{\varrho})]\}^2|\boldsymbol{\varphi}_0]}{\lambda_\infty(\boldsymbol{\varrho})[E\{h_{ic}[\varepsilon_t^*, \lambda_\infty^{1/2}(\boldsymbol{\varrho})]|\boldsymbol{\varphi}_0\}]^2},$$

where

$$\begin{aligned} h(x,s) &= \ln\left[\frac{1}{s}f\left(\frac{x}{s}\right)\right] = \ln f\left(\frac{x}{s}\right) - \ln s \\ h_1(x,s) &= \frac{\partial h(x,s)}{\partial s} = -\frac{1}{s}\left[1 + \frac{x}{s}\frac{\partial \ln f(x/s)}{\partial \varepsilon}\right] \\ h_2(x,s) &= \frac{\partial^2 h(x,s)}{\partial s \partial s} = \frac{\partial h_c(x,s)}{\partial s} = \frac{1}{s^2}\left\{2\left[\frac{x}{s}\frac{\partial \ln f(x/s)}{\partial \varepsilon} + 1\right] + \frac{x^2}{s^2}\frac{\partial^2 \ln f(x/s)}{\partial \varepsilon \partial \varepsilon} - 1\right\}. \end{aligned}$$

It is then easy to see that

$$E[\{h_c[\varepsilon_t^*, \lambda_\infty^{1/2}(\boldsymbol{\varrho})]\}^2|\boldsymbol{\varphi}_0] = \lambda_\infty^{-1}(\boldsymbol{\varrho})\mathcal{M}_{ss}^O(\boldsymbol{\varrho}, \boldsymbol{\varphi}_0)$$

and

$$E\{h_2[\varepsilon_t^*, \lambda_\infty^{1/2}(\boldsymbol{\varrho})]|\boldsymbol{\varphi}_0\} = \lambda_\infty^{-1}(\boldsymbol{\varrho})\mathcal{M}_{ss}^H(\boldsymbol{\varrho}; \boldsymbol{\varphi}_0)$$

so that the asymptotic variance of their estimator of $\boldsymbol{\psi}_c$ will be

$$\frac{\mathcal{M}_{ss}^{O}(\boldsymbol{\varrho}, \boldsymbol{\varphi}_0)}{[\mathcal{M}_{ss}^{H}(\boldsymbol{\varrho}; \boldsymbol{\varphi}_0)]^2} \cdot \mathbf{V}^{-1},$$

which coincides with (D96).

Similarly, if we re-write (D97) as

$$4\psi_{ic0}^2 \left[\frac{\mathcal{M}_{ss}^{O}(\boldsymbol{\phi}; \boldsymbol{\varphi}_0)}{[\mathcal{M}_{ss}^{H}(\boldsymbol{\phi}; \boldsymbol{\varphi}_0)]^2} (1 + \mathbf{W}'\mathbf{V}^{-1}\mathbf{W}) + \left(\frac{\varkappa - 1}{4} - \frac{\mathcal{M}_{ss}^{O}(\boldsymbol{\phi}; \boldsymbol{\varphi}_0)}{[\mathcal{M}_{ss}^{H}(\boldsymbol{\phi}; \boldsymbol{\varphi}_0)]^2} \right) \right],$$

it is clear that the Fan, Qi and Xiu (2014) estimator of ψ_{ic} also has the same asymptotic variance as our counterpart.

TABLE 1: Monte Carlo simulation results of univariate GARCH-M model.

Parameter		β	γ	δ, ψ_{im}	ϑ_i, ψ_{ic}	α	τ
True value		0.85	0.1	0.05	1.0	0.1	0.05
Normal	G-PML	0.8424	0.0909	0.0518	1.0953	0.0998	0.0508
		(0.0538)	(0.0463)	(0.0473)	(0.6003)	(0.0332)	(0.0433)
	t-PML	0.8424	0.0909	0.0519	1.0951	0.0998	0.0508
		(0.0538)	(0.0465)	(0.0474)	(0.6034)	(0.0333)	(0.0433)
	FS Asy.	0.8424	0.0909	0.0518	1.0950	0.0998	0.0507
		(0.0538)	(0.0465)	(0.0473)	(0.6034)	(0.0332)	(0.0433)
	FS Sym.	0.8424	0.0909	0.0519	1.0950	0.0998	0.0508
		(0.0538)	(0.0465)	(0.0474)	(0.6034)	(0.0332)	(0.0433)
Student t_{10}	G-PML	0.8418	0.0911	0.0513	1.0938	0.1002	0.0500
		(0.0593)	(0.0516)	(0.0477)	(0.6699)	(0.0382)	(0.0429)
	t-PML	0.8426	0.0913	0.0513	1.0892	0.1000	0.0505
		(0.0557)	(0.0480)	(0.0458)	(0.6152)	(0.0361)	(0.0417)
	FS Asy.	0.8426	0.0913	0.0512	1.0888	0.0999	0.0500
		(0.0557)	(0.0480)	(0.0471)	(0.6163)	(0.0361)	(0.0427)
	FS Sym.	0.8426	0.0913	0.0513	1.0889	0.0999	0.0504
		(0.0557)	(0.0480)	(0.0458)	(0.6163)	(0.0361)	(0.0417)
GC(0,3.2)	G-PML	0.8411	0.0911	0.0518	1.0933	0.0997	0.0509
		(0.0641)	(0.0527)	(0.0474)	(0.6727)	(0.0436)	(0.0434)
	t-PML	0.8442	0.0934	0.0511	**1.4512**	**0.1355**	**0.0433**
		(0.0548)	(0.0463)	(0.0366)	(0.8458)	(0.0607)	(0.0290)
	FS Asy.			0.0511	1.0623	0.0995	0.0507
				(0.0459)	(0.5632)	(0.0385)	(0.0432)
	FS Sym.				1.0626	0.0995	0.0509
					(0.5640)	(0.0385)	(0.0334)
GC(-.9,3.2)	G-PML	0.8419	0.0913	0.0528	1.0929	0.1003	0.0522
		(0.0644)	(0.0524)	(0.0480)	(0.6844)	(0.0441)	(0.0442)
	t-PML	0.8442	0.0940	**0.1748**	**1.6489**	**0.1531**	**0.1366**
		(0.0531)	(0.0452)	(0.0560)	(1.0232)	(0.0727)	(0.0340)
	FS Asy.			0.0518	1.0602	0.0996	0.0516
				(0.0460)	(0.5650)	(0.0363)	(0.0441)
	FS Sym.				**1.0754**	**0.1010**	**0.1698**
					(0.5732)	(0.0368)	(0.0335)

Monte Carlo medians and (interquartile ranges) of Gaussian PMLE (G-PML), Student t-based PMLE (t-PML), our consistent estimator of mean and scale parameters (FS Asy.) and our consistent estimator of the overall scale parameter (FS Sym.). Bold figures refer to inconsistent estimators. Random draws of innovations are standard normal (Normal), standardised Student t with ν degrees of freedom (Student t_ν), and standardised fourth-order Gram-Charlier expansion with skewness equal to c_3 and excess kurtosis equal to c_4 (GC(c_3,c_4)). Sample length=1000. Replications=10,000.

TABLE 2: Monte Carlo simulation results of multivariate market model.

	Parameter	**a**	**b**	ϑ_i	ω_{ii}°	ω_{ij}°	ω_{ii}	ω_{ij}
	True value	0.112	1.0	2.7602	1.1362	0.3408	3.136	0.9408
	G-PML	0.1144	0.9997	2.7302	1.1407	0.3422	3.1195	0.9352
		(0.1053)	(0.1200)	(0.0998)	(0.0870)	(0.0714)	(0.2616)	(0.1994)
	t-PML(t)	0.1145	0.9997	2.7302	1.1406	0.3424	3.1198	0.9350
Normal		(0.1056)	(0.1201)	(0.0999)	(0.0869)	(0.0714)	(0.2612)	(0.1993)
	FS Asy.	0.1145		2.7302	1.1407	0.3422	3.1195	0.9352
		(0.1052)		(0.0998)	(0.0870)	(0.0714)	(0.2616)	(0.1994)
	FS Sym.			2.7303			3.1198	0.9350
				(0.0998)			(0.2609)	(0.1993)
	G-PML	0.1158	0.9950	2.7190	1.1422	0.3413	3.1107	0.9278
		(0.1058)	(0.1189)	(0.1651)	(0.1049)	(0.0840)	(0.3420)	(0.2351)
	t-PML	0.1149	0.9959	2.7273	1.1412	0.3417	3.1120	0.9303
Student t_8		(0.0969)	(0.1121)	(0.1682)	(0.0903)	(0.0736)	(0.3123)	(0.2107)
	FS Asy.	0.1159		2.7200	1.1423	0.3414	3.1114	0.9280
		(0.1058)		(0.1655)	(0.1048)	(0.0840)	(0.3425)	(0.2351)
	FS Sym.			2.7247			3.1081	0.9302
				(0.1648)			(0.3152)	(0.2091)
	G-PML	0.1150	1.0017	2.7064	1.1454	0.3436	3.1060	0.9271
		(0.1051)	(0.1207)	(0.2559)	(0.1418)	(0.1168)	(0.4889)	(0.3240)
	t-PML	0.1134	1.0001	**3.1655**	1.1413	0.3411	**3.6249**	**1.0848**
DSMN(0.2,0.1)		(0.0766)	(0.0851)	(0.6102)	(0.0960)	(0.0786)	(0.7718)	(0.3298)
	FS Asy.	0.1153		2.7099	1.1456	0.3436	3.1105	0.9282
		(0.1053)		(0.2562)	(0.1420)	(0.1167)	(0.4885)	(0.3244)
	FS Sym.			2.7381			3.1190	0.9320
				(0.2592)			(0.4037)	(0.2336)
	G-PML	0.1134	1.0015	2.6770	1.1309	0.3171	3.0191	0.8433
		(0.1067)	(0.1212)	(0.2758)	(0.1555)	(0.1494)	(0.5848)	(0.4558)
	t-PML	**0.3927**	1.0014	**2.6784**	**1.0394**	**0.1395**	**2.7895**	**0.3742**
Asymmetric		(0.1153)	(0.1001)	(0.2521)	(0.1695)	(0.1117)	(0.5015)	(0.3059)
Student $t_{8,-1000}$	FS Asy.	0.1136		2.6795	1.1309	0.3175	3.0212	0.8449
		(0.1075)		(0.2763)	(0.1560)	(0.1493)	(0.5854)	(0.4560)
	FS Sym.			**2.8446**			**2.9740**	**0.3996**
				(0.4007)			(0.6082)	(0.3280)

Monte Carlo medians and (interquartile ranges) of Gaussian PMLE (G-PML), Student t-based PMLE (t-PML), our consistent estimator of mean and scale parameters (FS Asy.) and our consistent estimator of the global scale parameter (FS Sym.). Bold figures refer to inconsistent estimators. Random draws of innovations are multivariate standard normal (Normal), multivariate standardised Student t with ν degrees of freedom (Student t_ν), standardised symmetric discrete scale mixture of two multivariate normals with mixing probability r and variance ratio equal to s (DSMN(r,s)) and standardised asymmetric Student t with ν degrees of freedom and asymmetry parameters equal to $\beta\ell_5$ (Asymmetric Student $t_{\nu,\beta)}$. Sample length=500. Replications=1,000.

FIGURE 1: Relative bias of non Gaussian PMLE of ϑ_i

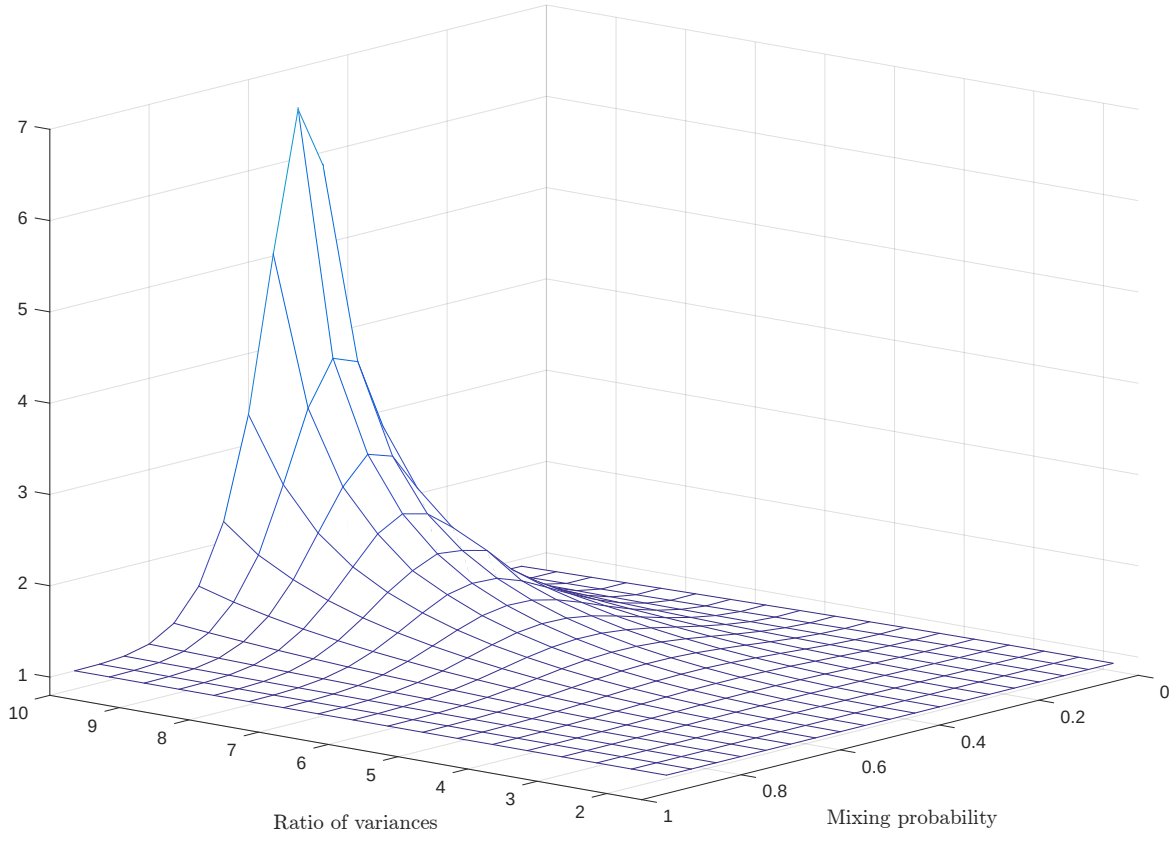


FIGURE 2A: Relative location bias of non Gaussian PMLE of ψ_{im}

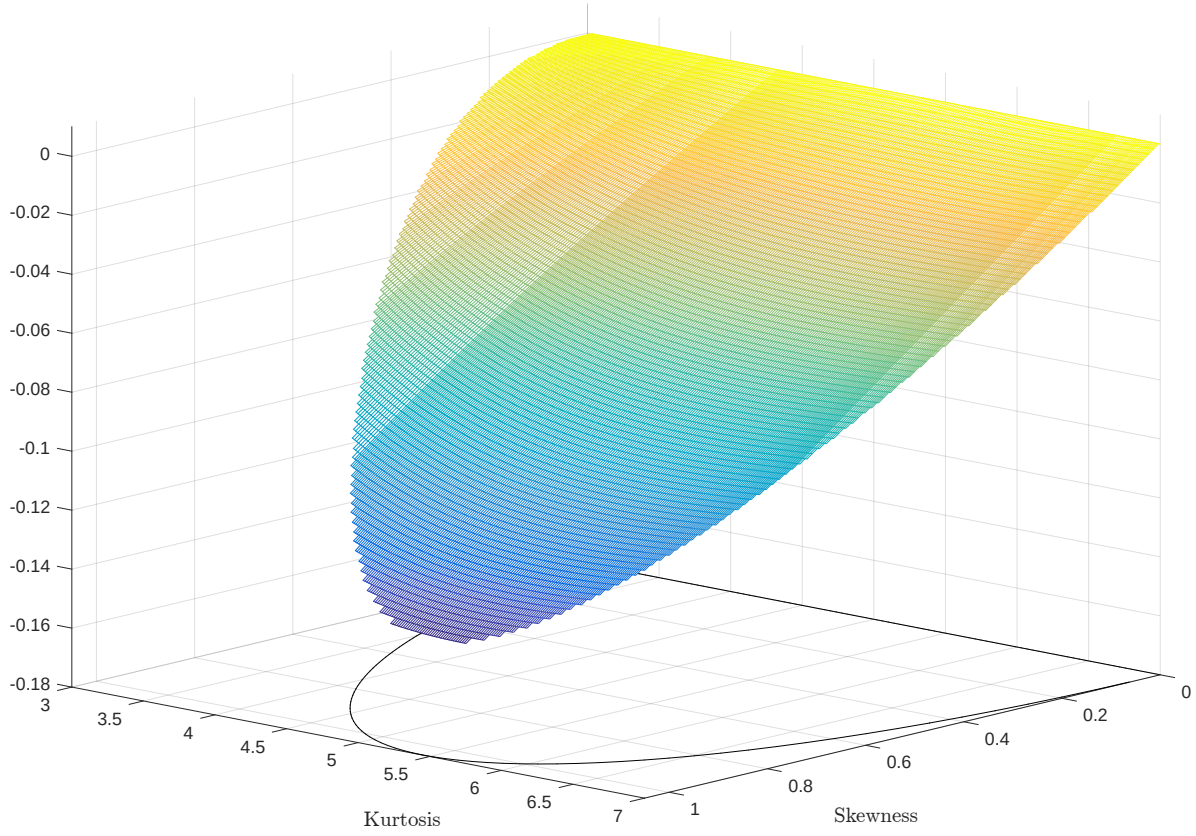


FIGURE 2B: Relative scale bias of non Gaussian PMLE of ψ_{ic}

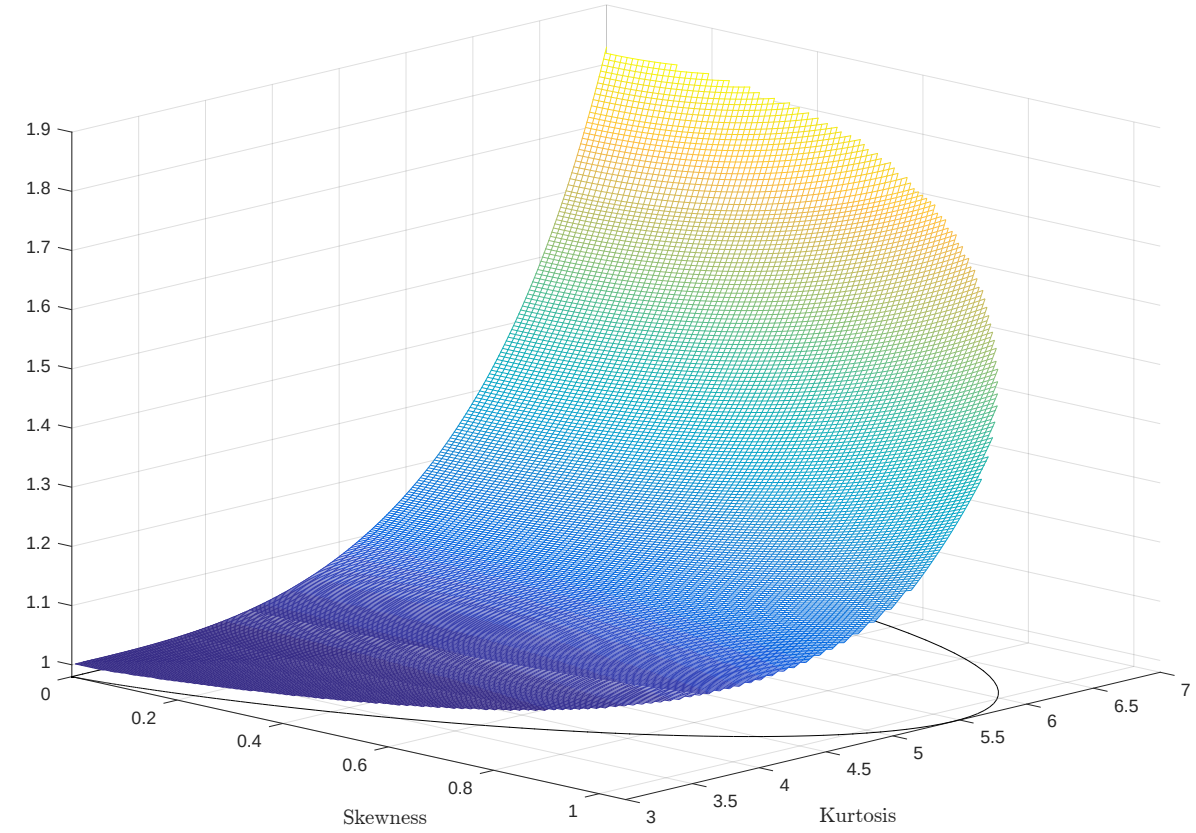


FIGURE 3: Relative efficiency of non Gaussian/Gaussian PMLEs of $\boldsymbol{\psi}_c$

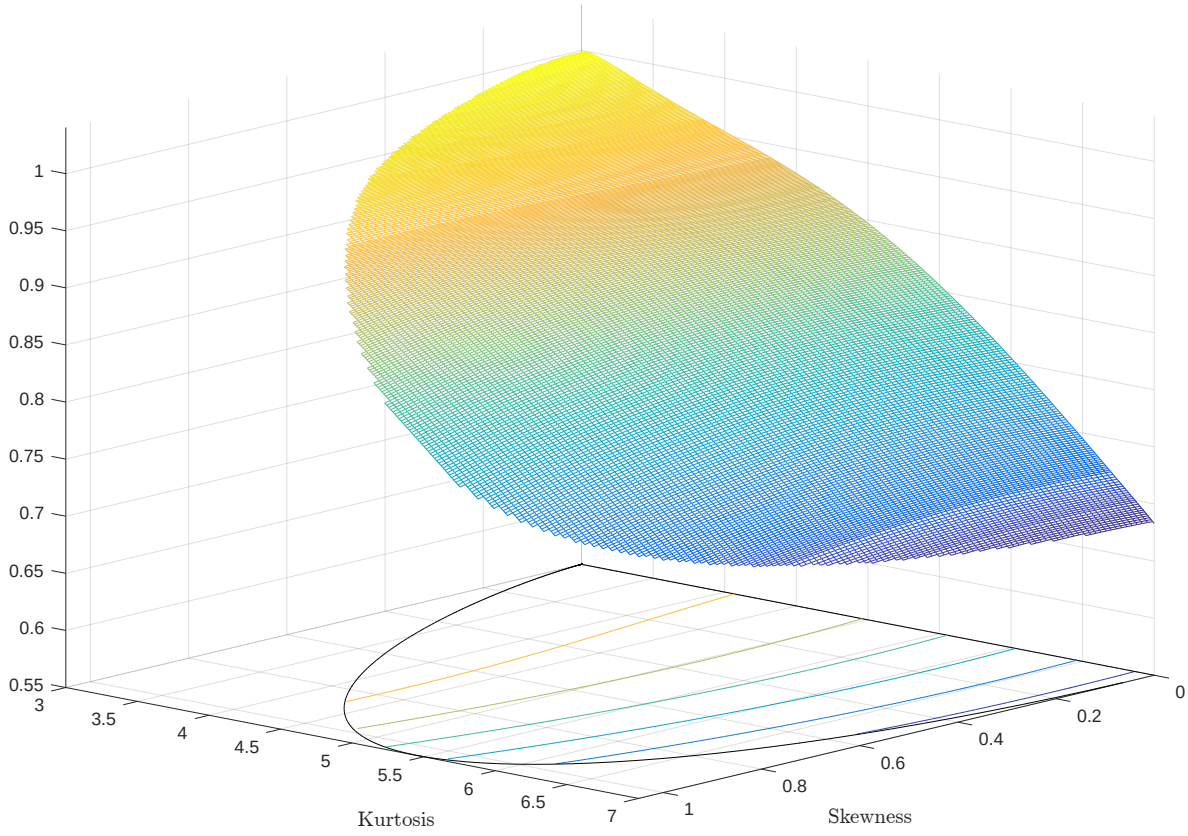


FIGURE 4: Efficiency of Gaussian PMLE and MLE of $\boldsymbol{\psi}_i$ relative to our consistent estimators

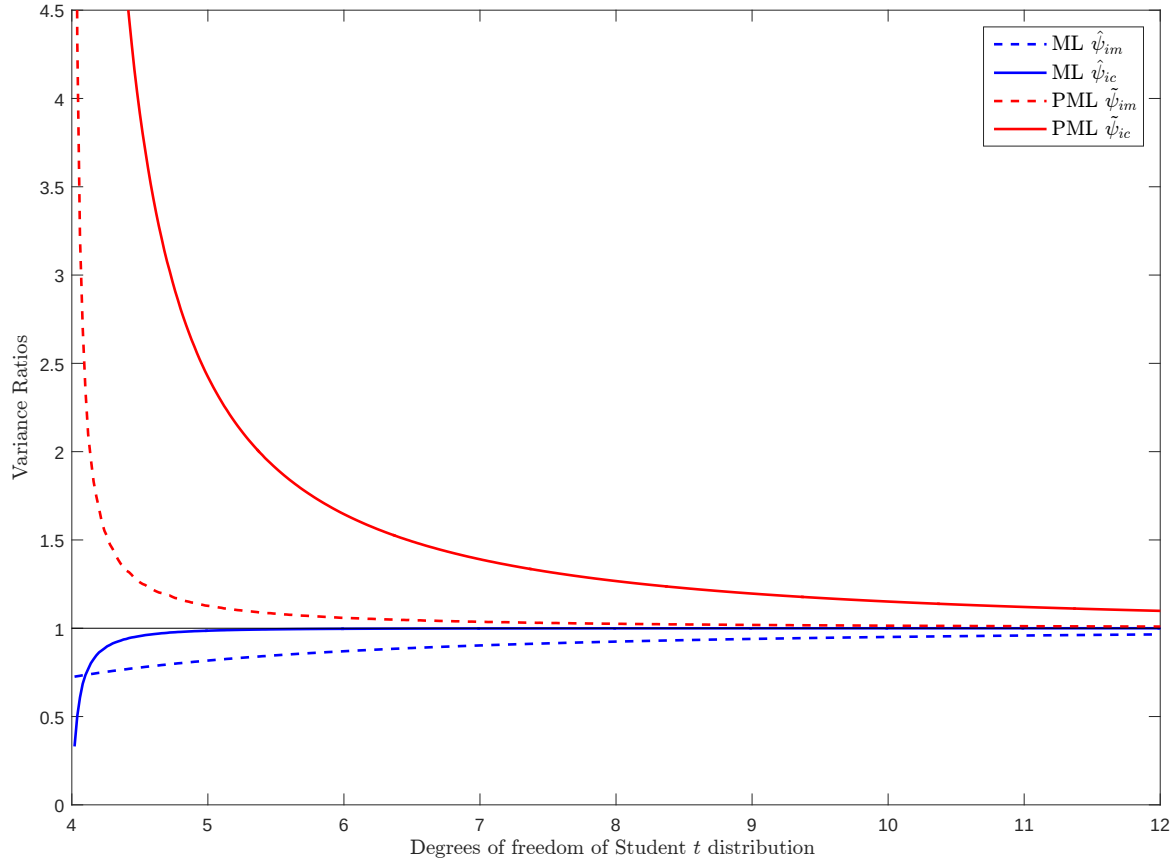


FIGURE 5A: Relative efficiency of non Gaussian/Gaussian PMLEs of **a** and **b**

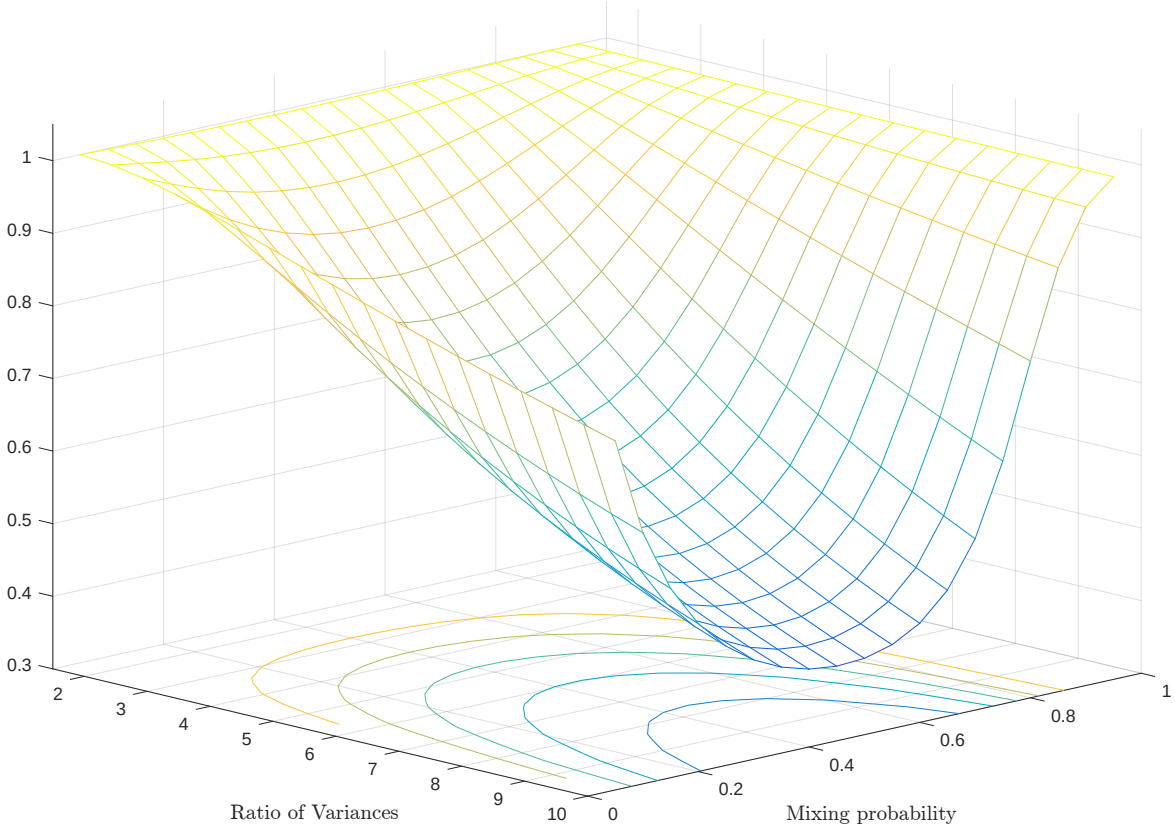


FIGURE 5B: Relative efficiency of non Gaussian/Gaussian PMLEs of $\boldsymbol{\varpi}$

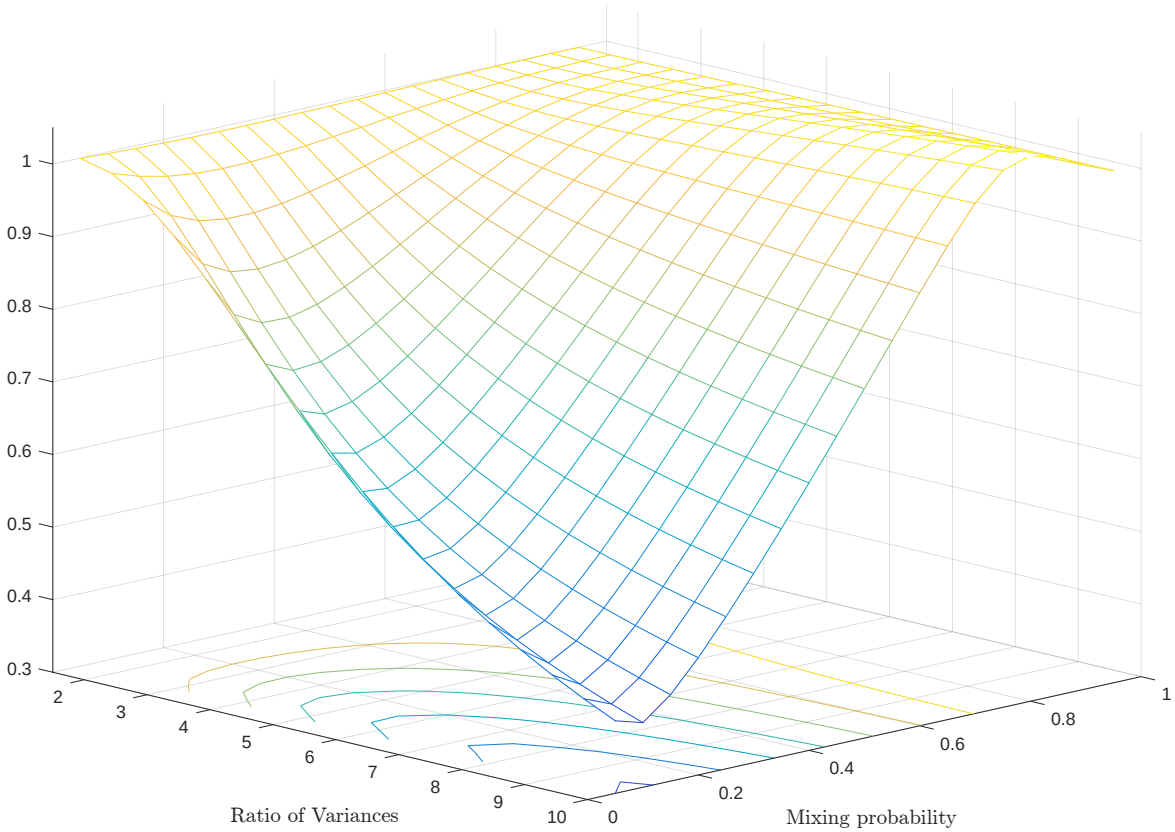


FIGURE 6: Efficiency of our consistent estimator relative to the MLE of ϑ_i