Econometric Theory, **23**, 2007, 251–280. Printed in the United States of America.
DOI: 10.1017/S0266466607070119

SEMIPARAMETRIC MULTIVARIATE VOLATILITY MODELS

CHRISTIAN M. HAFNER
Université Catholique de Louvain

JEROEN V.K. ROMBOUTS
HEC Montreal
and
CIRANO

We consider a model for a multivariate time series where the conditional covariance matrix is a function of a finite-dimensional parameter and the innovation distribution is nonparametric. The semiparametric lower bound for the estimation of the euclidean parameter is characterized, and it is shown that adaptive estimation without reparametrization is not possible. Based on a consistent first-stage estimator (such as quasi maximum likelihood), we propose a semiparametric estimator that estimates the efficient influence function using kernel estimators. We state conditions under which the estimator attains the semiparametric lower bound. For particular models such as the constant conditional correlation model, adaptive estimation of the dynamic part of the model is shown to be possible. To avoid the curse of dimensionality one can, e.g., restrict the multivariate density to the class of spherical distributions, for which we also derive the semiparametric efficiency bound and an estimator that attains this bound. A simulation experiment demonstrates the efficiency gain of the proposed estimator compared with quasi maximum likelihood estimation.

1. INTRODUCTION

Modeling correlations of multivariate financial time series has attracted considerable interest recently, as computing power more and more enables the researcher to model large covariance matrices in flexible ways. For example, two quite general, nonnested classes of models are the so-called VEC GARCH model described by Engle and Kroner (1995) and the dynamic conditional correlation (DCC) model of Engle (2002), including extensions allowing, e.g.,

Rombouts' work was supported by the Centre for Research on e-Finance, HEC Montreal. Hafner gratefully acknowledges financial support by the Fonds Spéciaux de Recherche (FSR 05) of the Université catholique de Louvain. The authors thank three anonymous referees for valuable comments and suggestions and Luc Bauwens, Geert Dhaene, Feico Drost, Wolfgang Härdle, Douglas Hodgson, Jens Peter Kreiss, Oliver Linton, and Bas Werker for helpful discussions. We also thank participants of the CORE econometrics seminar, the York annual meeting in econometrics, the annual econometric study group meeting 2002 in Bristol, the 2003 workshop "The Art of Semiparametrics" in Berlin, and the statistics seminar of the Stockholm School of Economics for valuable comments. Address correspondence to Christian M. Hafner, Institut de statistique and CORE, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium; e-mail: hafner@stat.ucl.ac.be.

© 2007 Cambridge University Press 0266-4666/07 $12.00

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

for asymmetries analogous to the univariate generalized autoregressive conditional heteroskedasticity (GARCH) literature. For a comparison and other models, see the survey by Bauwens, Laurent, and Rombouts (2006).

In this paper we assume that the functional form of the conditional covariance matrix is parametrically specified without restricting attention to specific models such as the DCC model. Our main interest is the estimation of the parameter that characterizes the conditional covariance matrix. Estimation by maximum likelihood is straightforward if one supposes a specific parametric distribution of the innovations. However, the choice of this distribution can be quite problematic. Usually one assumes normality, which provides the so-called quasi maximum likelihood estimator. It is now well known that quasi maximum likelihood estimation (QMLE) is consistent under quite general conditions, even if the true underlying distribution is not normal; see, e.g., Bollerslev and Wooldridge (1992) and Jeantheau (1998). However, in the case of misspecification there may be a substantial efficiency loss of QMLE compared with the correctly specified maximum likelihood estimator (MLE). On the other hand, assuming a nonnormal distribution entails the risk of inconsistent parameter estimation if the distribution is misspecified.

We follow a nonparametric approach by letting the data determine the distribution of the innovations. With the typically large data sets in finance we would expect to obtain density estimates that are sufficiently close to the true distribution of the innovations. As our model consists of a finite-dimensional parameter describing the volatility and correlation dynamics and an infinite-dimensional parameter describing the innovation distribution, it may be called semiparametric (SP) as in Engle and Gonzalez-Rivera (1991), who consider univariate GARCH models.

The SP approach is typically more efficient than QMLE. The case where SP estimators are asymptotically as efficient as MLEs where the true distribution is known is usually referred to as adaptivity. For example, in a univariate framework, adaptive estimation of autoregressive moving average (ARMA) models has been treated by Kreiss (1987). Often, parameters describing the autoregressive dynamics of a model can be estimated adaptively, whereas scale parameters cannot. For univariate GARCH models, Linton (1993) and Drost and Klaassen (1997) show that adaptive estimators of the autoregressive parameters can be constructed by reparameterizing the volatility process. As we want the model to be sufficiently general, the best we can do is to construct estimators that achieve the SP lower bound. We show that the SP lower bound is in general different from the parametric lower bound, so that adaptive estimation of the parameter vector is not possible. However, for particular models such as the constant conditional correlation (CCC) model of Bollerslev (1990), adaptive estimation of the part of the parameter vector that describes the dynamics of volatilities is possible.

To facilitate the applicability of the model, one can, e.g., assume a priori that the innovation distribution belongs to the class of spherical distributions. This

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

bears the important advantage that a nonparametric estimator of the innovation distribution can be constructed in such a way that it has the univariate convergence rate. Hence, there is no "curse of dimensionality" and the proposed procedure can be applied to high-dimensional systems, if sufficient structure is put on the multivariate volatility model to keep the number of parameters under control. Of course, restricting the class of distributions increases the lower bound of the SP estimator, but it will remain more efficient than QMLE. The idea of considering efficiency within the class of spherical (or elliptical) distributions has also been used in Brown and Hodgson (2004) in the context of simultaneous dynamic nonlinear equations. In a simulation study using a multivariate t distribution, we show that there are substantial efficiency gains of SP estimation over QMLE.

The paper is organized as follows. First, the model framework and the traditional estimation method are introduced. The third section discusses the SP lower bounds and shows the properties of SP estimators. In the fourth section a simulation study is provided, and Section 5 concludes. Appendix A contains index set definitions that are used in the proofs of Lemmas 1 and 2. All lemmas are collected in Appendix B, and the proofs of the propositions are in Appendix C.

2. THE MODEL AND ASSUMPTIONS

Consider a vector stochastic process $\{\varepsilon_t, t \in \mathbb{N}, \varepsilon_t \in \mathbb{R}^N\}$ and assume that ε_t has the properties of a conditionally heteroskedastic error term, i.e., mean zero and serially uncorrelated. Our basic model can be written as

$$\varepsilon_t = H_t(\theta) v_t, \tag{1}$$

where $H_t(\theta)$ is an $N \times N$ matrix of full rank depending on a finite-dimensional parameter $\theta \in \Theta \subset R^K$ and where $\{v_t\}$ is independent and identically distributed (i.i.d.) with $\mathrm{E}[v_t] = 0$ and $\mathrm{E}[v_t v_t'] = I_N$. Let $\mathcal{F}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_1)$ and assume that $H_t(\theta)$ is $\mathcal{F}_{t-1}$-measurable. The model implies that the covariance matrix of ε_t conditional on $\mathcal{F}_{t-1}$ is given by $\Sigma_t(\theta) = H_t(\theta) H_t(\theta)'$, which is symmetric. It is of full rank because $H_t(\theta)$ is of full rank. This also implies that $\Sigma_t(\theta)$ is positive definite. Occasionally, we will suppress the dependence of H_t and Σ_t on θ for notational convenience.

Multivariate volatility models either specify H_t directly or, more commonly, they specify Σ_t and say how H_t is defined as a function of Σ_t. Often, H_t will be assumed to be symmetric, in which case a definition of the matrix square root can be employed. For example, using the decomposition $\Sigma_t = \Gamma_t \Lambda_t \Gamma_t'$ where the columns of Γ_t contain the eigenvectors of Σ_t and Λ_t is diagonal with the (positive) eigenvalues on its diagonal, then H_t could be defined as

$$H_t = \Gamma_t \Lambda_t^{1/2} \Gamma_t'. \tag{2}$$

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

Note that H_t in (2) is symmetric and positive definite. In general, however, we impose neither symmetry nor positive definiteness. The benefit of this generality will become obvious later. In the following discussion we make assumptions about the innovation term v_t, where we denote by I_N the identity matrix of dimension N.

Assumption 1. The stochastic error $\{v_t\}$ is an i.i.d. sequence with $\mathrm{E}[v_t] = 0$, $\mathrm{E}[v_t v_t'] = I_N$, and $\mathrm{E}[\|v_t\|^4] < \infty$, where $\|\cdot\|$ denotes the euclidean norm.

Note that Assumption 1 excludes, e.g., a multivariate t distribution with four or less degrees of freedom. The assumption of finite fourth moments of v_t does not restrict ε_t to have finite fourth moments. However, to prove consistency and asymptotic normality of estimators such as the quasi maximum likelihood estimator one typically needs higher moment conditions for ε_t also. Next we make assumptions about the distribution of v_t.

DEFINITION 1. *Let $\mathcal{D}$ denote the class of absolutely continuous N-variate density functions g such that $\forall x \in \mathbb{R}^N : g(x) > 0$, $\int x g(x)\, dx = 0$, $\int xx' g(x)\, dx = I_N$, and $M_{\psi\psi} < \infty$, where $M_{\psi\psi}$ is the Fisher information for scale, i.e., $M_{\psi\psi} = \int \psi(x)\psi(x)' g(x)\, dx < \infty$, and*

$$\psi(x) = -\mathrm{vec}\left(I_N + \frac{\partial \log g(x)}{\partial x}\, x'\right) \tag{3}$$

is the score vector with respect to the scale parameters.

Absolute continuity of g with respect to the Lebesgue measure is defined as the total differential of g being finite almost everywhere on its support.

Assumption 2. Let v_t have density function $g \in \mathcal{D}$.

The assumption of a finite Fisher information for scale is standard in the literature on SP scale models; see, e.g., Linton (1993) and Drost and Klaassen (1997). Note that g is not required to be in a parametric class of densities. Model (1) can now be characterized as a set of probability measures $P = \{P_{\theta g} : \theta \in \Theta, g \in \mathcal{D}\}$. Assume that the true parameter θ_n converges to some fixed $\theta_0 \in int(\Theta)$ and that the true density $g_0 \in \mathcal{D}$. Let $P_0 = P_{\theta_0 g_0}$ and denote by E_0 the expectation under P_0.

Next, we summarize regularity conditions that are used by Jeantheau (1998) and Comte and Lieberman (2003) to show consistency of quasi maximum likelihood estimators.

Assumption 3. We assume the following conditions for all $g(v_t) \in \mathcal{D}$.

(a) Θ is compact.
(b) $\forall \theta_0 \in \Theta$, model (1) admits a unique strictly stationary and ergodic solution with stationary distribution given by P_0. The initial value $H_1(\theta)$ is drawn from P_0.

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

(c) There exists a deterministic constant $c > 0$ such that $\forall t$, $\forall \theta \in \Theta$, $\det \Sigma_t(\theta) \geq c$.
(d) $\forall \theta_0 \in \Theta$, $\mathrm{E}_0[|\log(\det \Sigma_t(\theta_0))|] < \infty$.
(e) The model is identifiable.
(f) $H_t(\theta)$ is a P_0-a.s. continuous function of θ.

Note that because of the generality of our model, the conditions of Assumption 3 are necessarily high level and therefore possibly difficult to check. For the BEKK model[1] proposed by Engle and Kroner (1995), Comte and Lieberman (2003) give simple conditions under which Assumption 3 holds.

Several specifications have been proposed for H_t or Σ_t; see Bauwens et al. (2006) for a survey. Our results concerning SP estimation of θ are sufficiently general to be applicable to any such specification as long as the regularity conditions of Assumption 3 hold. However, in our simulation study of Section 4 we will work with the so-called vector error correction (VEC) representation of a multivariate GARCH(p,q) model, which is given by

$$\mathrm{vech}(\Sigma_t) = \omega + \sum_{i=1}^{q} A_i \,\mathrm{vech}(\varepsilon_{t-i}\varepsilon_{t-i}') + \sum_{j=1}^{p} B_j \,\mathrm{vech}(\Sigma_{t-j}), \tag{4}$$

where A_i and B_i are $N^* \times N^*$ parameter matrices and ω is an N^* parameter vector with $N^* = N(N+1)/2$. If one imposes restrictions on the parameters ω, A_i, and B_j that guarantee positive definite Σ_t, such as the BEKK model of Engle and Kroner (1995), then a sufficient condition for stationarity of ε_t is that the eigenvalues of $\sum_{i=1}^{q} A_i + \sum_{j=1}^{p} B_j$ have modulus smaller than one.

We now turn to the problem of estimating θ. If one supposes that the distribution of v_t is known then maximum likelihood estimation is in principle straightforward. Nevertheless, because the number of parameters is often large, estimation can become a tedious exercise. If v_t is assumed to be normally distributed with zero mean vector and I_N variance matrix then ε_t will be conditionally normally distributed with zero mean vector and Σ_t as covariance matrix. The likelihood, up to an additive constant, for a sample of n observations then takes the form

$$L^{qml}(\theta) = -\frac{1}{2}\sum_{t=1}^{n} (\log\det \Sigma_t(\theta) + \varepsilon_t'\Sigma_t^{-1}(\theta)\varepsilon_t). \tag{5}$$

Defining $l_t^{qml}(\theta) = -\frac{1}{2}(\log\det \Sigma_t(\theta) + \varepsilon_t'\Sigma_t^{-1}(\theta)\varepsilon_t)$, we can write $L^{qml}(\theta) = \sum_{t=1}^{n} l_t^{qml}(\theta)$. Define the quasi maximum likelihood estimator as $\tilde{\theta}_n = \arg\max_{\theta\in\Theta} L^{qml}(\theta)$. As shown by Bollerslev and Wooldridge (1992) in a general conditional heteroskedasticity framework, $\tilde{\theta}_n$ is consistent even if the likelihood is misspecified under fairly general conditions. Therefore this method has been termed quasi maximum likelihood estimation (QMLE). We next assume finiteness of expectations of the Hessian and the outer product of the gradients.

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

Assumption 4. We assume that

(a) $\mathcal{I}(\theta_0) = \mathrm{E}_0[(\partial l_t^{qml}/\partial\theta)(\partial l_t^{qml}/\partial\theta')] < \infty$.
(b) $\mathcal{J}(\theta_0) = -\mathrm{E}_0[(\partial^2 l_t^{qml}/\partial\theta\partial\theta')] < \infty$.
(c) For all i,j,k, $\mathrm{E}_0[\sup_{|\theta-\theta_0|\leq\delta}|(\partial^3 l_t^{qml}(\theta)/\partial\theta_i\partial\theta_j\partial\theta_k)|] < \infty$ for all $\delta > 0$,

where derivatives are evaluated at θ_0.

Under Assumptions 1–4, Comte and Lieberman (2003) prove that the asymptotic distribution of $\tilde{\theta}_n$ is given by

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N(0, V_{qml})$$

with $V_{qml} = \mathcal{J}(\theta_0)^{-1}\mathcal{I}(\theta_0)\mathcal{J}(\theta_0)^{-1}$. In the case of correct specification, i.e., the distribution of v_t is indeed multinormal, $\mathcal{J} = \mathcal{I}$ and $V_{qml} = \mathcal{I}^{-1}$. Comte and Lieberman (2003) assume that the components of v_t for a given t are independent, which seems to be a rather strong assumption. We believe that it should be possible to prove their asymptotic result without this independence assumption but leave it open to debate.

Although the quasi maximum likelihood estimator is consistent, it is inefficient if the likelihood is misspecified. Therefore one sometimes considers the multivariate t distribution as an appropriate choice because of potential fat tails in the innovations. The drawback of this assumption is that if the assumption of a specific nonnormal distribution is not correct, then in general the estimator may not even be consistent; see, e.g., Bollerslev and Wooldridge (1992). Therefore, we will not pursue the assumption of a specific parametric distribution in our paper. In the next section we formalize our motivation for giving all the weight to the data in search of a suitable distribution.

3. SEMIPARAMETRIC ESTIMATION

This section first characterizes the SP lower bound and compares it with the parametric and quasi maximum likelihood lower bounds. Then, construction of efficient or even adaptive estimators is considered. Finally, we discuss the case of imposing more structure on g, namely, assuming that it belongs to the class of spherical distributions.

3.1. The Semiparametric Lower Bound

In the following discussion we characterize the SP lower bound for multivariate volatility models of the form (1). In the context of univariate GARCH models, standard references are Linton (1993), Drost and Klaassen (1997), and Gonzalez-Rivera and Drost (1999). A detailed description of general SP estimation theory and adaptivity is beyond the scope of this paper. We refer to Bickel (1982), Newey (1990), Steigerwald (1992), Bickel et al. (1993), and Drost,

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

Klaassen, and Werker (1997) for details. We follow a standard approach to derive the SP efficiency bound. Alternative methods such as that proposed by Severini and Tripathi (2001) may be possible also.

One can estimate the SP model using a two-step procedure that uses the so-called influence function to correct an initial consistent estimator such as the quasi maximum likelihood estimator. The correction is essentially a one-step Newton–Raphson algorithm based on the score vector of the likelihood. Assume for now that the unknown density g of v_t is parameterized by some nuisance parameter η and write this density as $g(v_t, \eta)$. A particular parameterization of g that contains the true parameterization as a special case is known as the parametric submodel. Let us write the log likelihood as

$$L(\theta, \eta) = -\sum_{t=1}^{n} \log|\det H_t(\theta)| + \sum_{t=1}^{n} \log g(H_t(\theta)^{-1}\varepsilon_t, \eta) = \sum_{t=1}^{n} l_t(\theta, \eta) \tag{6}$$

with

$$l_t(\theta, \eta) = -\log|\det H_t(\theta)| + \log g(H_t(\theta)^{-1}\varepsilon_t, \eta) \tag{7}$$

and denote by $\dot{l}_t(\theta, \eta) = \partial l_t/\partial\theta$ the score vector with respect to the parameter of interest and by $s_t(\theta, \eta) = \partial l_t/\partial\eta$ the score with respect to the nuisance parameter. Note that $s_t(\theta, \eta)$ is a vector of length $\dim(\eta)$, so that its dimension differs among parametric submodels. It is easily seen that $\mathrm{E}_\theta[s_t(\theta, \eta)] = 0$. Also, recall that g is restricted to be density of a mean-zero, identity covariance matrix random vector, so that $s_t(\theta, \eta)$ is orthogonal to the vector $F(v_t)$ where $F(x) = (x', \mathrm{vech}(xx' - I_N)')'$. In the following discussion we suppress the dependence of l_t and $\dot{l}_t$ on η for notational convenience.

To obtain efficient estimates, it is required to eliminate the variation of $\dot{l}_t(\theta)$ that is due to the nuisance parameter η. This is achieved by projecting the score on the so-called tangent set, which is the mean square closure of $\kappa' s_t(\theta, \eta)$, where $\kappa \in \mathbb{R}^{\dim(\eta)\times K}$; see, e.g., Newey (1989, 1990). Because of the properties of $s_t(\theta, \eta)$, the elements of the tangent set are characterized by zero mean and orthogonality to F_t. Thus, the tangent set is defined by

$$\mathcal{T} = \{f: \mathbb{R}^N \to \mathbb{R}^K | \mathrm{E}_g[f(x)] = 0, \mathrm{E}_g[f(x)F'(x)] = 0, \mathrm{E}_g[f(x)f(x)'] < \infty\}.$$

To do the projection on $\mathcal{T}$ it is often crucial to factorize the score $l_t(\theta)$ into a term that depends on θ and $\mathcal{F}_{t-1}$ and another that depends on θ and the nuisance parameter. The next proposition shows that this is possible for the model under study.

PROPOSITION 1. *For the model (1), the score vector takes the form*

$$\dot{l}_t(\theta) = W_t(\theta)\psi(v_t(\theta)), \tag{8}$$

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

where

$$W_t(\theta) = \frac{\partial \operatorname{vec}(H_t)'}{\partial \theta} (I_N \otimes H_t'^{-1}) \tag{9}$$

and $\psi_t = \psi(v_t(\theta))$ *is given by (3).*

Note that W_t is $\mathcal{F}_{t-1}$-measurable and depends only on the specification of $H_t(\theta)$. The other term ψ_t depends on the innovation v_t, so that W_t and ψ_t are independent. Note also that the score vector is a martingale difference sequence. The reason is that $\mathrm{E}_\theta[\dot{l}_t(\theta)|\mathcal{F}_{t-1}] = W_t(\theta)\mathrm{E}_\theta[\psi(v_t(\theta))|\mathcal{F}_{t-1}] = W_t(\theta)\mathrm{E}_\theta[\psi(v_t(\theta))] = 0$ because $\mathrm{E}_\theta[\psi_t] = 0$ as implied by Assumptions 1 and 2.

Remark 1. If one has a model for Σ_t, the score can also be expressed in terms of derivatives of Σ_t rather than H_t. However, it turns out that the score will in general have a different form depending on whether H_t is restricted to be symmetric or not. In the first case ($H_t = H_t'$), W_t takes the form

$$W_t(\theta) = \frac{\partial \operatorname{vec}(\Sigma_t)'}{\partial \theta} (I_N \otimes H_t^2 + H_t \otimes H_t)^{-1} \tag{10}$$

and in the second case

$$W_t(\theta) = \frac{1}{2} \frac{\partial \operatorname{vec}(\Sigma_t)'}{\partial \theta} (H_t'^{-1} \otimes H_t'^{-1}). \tag{11}$$

If g has certain symmetry properties, e.g., sphericity or componentwise independence, then the score in (8) can be equivalently written as $\dot{l}_t(\theta) = W_t(\theta) D_N D_N^+ \psi_t(v_t(\theta))$.[2] In that case, $\dot{l}_t(\theta)$ does not depend on H_t being symmetric or not, because $W_t(\theta)D_N$ is the same whether (10) or (11) is used.

In the following paragraphs we discuss the conditions under which one obtains local asymptotic normality (LAN) of the SP multivariate volatility model. Consider two local parameterizations $\theta_n = \theta_0 + O(n^{-1/2})$ and $\tilde{\theta}_n = \theta_0 + O(n^{-1/2})$ such that $n^{1/2}(\theta_n - \tilde{\theta}_n) \to \lambda$. Suppose the log-likelihood ratio $\Lambda_n = L(\tilde{\theta}_n) - L(\theta_n)$ can be written in the form

$$\Lambda_n = \sum_{t=1}^{n} \{l(I_N + \Xi_{nt}) - l(I_N)\} + R_n,$$

where $\Xi_{nt} = H_t^{-1}(\theta_n) H_t(\tilde{\theta}_n) - I_N$ and where $l(H_t(\theta)) = l_t(\theta, \eta)$ is given in (7). Let W_{nt} be defined such that

$$W_{nt}'(\tilde{\theta}_n - \theta_n) = \operatorname{vec}(\Xi_{nt}). \tag{12}$$

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

As we assumed continuity of H_t in θ, we see that W_{nt} is asymptotically equivalent to W_t. Note also that, by definition, W_{nt} is measurable with respect to $\mathcal{F}_{t-1}$.

We now have the following conditions to obtain LAN.

Assumption 5. Let W_t and W_{nt} be defined as in (9) and (12), respectively.

(a) $\sum_{t=1}^{n} \|(W_{nt} - W_t)'(\tilde{\theta}_n - \theta_n)\|^2 = o_{P_{\theta_n}}(1)$
(b) $n^{-1} \sum_{t=1}^{n} \|W_t(\theta_n) - W_t(\theta_0)\|^2 = o_{P_{\theta_n}}(1)$
(c) $n^{-1} \sum_{t=1}^{n} W_t(\theta_n) M_{\psi\psi} W_t(\theta_n)' \to_{P_{\theta_n}} J(\theta_0) > 0$, where $J(\theta) = \mathrm{E}_\theta[\dot{l}_t(\theta)\dot{l}_t(\theta)']$
(d) $n^{-1} \sum_{t=1}^{n} \|W_t(\theta_0)\|^2 1_{(n^{-1/2}|W_t(\theta_0)|>\delta)} = o_{P_{\theta_n}}(1)$ for all $\delta > 0$.

PROPOSITION 2. *Under Assumption 5, it follows that* $R_n = o_{P_{\theta_n}}(1)$ *and*

$$\Lambda_n \xrightarrow{\mathcal{D}} N\left(-\frac{1}{2}\lambda' J(\theta_0)\lambda, \lambda' J(\theta_0)\lambda\right).$$

Furthermore, the probability measures $P_{\tilde{\theta}_n}$ *and* P_{θ_n} *are contiguous.*

Denote $M_{\psi F}(\theta) = \mathrm{E}_\theta[\psi_t F_t']$, $M_{F\psi}(\theta) = M_{\psi F}'(\theta)$, and $M_{FF}(\theta) = \mathrm{E}[F_t F_t']$, where $\psi_t = \psi(v_t(\theta))$ and $F_t = F(v_t(\theta))$. We can now derive the projection of $\dot{l}_t(\theta)$ on the tangent set. The orthogonal complement of this projection is the so-called efficient score function, which can be used to do a one-step Newton–Raphson improvement of the quasi maximum likelihood estimator.

PROPOSITION 3. *The projection of* $\dot{l}_t(\theta)$ *on* $\mathcal{T}$ *is given by*

$$P_t(\theta) = \mathcal{P}(\dot{l}_t(\theta)|\mathcal{T}) = \mathrm{E}_\theta[W_t(\theta)](\psi(v_t(\theta)) - M_{\psi F}(\theta)M_{FF}^{-1}(\theta)F(v_t(\theta))). \quad \textbf{(13)}$$

Define by $l_t^*(\theta) = \dot{l}_t(\theta) - P_t(\theta)$ the efficient score function. The SP lower bound is then given by

$$V_{sp} = J^*(\theta_0)^{-1}, \quad \textbf{(14)}$$

where $J^*(\theta) = \mathrm{E}_\theta[l_t^*(\theta)l_t^{*\prime}(\theta)]$ is the SP information matrix. By definition, adaptive estimation is possible if and only if $P_t(\theta_0) = 0$, which means that the SP efficient score, l_t^*, is equal to the parametric score $\dot{l}_t$ and, hence, V_{sp} is equal to the parametric lower bound, $V_{ml} = J(\theta_0)^{-1}$. In the following proposition we characterize the asymptotic covariance matrices of the three estimation methods in terms of W_t and ψ_t.

PROPOSITION 4. *Under Assumptions 1–5, the information matrices can be expressed as*

$$J(\theta) = \mathrm{E}_\theta[W_t M_{\psi\psi} W_t'], \quad \textbf{(15)}$$

$$J^*(\theta) = \mathrm{E}_\theta[W_t M_{\psi\psi} W_t'] - \mathrm{E}_\theta[W_t]Q\mathrm{E}_\theta[W_t'], \quad \textbf{(16)}$$

$$\mathcal{J}(\theta)\mathcal{I}(\theta)^{-1}\mathcal{J}(\theta) = \mathrm{E}_\theta[W_t M_{\psi F} W_t']\mathrm{E}_\theta[W_t M_{FF} W_t']^{-1}\mathrm{E}_\theta[W_t M_{F\psi} W_t'], \quad \textbf{(17)}$$

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

where dependence of W_t, $M_{\psi\psi}$, $M_{\psi F}$, and M_{FF} on θ is suppressed and where

$$Q = M_{\psi\psi} - M_{\psi F} M_{FF}^{-1} M_{F\psi}. \tag{18}$$

As a corollary, we obtain that the difference of the information between the MLE and the SP estimator is given by the positive semidefinite matrix

$$J(\theta_0) - J^*(\theta_0) = \mathrm{E}_0[W_t(\theta_0)] Q \mathrm{E}_0[W_t'(\theta_0)].$$

The matrix Q determines the inefficiency of the SP estimator with respect to maximum likelihood estimation. Adaptive estimation is possible if and only if $Q = 0$. This would clearly be a special case, and we show later that $Q = 0$ happens only if g is Gaussian. The conclusion is that adaptive estimation of the model in (1) without reparameterization of H_t is not possible. Similar to Gonzalez-Rivera and Drost (1999) it can also be verified that $V_{sp}^{-1} - V_{qml}^{-1}$ is positive semidefinite, meaning that the SP estimator is at least as efficient as the quasi maximum likelihood estimator.

To characterize the distance of the SP estimator from adaptivity, it will be necessary to evaluate the matrices $M_{\psi\psi}$ and M_{FF} by numerical integration. However, there are special cases where they take simple forms. In the following discussion we consider two subclasses of $\mathcal{D}$ for which calculation turns out to be particularly simple. The first one is the class of spherical distributions:

$$\mathcal{D}_s = \{g : \mathbb{R}^N \to \mathbb{R}_{++} \,|\, g \in \mathcal{D}, \exists f : g(x) = f(x'x)\}, \tag{19}$$

and the second one is the class where the components are independent with identical and symmetric marginal densities:

$$\mathcal{D}_i = \left\{ g : \mathbb{R}^N \to \mathbb{R}_{++} \,|\, g \in \mathcal{D}, g(x) = \prod_{i=1}^{N} h(x_i), h(x_i) = h(-x_i) \right\}, \tag{20}$$

where $h(\cdot)$ is the marginal density of any component of v_t.

The following discussion analyzes the properties of the SP estimator with respect to maximum likelihood estimation in the case that g happens to be in $\mathcal{D}_s$ or in $\mathcal{D}_i$ without having made this assumption to construct the estimator. The situation where one has superior knowledge about g and, e.g., knows a priori that g is in $\mathcal{D}_s$, is different and is treated in Section 3.4.

The matrix M_{FF} depends on the structure of fourth moments of v_t. Lemma 3 in the Appendix implies that, for spherical distributions, the marginal kurtosis $\kappa = \mathrm{E}[v_{ti}^4]$ is linked to any co-kurtosis $c = \mathrm{E}[v_{ti}^2 v_{tj}^2], j \neq i$ by $\kappa = 3c$, and M_{FF} depends on only one parameter.

PROPOSITION 5.

(1) If $g \in \mathcal{D}$, then

$$M_{\psi F} = [0_{N^2 \times N} \quad 2D_N D_N^+ D_N^{+\prime}]. \tag{21}$$

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

(2) If $g \in \mathcal{D}_s$, then

$$M_{FF} = \begin{bmatrix} I_N & 0_{N\times N^*} \\ 0_{N^*\times N} & 2cD_N^+ D_N^{+\prime} + (c-1)\mathrm{vech}(I_N)\mathrm{vech}(I_N)' \end{bmatrix}, \tag{22}$$

$$M_{\psi\psi} = 2\tau D_N D_N^+ + (\tau - 1)\mathrm{vec}(I_N)\mathrm{vec}(I_N)', \tag{23}$$

where

$$\tau = \mathrm{E}\left[\left(\frac{\partial \log g(x)}{\partial x_1}\right)^2 x_1^2\right] \Big/ 3. \tag{24}$$

(3) If $g \in \mathcal{D}_i$, then

$$M_{FF} = \begin{bmatrix} I_N & 0_{N\times N^*} \\ 0_{N^*\times N} & I_{N^*} + [I_{N^*} \odot \{(\kappa - 2)\mathrm{vech}(I_N)\mathrm{vech}(I_N)'\}] \end{bmatrix}, \tag{25}$$

$$M_{\psi\psi} = \tau_2 D_N D_N' + D_N[I_{N^*} \odot \{(\tau_1 - 1 - \tau_2)\mathrm{vech}(I_N)\mathrm{vech}(I_N)'\}]D_N', \tag{26}$$

where $\tau_1 = 3\tau$, $\odot$ is the Hadamard product (elementwise multiplication), κ is the marginal kurtosis, and

$$\tau_2 = \mathrm{E}\left[\left(\frac{\partial \log g(x)}{\partial x_1}\right)^2\right]. \tag{27}$$

Note that Assumption 2 implies that τ in (23) and τ_2 in (26) are finite. The scalar τ_2 can be interpreted as the Fisher information for location that for distributions in $\mathcal{D}_i$ is the same for all components.

Clearly, the parametric and SP lower bounds coincide in the Gaussian case because then $c = 1$, $\tau = 1$, and therefore $Q = 0$. In a univariate framework, Gonzalez-Rivera (1997) has shown that a class of symmetric bimodal distributions allows one to attain the parametric lower bound. The following proposition states that, for higher dimensions, the Gaussian distribution is the only one for which parametric efficiency can be attained.

PROPOSITION 6. *If $N \geq 2$, then the multinormal distribution is the only one for which the parametric and SP lower bounds for the estimation of model (1) coincide.*

See however Section 3.3 for examples where reparameterizations may yield adaptive estimates of particular linear combinations of θ.

3.2. Efficient Estimation

To construct an estimator that attains the lower bound J^{*-1}, we rely on a discretization of the first-stage estimator and on sample splitting as in Linton (1993)

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

and Drost et al. (1997). This is convenient for the proofs but may not be necessary in practice. Let $\tilde{\theta}_n$ be the point in $\{\theta : \theta = n^{-1/2}\mathbb{Z}^K\}$ that is closest to the quasi maximum likelihood estimator. The vector $\tilde{\theta}_n$ can then be considered deterministic with $\tilde{\theta}_n = \theta_0 + O(n^{-1/2})$. Furthermore, denote $\tilde{v}_t = v_t(\tilde{\theta}_n) = H_t(\tilde{\theta}_n)^{-1}\varepsilon_t$.

Let α_n be an integer sequence such that $\alpha_n/n \to \alpha \in (0,1)$. Defining $\tilde{F}_t = F(\tilde{v}_t)$, consistent estimators of M_{FF} are given by $\widetilde{M}_{1FF} = \alpha_n^{-1}\sum_{t=1}^{\alpha_n}\tilde{F}_t\tilde{F}_t'$ and $\widetilde{M}_{2FF} = (n-\alpha_n)^{-1}\sum_{t=\alpha_n+1}^{n}\tilde{F}_t\tilde{F}_t'$. For the matrix $M_{\psi F}$ an explicit expression is available for all $g \in \mathcal{D}$, given by (21).

To estimate the unknown score function ψ, we use a kernel estimator of the multivariate density of $\tilde{v}_t$; see, e.g., the monograph of Scott (1992). Because the scale of the components of v_t is the same by construction, it is reasonable to use a scalar bandwidth, $h_n > 0$, and a product kernel with univariate kernel function K. Our density estimators for the two parts of the sample are

$$\hat{g}_{1h}(x;\tilde{\theta}_n) = \frac{1}{\alpha_n h_n^N}\sum_{t=1}^{\alpha_n}\prod_{i=1}^{N}K\left(\frac{\tilde{v}_{i,t}-x_i}{h_n}\right)$$

and

$$\hat{g}_{2h}(x;\tilde{\theta}_n) = \frac{1}{(n-\alpha_n)h_n^N}\sum_{t=\alpha_n+1}^{n}\prod_{i=1}^{N}K\left(\frac{\tilde{v}_{i,t}-x_i}{h_n}\right),$$

where $K(u) = (2\pi)^{-1/2}\exp(-u^2/2)$.

Using $\hat{g}$, we can estimate ψ by

$$\hat{\psi}_{it} = \begin{cases} -\mathrm{vec}\left(I_N + \dfrac{\partial \log \hat{g}_{ih}(\tilde{v}_t;\tilde{\theta}_n)}{\partial \tilde{v}_t}\,\tilde{v}_t'\right) & \text{if } \begin{cases} \hat{g}_{ih}(\tilde{v}_t) \geq d_n \\ \|\partial \log \hat{g}_{ih}(\tilde{v}_t)/\partial \tilde{v}_t\| \leq c_n \\ \|\tilde{v}_t\| \leq e_n, \end{cases} \\ -\mathrm{vec}(I_N) & \text{otherwise,} \end{cases}$$

$i = 1,2$, where $d_n \to 0$, $c_n \to \infty$, and $e_n \to \infty$. Further, let $\hat{\hat{l}}_{1t} = W_t(\tilde{\theta}_n)\hat{\psi}_{2t}$ and $\hat{\hat{l}}_{2t} = W_t(\tilde{\theta}_n)\hat{\psi}_{1t}$. Finally, we can estimate $P_t(\tilde{\theta}_n)$ in (13) by

$$\hat{P}_{1t}(\tilde{\theta}_n) = \left(\frac{1}{\alpha_n}\sum_{t=1}^{\alpha_n}W_t(\tilde{\theta}_n)\right)(\hat{\psi}_{2t} - M_{\psi F}\widetilde{M}_{1FF}^{-1}\tilde{F}_t)$$

and

$$\hat{P}_{2t}(\tilde{\theta}_n) = \left(\frac{1}{n-\alpha_n}\sum_{t=\alpha_n+1}^{n}W_t(\tilde{\theta}_n)\right)(\hat{\psi}_{1t} - M_{\psi F}\widetilde{M}_{2FF}^{-1}\tilde{F}_t)$$

and the efficient score function by $\hat{l}_{it}^*(\tilde{\theta}_n) = \hat{\hat{l}}_{it}(\tilde{\theta}_n) - \hat{P}_{it}(\tilde{\theta}_n)$, $i = 1,2$.

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

Assumption 6. There exists a square integrable matrix $W_0(\theta_0)$ such that $(1/n)\sum_{t=1}^{n} W_t(\tilde{\theta}_n) \to_{P_{\theta_n}} W_0(\theta_0)$.

Our estimator for the SP information matrix is

$$\hat{J}_n^*(\tilde{\theta}_n) = \frac{1}{n}\left(\sum_{t=1}^{\alpha_n} \hat{l}_{1t}^*(\tilde{\theta}_n)\hat{l}_{1t}^*(\tilde{\theta}_n)' + \sum_{t=\alpha_n+1}^{n} \hat{l}_{2t}^*(\tilde{\theta}_n)\hat{l}_{2t}^*(\tilde{\theta}_n)' \right).$$

We propose the following SP estimator for θ:

$$\hat{\theta}_n = \tilde{\theta}_n + \hat{J}_n^*(\tilde{\theta}_n)^{-1} \frac{1}{n}\left(\sum_{t=1}^{\alpha_n} \hat{l}_{1t}^*(\tilde{\theta}_n) + \sum_{t=\alpha_n+1}^{n} \hat{l}_{2t}^*(\tilde{\theta}_n) \right). \tag{28}$$

For the estimator $\hat{\theta}_n$ in (28) we now have the following result.

PROPOSITION 7. *If Assumptions 1–6 hold,* $c_n h_n \to 0$, *and* $c_n^2 e_n^2 n^{-1} \times h_n^{-(2+N)} \to 0$, *then* $\hat{J}_n^*(\tilde{\theta}_n) \to_{P_{\theta_n}} J^*(\theta_0)$ *and*

$$\sqrt{n}(\hat{\theta}_n - \theta_n) \xrightarrow{\mathcal{D}} N(0, J^*(\theta_0)^{-1}),$$

where $J^*(\theta_0)^{-1}$ *is the SP lower bound defined in (14).*

3.3. Adaptive Estimation

As we have seen in Proposition 6, adaptive estimation of the euclidean parameter, θ, is not possible. For a given model it may be possible, however, to find linear combinations of θ, possibly after reparameterization, such that adaptive estimation of these linear combinations is possible. Technically, this amounts to an additional orthogonality condition so that the efficient influence function is equal to the heuristic influence function. This condition is stated in the following assumption.

Assumption 7. There exists a matrix $A \in \mathbb{R}^{K\times r}$, $1 \le r \le K$, of rank r, such that

$$A' J(\theta)^{-1} W_0 = 0, \qquad P_0 - a.s.$$

Define the estimator of the information matrix by

$$\hat{J}_n(\tilde{\theta}_n) = \frac{1}{n}\left(\sum_{t=1}^{\alpha_n} \hat{\hat{l}}_{1t}(\tilde{\theta}_n)\hat{\hat{l}}_{1t}(\tilde{\theta}_n)' + \sum_{t=\alpha_n+1}^{n} \hat{\hat{l}}_{2t}(\tilde{\theta}_n)\hat{\hat{l}}_{2t}(\tilde{\theta}_n)' \right)$$

and the estimator of $\nu = A'\theta$ by

$$\hat{\nu}_n = A'\tilde{\theta}_n + A'\hat{J}_n(\tilde{\theta}_n)^{-1} \frac{1}{n}\left(\sum_{t=1}^{\alpha_n} (\overline{W}_t - \overline{W}_1)\hat{\psi}_{2t} + \sum_{t=\alpha_n+1}^{n} (\overline{W}_t - \overline{W}_2)\hat{\psi}_{1t} \right).$$

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

PROPOSITION 8. *Under Assumptions 1–7, $\hat{\nu}_n$ is an adaptive estimator of ν, i.e.,*

$$\sqrt{n}(\hat{\nu}_n - \nu) \xrightarrow{\mathcal{D}} N(0, A'J(\theta_0)^{-1}A).$$

By Corollary 3.1 of Drost et al. (1997), a matrix A that satisfies Assumption 7 can be found if $B'W_t(\theta)$ is invertible and does not depend on t, where B is a $K \times N^2$ matrix such that $A'B = 0$. We consider two examples.

Example 1.

Suppose a model for $H_t(\theta)$ can be parameterized such that $H_t(\theta) = Z_t(\nu)\Omega$, where $\theta = (\nu', \mathrm{vec}(\Omega)')'$ and $Z_t, \Omega \in \mathbb{R}^{N\times N}$ symmetric and invertible. Define $A = (I_{K-N^2}, 0_{K-N^2\times N^2})'$ and $B = (0_{N^2\times K-N^2}, I_{N^2})'$. Then, using (9),

$$\begin{aligned} B'W_t(\theta) &= \frac{\partial\, \mathrm{vec}(H_t)'}{\partial\, \mathrm{vec}(\Omega)} (I_N \otimes H_t'^{-1}) \\ &= (I_N \otimes Z_t)(I_N \otimes Z_t^{-1}\Omega^{-1}) \\ &= (I_N \otimes \Omega^{-1}), \end{aligned}$$

and therefore $A'\theta = \nu$ is adaptively estimable. Most of the well-known multivariate GARCH-type models can be parameterized in this form, but unfortunately they are all for Σ_t and not for H_t. For example, if the BEKK model was formulated for H_t, then the "dynamic part" of the model could be estimated adaptively after a similar reparametrization as in the univariate case.

The second example is for a model for Σ_t.

Example 2.

Let Z_t, Ω, θ, ν, A, and B be defined as in Example 1. Suppose a model for $\Sigma_t(\theta)$ can be parameterized such that $\Sigma_t(\theta) = Z_t(\nu)\Omega^2 Z_t(\nu)$. Let the square root be $H_t(\theta) = Z_t(\nu)\Omega$, which is *not* restricted to be symmetric. Thus, we can use (11) to obtain

$$B'W_t(\theta) = \frac{1}{2}(\Omega^{-1} \otimes \Omega^{-1}),$$

and therefore $A'\theta = \nu$ is adaptively estimable. For example, the well-known CCC model of Bollerslev (1990) falls into this category. Assuming univariate GARCH(1,1) models for the conditional variances, define $Z_t = diag(1 + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i Z_{ii,t-1})_{i=1}^N$, $D = diag(\omega_1, \ldots, \omega_N)$, and R the CCC matrix. Then the CCC model can be written in the form $\Sigma_t(\theta) = Z_t(\nu)\Omega^2 Z_t(\nu)$ with $\Omega = (DRD)^{1/2}$. The parameter $\nu = (\alpha_1, \ldots, \alpha_N, \beta_1, \ldots, \beta_N)'$ is adaptively estimable.

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

3.4. Efficient Estimation in $\mathcal{D}_s$

In practice one may want to impose more structure on the multivariate density g to avoid the curse of dimensionality, e.g., the data sparseness problem described by Silverman (1986). One way to do so is to assume that g belongs to the class of spherical distributions, which we do in Assumption 8.

Assumption 8. For the density g, assume $g \in \mathcal{D}_s$, where $\mathcal{D}_s$ is defined by (19).

The density $g(v_t)$ is said to be spherical if there exists a positive function f such that $g(v_t) = f(w_t)$ with $w_t = v_t' v_t$. Examples of spherical distributions are the multivariate versions of the normal, the t and Laplace distributions. The class of spherical distributions (or elliptical, respectively, if the conditional distribution of ε_t is considered) is the most general one that is consistent with the conditional capital asset pricing model, as shown by Berk (1997); see also Hodgson and Vorkink (2003) and Brown and Hodgson (2004). In our framework, the main advantage of Assumption 8 is that it allows for univariate convergence rates of nonparametric estimators of g.

SP estimators of model (1) that are based on Assumption 8 have a lower bound that is larger than the bound in the general case. The score in the nonparametric direction, $s(\theta, \eta)$, is now a function of v_t only through w_t, and therefore, the tangent set contains functions that all depend on w_t only. But still, these functions have to be orthogonal to F_t because g is restricted to have mean zero and identity covariance matrix. Thus, the tangent set under Assumption 8 is defined by

$$\mathcal{T}_s = \{f: \mathbb{R}_+ \to \mathbb{R}^K | \mathrm{E}_g[f(y)] = 0, \mathrm{E}_g[f(y)F'(x)] = 0, \mathrm{E}_g[f(y)f(y)'] < \infty\},$$
$$y = x'x.$$

Note that $\mathcal{T}_s \subset \mathcal{T}$. We can now derive the projection of $\dot{l}_t(\theta)$ on $\mathcal{T}_s$.

PROPOSITION 9. *Under Assumption 8, the projection of $\dot{l}_t(\theta)$ on $\mathcal{T}_s$ is given by*

$$P_t^s(\theta) = \mathcal{P}(\dot{l}_t(\theta)|\mathcal{T}_s) = \mathrm{E}[W_t(\theta)](\psi_t^s - M_{\psi^s F} M_{F^s F}^{-1} F_t^s), \tag{29}$$

where $\psi_t^s = \psi^s(w_t(\theta)) = \mathrm{E}_\theta[\psi(v_t(\theta))|w_t]$, $F_t^s = F^s(w_t(\theta)) = \mathrm{E}_\theta[F(v_t)|w_t]$, $M_{\psi^s F} = \mathrm{E}_\theta[\psi_t^s F_t']$, and $M_{F^s F} = \mathrm{E}_\theta[F_t^s F_t']$.

Note that F_t^s contains the conditional expectations $\mathrm{E}[v_t|w_t]$ and $\mathrm{E}[v_t v_t'|w_t]$. The first one is zero because the distribution of v_t conditional on w_t is symmetric in v_t. The second one, $\mathrm{E}[v_t v_t'|w_t]$, can be estimated with arbitrary precision as described by Hodgson and Vorkink (2003). We can therefore consider it known and denote it by $m(w_t)$.

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

In the following discussion we propose an estimator of θ in the spherical case. To simplify the presentation we do not use sample splitting here. Because y has a positive support, one faces the problem of estimating its density near the boundary. To solve this problem, Hodgson, Linton, and Vorkink (2002) and Hodgson and Vorkink (2003) apply a Box–Cox transformation to y and then use the standard kernel density estimator to the transformed variable. We use an alternative method by applying a gamma kernel estimator; see the paper by Chen (2000), who shows that gamma kernels are particularly suited for the estimation of density functions with bounded support. As before we use a first-stage consistent estimator, $\tilde{\theta}_n$, say, and let $\tilde{v}_t = v_t(\tilde{\theta}_n)$ and $\tilde{w}_t = \tilde{v}_t'\tilde{v}_t$. First, the gamma kernel estimator of $g(x)$ is given by

$$\hat{g}_b(x) = \frac{\Gamma(N/2)}{\pi^{N/2}} (x'x)^{1-N/2} \frac{1}{n} \sum_{t=1}^{n} G_{x'x/b+1,b}(\tilde{w}_t),$$

where $G_{p,q}$ is the density function of a gamma(p,q) random variable. As $\hat{g}_b(x)$ depends only on $y = x'x$, we can write $\hat{g}_b(x) = \hat{f}_b(y)$. Using $\hat{f}$, we can estimate ψ by

$$\hat{\psi}_t^s = \begin{cases} -\mathrm{vec}\left(I_N + 2 \dfrac{\partial \log \hat{f}_b(\tilde{w}_t)}{\partial \tilde{w}_t} m(\tilde{w}_t)\right) & \text{if} \begin{cases} \hat{f}_b(\tilde{w}_t) \geq d_n \\ |\partial \log \hat{f}_b(\tilde{w}_t)/\partial \tilde{w}_t| \leq c_n \\ \tilde{w}_t \leq e_n \\ \tilde{w}_t \geq f_n, \end{cases} \\ -\mathrm{vec}(I_N) & \text{otherwise,} \end{cases}$$

where $d_n \to 0$, $c_n \to \infty$, $e_n \to \infty$, and $f_n \to 0$. Further, let $\hat{\tilde{l}}_t^s = W_t(\tilde{\theta}_n)\hat{\psi}_t^s$. Finally, we can estimate $P_t^s(\tilde{\theta}_n)$ in (29) by

$$\hat{P}_t^s(\tilde{\theta}_n) = \left(\frac{1}{n} \sum_{t=1}^{n} W_t(\tilde{\theta}_n)\right)(\hat{\psi}_t^s - M_{\psi^s F} \widetilde{M}_{1F^s F}^{-1} \tilde{F}_t^s)$$

and the efficient score function by $\hat{l}_t^{*s}(\tilde{\theta}_n) = \hat{\tilde{l}}_t^s(\tilde{\theta}_n) - \hat{P}_t^s(\tilde{\theta}_n)$. Analogous to the general case define

$$\hat{J}_n^{*s}(\tilde{\theta}_n) = \frac{1}{n}\left(\sum_{t=1}^{n} \hat{l}_t^{*s}(\tilde{\theta}_n)\hat{l}_t^{*s}(\tilde{\theta}_n)'\right)$$

and the estimator

$$\hat{\theta}_n = \tilde{\theta}_n + \hat{J}_n^{*s}(\tilde{\theta}_n)^{-1} \frac{1}{n} \sum_{t=1}^{n} \hat{l}_t^{*s}(\tilde{\theta}_n). \tag{30}$$

PROPOSITION 10. *If Assumptions 1–8 hold, $c_n b_n \to 0$, and either $c_n^2 n^{-1}\{(b_n f_n)^{-1/2} + (b_n e_n)^{1/2}\} \to 0$ for $N = 1$ or $c_n^2 n^{-1} b_n^{-1/2} f_n^{1/2-N} \to 0$*

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

for $N \geq 2$, *then* $\hat{J}_n^{*s}(\tilde{\theta}_n) \to_{P_{\theta_n}} J^{*s}(\theta_0)$, *where* $J^{*s}(\theta_0) = \mathrm{E}_0[\{\dot{l}_t(\theta_0) - P_t^s(\theta_0)\}\{\dot{l}_t(\theta_0) - P_t^s(\theta_0)\}']$ *and*

$$\sqrt{n}(\hat{\theta}_n - \theta_n) \xrightarrow{\mathcal{D}} N(0, J^s(\theta_0)^{-1}).$$

Note that this result is similar to that of Proposition 7, but because the gamma kernel is unbounded near zero, we have to take care that $\hat{\psi}_t^s$ is bounded by using appropriate trimming conditions.

4. FINITE-SAMPLE PERFORMANCE

In this section we are interested in the performance of the proposed SP estimator (28) relative to the quasi maximum likelihood estimator and MLE in finite samples. Intuitively, the SP method should perform better than quasi maximum likelihood, but worse than maximum likelihood, when there are strong departures from normality.

For convenience we consider the VEC model for Σ_t in (4) of order $p = q = 1$ and define H_t as in (2). To calculate $W_t(\theta)$ in (10) one needs to evaluate $\partial \operatorname{vec}(\Sigma_t)'/\partial\theta$. We obtain

$$\frac{\partial \operatorname{vech}(\Sigma_t)}{\partial \omega'} = (I_{N^*} - B)^{-1},$$

$$\frac{\partial \operatorname{vech}(\Sigma_t)}{\partial \operatorname{vec}(A)'} = \eta_{t-1}' \otimes I_{N^*} + B \frac{\partial \operatorname{vech}(\Sigma_{t-1})}{\partial \operatorname{vec}(A)'},$$

$$\frac{\partial \operatorname{vech}(\Sigma_t)}{\partial \operatorname{vec}(B)'} = h_{t-1}' \otimes I_{N^*} + B \frac{\partial \operatorname{vech}(\Sigma_{t-1})}{\partial \operatorname{vec}(B)'},$$

where $\eta_{t-1} = \operatorname{vech}(\varepsilon_{t-1}\varepsilon_{t-1}')$. Empirical estimates using financial time series typically imply positive unconditional correlation, small off-diagonal elements of A and B, and diagonal elements of B larger than those of A. Motivated by this observation, we choose our process as

$$h_t = \begin{bmatrix} 1 \\ 0.7 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.2 \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.6 \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{bmatrix}.$$

For the distributional assumption on v_t we take the bivariate standardized t distribution with five degrees of freedom. We assume that it is known that g belongs to the spherical class. For the nonparametric estimation of $g(\cdot)$, this assumption allows us to use the gamma kernel estimator of Section 3.4. The employed estimator is given by (30). Chen (2000) provides formulas for the

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

TABLE 1. Monte Carlo results based on 500 replications of the diagonal VEC model with $n = 2{,}000$

	ML			QML			SP		
Population	Mean	SD	MSE	Mean	SD	MSE	Mean	SD	MSE
$c_{11} = 1$	1.021	0.209	0.0441	1.043	0.286	0.0841	1.0121	0.2394	0.0574
$c_{21} = 0.7$	0.725	0.245	0.0609	0.739	0.309	0.0975	0.7227	0.2964	0.0883
$c_{22} = 1$	1.039	0.218	0.0493	1.060	0.285	0.0853	1.0302	0.2328	0.0551
$\beta_{11} = 0.5$	0.492	0.081	0.0066	0.483	0.113	0.0131	0.4961	0.0930	0.0086
$\beta_{22} = 0.1$	0.070	0.269	0.0735	0.060	0.324	0.1066	0.0824	0.3211	0.1034
$\beta_{33} = 0.6$	0.591	0.062	0.0040	0.581	0.084	0.0075	0.5950	0.0690	0.0047
$\alpha_{11} = 0.2$	0.201	0.038	0.0014	0.201	0.055	0.0030	0.1942	0.0400	0.0016
$\alpha_{22} = 0.1$	0.099	0.032	0.0010	0.102	0.043	0.0018	0.0962	0.0347	0.0012
$\alpha_{33} = 0.2$	0.201	0.035	0.0012	0.208	0.052	0.0028	0.1961	0.0378	0.0014

Note: The innovation density is a bivariate t distribution with five degrees of freedom. ML = maximum likelihood estimation, QML = quasi maximum likelihood estimation, and SP = semiparametric estimation. MSE = mean squared error, and SD = standard deviation.

bandwidth b that minimize the mean integrated squared error that we use here. The results are displayed in Table 1.

Concerning the bias, the three estimation procedures perform similarly; perhaps one could notice that SP estimation performs better than QMLE for most of the parameters. There are, however, substantial differences between the standard deviations. Clearly maximum likelihood estimation performs best for all the parameters. SP estimation is as expected in between the two other procedures, also for all the parameters. The same holds true for the mean square error of $\hat{\theta}$. One can see that a good part of the loss of the inefficient QMLE (compared to maximum likelihood estimation) is recaptured by SP estimation. We also compared the performance of the univariate gamma kernel estimator with the bivariate product kernel estimator when the assumption of sphericity is dropped and the estimator (28) is used. The latter estimator had higher mean squared errors for all parameters but still considerably less than the quasi maximum likelihood mean squared errors. However, the relative performance of the multivariate kernel estimator for $N > 2$ is likely to become worse if the dimension increases and if the true density is close to the spherical class.

5. CONCLUSIONS AND OUTLOOK

This paper shows that efficiency gains of SP univariate volatility models over QMLE carry over to the multivariate case. We suggest two alternative types of SP estimator. The first one applies to general innovation densities, whereas the second one is based on the assumption of sphericity. An advantage of the sphericity assumption is that a nonparametric density estimator can be constructed

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

that has univariate convergence rates. Both estimators are efficient, but the SP and parametric lower bounds are different in general, so that adaptive estimation is not possible.

In practice, the general approach outlined in Sections 3.1 and 3.2 will only be feasible in small dimensions. In high dimensions one would like to reduce the dimensionality, and the assumption of sphericity is only one way of doing this. Another one would be to assume that innovations are componentwise i.i.d. One would again obtain univariate convergence rates and hence avoid the curse of dimensionality. Another advantage of componentwise independence of v_t is that it is possible to define impulse response functions for volatility that avoid typical orthogonalization and ordering problems as shown by Hafner and Herwartz (2006).

Yet another possibility is to work with copulas. Copulas allow one to decompose the information captured in the joint distribution into information concerning the marginal distributions and information about the dependence structure. To keep a flexible innovation distribution one could estimate the univariate marginal distributions nonparametrically and fit a parametric copula. Restricting the innovation distribution to have identity covariance matrix can be achieved as in Lee and Long (2005). This seems a promising direction for future research.

NOTES

1. After an unpublished paper by Baba, Engle, Kraft, and Kroner.

2. Here D_N denotes the $N^2 \times N(N+1)/2$ duplication matrix, defined by the property $D_N \operatorname{vech}(A) = \operatorname{vec}(A)$ for any symmetric $(N \times N)$ matrix A, and D_N^+ denotes its generalized inverse.

REFERENCES

Bauwens, L., S. Laurent, & J.V.K. Rombouts (2006) Multivariate GARCH models: A survey. *Journal of Applied Econometrics* 21, 79–109.

Berk, J. (1997) Necessary conditions for the CAPM. *Journal of Economic Theory* 73, 245–257.

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

Bickel, P.J., C.A.J. Klaassen, Y. Ritov, & J.A. Wellner (1993) *Efficient and Adaptive Statistical Inference for Semiparametric Models*. Johns Hopkins University Press.

Bollerslev, T. (1990) Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach. *Review of Economics and Statistics* 72, 498–505.

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

Box, G.E.P. & J.S. Hunter (1957) Multi-factor experimental designs for exploring response surfaces. *Annals of Mathematical Statistics* 28, 195–241.

Brown, B.W. & D.J. Hodgson (2004) Efficient Semiparametric Estimation of Dynamic Nonlinear Systems under Elliptical Symmetry. Mimeo, Rice University.

Chen, S.X. (2000) Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics* 52, 471–480.

Comte, F. & O. Lieberman (2003) Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* 84, 61–84.

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

Drost, F.C. & C.A.J. Klaassen (1997) Efficient estimation in semiparametric GARCH models. *Journal of Econometrics* 81, 193–221.
Drost, F.C., C.A.J. Klaassen, & B.J.M. Werker (1997) Adaptive estimation in time series models. *Annals of Statistics* 25, 786–817.
Engle, R.F. (2002) Dynamic conditional correlation—A simple class of multivariate GARCH models. *Journal of Business & Economic Statistics* 20, 339–350.
Engle, R. & G. Gonzalez-Rivera (1991) Semiparametric ARCH models. *Journal of Business & Economic Statistics* 9, 345–359.
Engle, R. & K. Kroner (1995) Multivariate simultaneous GARCH. *Econometric Theory* 11, 122–150.
Fang, K.T., S. Kotz, & K.W. Ng (1990) *Symmetric Multivariate and Related Distributions*. Chapman and Hall.
Gonzalez-Rivera, G. (1997) A note on adaptation in ARCH models. *Econometric Reviews* 16, 55–68.
Gonzalez-Rivera, G. & F.C. Drost (1999) Efficiency comparisons of maximum-likelihood-based estimators in GARCH models. *Journal of Econometrics* 93, 93–111.
Hafner, C.M. & H. Herwartz (2006) Volatility impulse response functions for multivariate GARCH models: An exchange rate illustration. *Journal of International Money and Finance* 25, 719–740.
Hodgson, D.J., O. Linton, & K.P. Vorkink (2002) Testing the capital asset pricing model efficiently under elliptical symmetry: A semiparametric approach. *Journal of Applied Econometrics* 17, 617–639.
Hodgson, D.J. & K.P. Vorkink (2003) Efficient estimation of conditional asset-pricing models. *Journal of Business & Economic Statistics* 52, 471–480.
Jeantheau, T. (1998) Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* 14, 70–86.
Lee, T. & X. Long (2005) Copula-Based Multivariate GARCH Model with Uncorrelated Dependent Errors. Working paper, University of California, Riverside.
Linton, O. (1993) Adaptive estimation of ARCH models. *Econometric Theory* 9, 539–569.
Kreiss, J.P. (1987) On adaptive estimation in stationary ARMA processes. *Annals of Statistics* 15, 112–133.
Lütkepohl, H. (1996) *Handbook of Matrices*. Wiley.
Newey, W.K. (1989) Locally Efficient, Residual-Based Estimation of Nonlinear Simultaneous Equations. Bellcore Economics Discussion Paper 59.
Newey, W.K. (1990) Semiparametric efficiency bounds. *Journal of Applied Econometrics* 5, 99–135.
Scott, D. (1992) *Multivariate Density Estimation: Theory, Practice, and Visualisation*. Wiley.
Severini, T. & G. Tripathi (2001) A simplified approach to computing efficiency bounds in semiparametric models. *Journal of Econometrics* 102, 23–66.
Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.
Steigerwald, D.G. (1992) Adaptive estimation in time series regression models. *Journal of Econometrics* 54, 251–275.

APPENDIX A: Index Set Definitions

Define the index

$$k_{ij}^N = i + (j-1)\left(N - \frac{j}{2}\right) \tag{A.1}$$

and the index sets

$$\mathcal{K}_{ij}^N = \begin{cases} \varnothing & N = 1 \\ \{k_{ij}^N \mid j = 1, \ldots, N-1; \quad i = j+1, \ldots, N\} & N \geq 2 \end{cases} \tag{A.2}$$

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

and

$$\mathcal{K}_{ii}^N = \{k_{ii}^N | i = 1, \ldots, N\}. \tag{A.3}$$

The index k_{ij}^N is the position of the (i,j)th element of an $(N \times N)$ symmmetric matrix A in the vector vech(A). Remember that vech(A) contains $N^* = N(N+1)/2$ elements. The set $\mathcal{K}_{ij}^N$ contains all indices of the elements below the diagonal of A in the vector vech(A); this set contains $N(N-1)/2$ elements. The set $\mathcal{K}_{ii}^N$ contains all indices of the N diagonal elements. For example, for $N = 2$, $\mathcal{K}_{ij}^2 = \{2\}$ and $\mathcal{K}_{ii}^2 = \{1,3\}$, and for $N = 3$, $\mathcal{K}_{ij}^3 = \{2,3,5\}$ and $\mathcal{K}_{ii}^3 = \{1,4,6\}$. Note that $\mathcal{K}_{ij}^N \cup \mathcal{K}_{ii}^N = \{1, \ldots, N^*\}$ and $\mathcal{K}_{ij}^N \cap \mathcal{K}_{ii}^N = \varnothing$.

APPENDIX B: Lemmas

LEMMA 1. *The matrix $D_N^+ D_N^{+\prime}$ is an $(N^* \times N^*)$ diagonal matrix with 1 at the (i,i)th position, $i \in \mathcal{K}_{ii}^N$, and $\frac{1}{2}$ at the (j,j)th position, $j \in \mathcal{K}_{ij}^N$, where $\mathcal{K}_{ij}^N$ is defined in (A.2) and $\mathcal{K}_{ii}^N$ in (A.3).*

Proof. The statement holds for $D_1^+ D_1^{+\prime} = 1$. Noting that $k_{ii}^{N+1} = k_{i-1,i-1}^N + N + 1$, where k_{ij}^N is defined in (A.1), and using the recursive equation for $D_{N+1}^+ D_{N+1}^{+\prime}$ in Lütkepohl (1996, p. 125), the statement follows by induction. ■

LEMMA 2. *The matrix* vech(I_N)vech$(I_N)'$ *is an $(N^* \times N^*)$ matrix with 1 at the (i,j)th position, $i,j \in \mathcal{K}_{ii}^N$, and 0 elsewhere, where $\mathcal{K}_{ii}^N$ is defined in (A.3).*

Proof. By definition, $\mathcal{K}_{ii}^N$ contains the positions of the diagonal elements of an $(N \times N)$ matrix A in the vector vech(A). Therefore, $\mathcal{K}_{ii}^N$ contains the positions of the ones in the vector vech(I_N). The matrix vech(I_N)vech$(I_N)'$ then contains ones at pairs of any permutations of these positions and zeros elsewhere. So there is a total of N^2 ones in the matrix. ■

LEMMA 3. *For any spherical distribution,*

$$\mathrm{E}\left[\prod_{j=1}^N X_j^{\alpha_j}\right] = \begin{cases} 0 & \text{if one (or more) } \alpha_j \text{ is odd} \\ K_\alpha \prod_{j=1}^N \dfrac{\alpha_j!}{(\alpha_j/2)!} & \text{if all } \alpha_j \text{ are even,} \end{cases}$$

where $\alpha = \sum_{j=1}^N \alpha_j$ and K_α depends on α only.

Proof. See Box and Hunter (1957).

LEMMA 4. *Under the conditions of Proposition 7, $\hat{J}_n^*(\tilde{\theta}_n) \to_{P_{\theta_0}} J^*(\theta_0)$.*

Proof. The proof is similar to Lemma 3.1 of Drost et al. (1997). Let $\overline{W}_1 = (1/\alpha_n)\sum_{t=1}^{\alpha_n} \widetilde{W}_t$ and $\tilde{\phi}_{1t} = \overline{W}_1 M_{\psi F} \widetilde{M}_{1FF}^{-1} \widetilde{F}_t$. Then, for some $a \in \mathbb{R}^K$,

$$\left| \frac{1}{\alpha_n} \sum_{t=1}^{\alpha_n} |a'\hat{l}_{1t}^*|^2 - a'J_0^* a \right| \le \left| \frac{1}{\alpha_n} \sum_{t=1}^{\alpha_n} |a'\tilde{l}_{1t}^*|^2 - a'J_0^* a \right| + \left| \frac{1}{\alpha_n} \sum_{t=1}^{\alpha_n} |a'\hat{l}_{1t}^*|^2 - |a'\tilde{l}_{1t}^*|^2 \right|. \tag{B.1}$$

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

The first term on the right-hand side of (B.1) is $o_p(1)$ by invoking a weak law of large numbers. The second term is

$$\left| \frac{1}{\alpha_n} \sum_{t=1}^{\alpha_n} |a'\hat{I}_{1t}^*|^2 - |a'\tilde{I}_{1t}^*|^2 \right|$$

$$= \left| \frac{1}{\alpha_n} \sum_{t=1}^{\alpha_n} |a'((\widetilde{W}_t - \overline{W}_1)\hat{\psi}_{2t} + \tilde{\phi}_{1t})|^2 - |a'((\widetilde{W}_t - \overline{W}_1)\tilde{\psi}_t + \tilde{\phi}_{1t})|^2 \right|$$

$$\leq \frac{1}{\alpha_n} \sum_{t=1}^{\alpha_n} |a'(\widetilde{W}_t - \overline{W}_1)(\hat{\psi}_{2t} - \tilde{\psi}_t)|^2$$

$$+ \left| \frac{1}{\alpha_n} \sum_{t=1}^{\alpha_n} 2a'(\widetilde{W}_t - \overline{W}_1)(\hat{\psi}_{2t} - \tilde{\psi}_t)(\tilde{\psi}_t'(\widetilde{W}_t - \overline{W}_1)' + \tilde{\phi}_{1t}')a \right|. \tag{B.2}$$

Applying Cauchy–Schwarz, and condition (3.1) and Lemma 2.2 of Drost et al. (1997), both terms on the right-hand side of (B.2) converge to zero under $P_{\tilde{\theta}_n}$. The same argument applies to the estimator using the second part of the sample. This implies that $\hat{J}_n^*(\tilde{\theta}_n)$ converges to $J^*(\theta_0)$ under $P_{\tilde{\theta}_n}$. By contiguity (Proposition 2), this implies convergence under P_{θ_n}. ■

LEMMA 5. *Define* $g_h(x) = h^{-N}\int_{\mathbb{R}^N} \prod_{i=1}^N K((x_i - y_i)/h)g(y)\,dy$. *Under the conditions of Proposition 7, there are suitable constants* κ_0 *and* κ_1 *such that*

$$\mathrm{E}_{\theta_n}[\|x\|^2 g_h^{-1}(x)\{\hat{g}_h(x,\theta_n) - g_h(x)\}^2] \leq \frac{1}{nh_n^N}\kappa_0\|x\|^2$$

and

$$\max_{i=1,\ldots,N} \mathrm{E}_{\theta_n}[\|x\|^2 g_h^{-1}(x)\{\hat{g}_h^{(i)}(x,\theta_n) - g_h^{(i)}(x)\}^2] \leq \frac{1}{nh_n^{N+2}}\kappa_1\|x\|^2,$$

where $\hat{g}_h^{(i)}(x,\theta_n)$ *and* $g_h^{(i)}(x)$ *are the ith partial derivatives of* $\hat{g}_h$ *and* g_h, *respectively, evaluated at x.*

Proof. See, e.g., Scott (1992).

LEMMA 6. *If* $G_{p,q}(t)$ *is the density function of a gamma* (p,q) *random variable, then* $\partial G_{p,q}(t)/\partial p = G_{p,q}(t)\{\ln(t/q) - \Psi(p)\}$, *where* $\Psi(\cdot)$ *is the digamma function.*

Proof. The proof consists of straightforward calculations that are omitted.

LEMMA 7. $\mathrm{Var}(n^{-1}\sum_t \partial G_{y/b+1,b}(w_t)/\partial x) = O(n^{-1}b^{1/2}y^{-3/2})$ *for* $y/b \to \infty$.

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

Proof. Using Lemma 6 and the results of Chen (2000),

$$\operatorname{Var}\left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial G_{y/b+1,b}(w_t)}{\partial y}\right) = n^{-1}B_b(y)\mathrm{E}[\,f(\eta_y)\{\log(\eta_y/b)-\Psi(y/b+1)\}^2] \le \left(\sup_x f(x)\right) n^{-1}B_b(y)\mathrm{E}[\{\log(\eta_y/b)-\Psi(y/b+1)\}^2], \tag{B.3}$$

where η_y is a gamma$(2y/b+1, b/2)$ r.v. and $B_b(y) = O(b^{-5/2}y^{-1/2})$. Note that $\mathrm{E}[\eta_y] = y + b/2$ and $\operatorname{Var}(\eta_y) = yb/2 + b^2/4$. We now expand the expectation on the right-hand side of (B.3) about $\mathrm{E}[\eta_y] = y + b/2$. The digamma function has asymptotic expansion $\Psi(z+1) = \log z + 1/(2z) - \sum_{i=1}^{\infty} B_{2i}/(2iz^{2i})$, where B_{2i} are Bernoulli numbers. Thus, $\Psi(z+1) = \log(z+\frac{1}{2}) + O(z^{-2})$, and hence

$$\begin{aligned}
&\mathrm{E}[\{\log(\eta_y/b)-\Psi(y/b+1)\}^2] \\
&\quad = \mathrm{E}\left[\left\{O((y/b)^{-2}) + \frac{\eta_y-\mu_y}{y/b+1/2} - \frac{1}{2}\left(\frac{\eta_y-\mu_y}{y/b+1/2}\right)^2 + \dots\right\}^2\right] \\
&\quad = O((y/b)^{-2}\operatorname{Var}(\eta_y)) = O((y/b)^{-2}(yb/2+b^2/4)) \\
&\quad = O(b^3y^{-1}).
\end{aligned}$$

Finally, $\operatorname{Var}((1/n)\sum_{t=1}^{n}(\partial G_{y/b+1,b}(w_t)/\partial y)) = O(n^{-1}B_b(y)b^3y^{-1}) = O(n^{-1}b^{1/2}y^{-3/2})$. ■

APPENDIX C: Proofs

Proof of Proposition 1. Let us write the likelihood as $L(\theta) = \sum_{t=1}^{n} l_t(\theta)$ with

$$l_t = -\frac{1}{2}\log\det\Sigma_t + \log g(H_t^{-1}\varepsilon_t),$$

where $\Sigma_t = H_tH_t'$ and $v_t = v_t(\theta) = H_t^{-1}(\theta)\varepsilon_t$. The score vector is given by

$$\frac{\partial l_t(\theta)}{\partial\theta} = -\frac{1}{2}\frac{\partial\log\det\Sigma_t}{\partial\theta} + \frac{\partial\log g(v_t)}{\partial\theta},$$

where the first term is

$$\frac{\partial\log\det\Sigma_t}{\partial\theta'} = 2\operatorname{vec}(H_t'^{-1})'\frac{\partial\operatorname{vec}(H_t)}{\partial\theta'}.$$

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

With the chain rule for matrix differentiation we can write

$$\frac{\partial \log g(v_t)}{\partial \theta_i} = \frac{\partial \log g(v_t)}{\partial v_t'} \frac{\partial v_t}{\partial \operatorname{vec}(H_t)'} \frac{\partial \operatorname{vec}(H_t)}{\partial \theta_i}.$$

We can further write

$$\begin{aligned}\frac{\partial(v_t)}{\partial \operatorname{vec}(H_t)'} &= -(\varepsilon_t' \otimes I_N)(H_t'^{-1} \otimes H_t^{-1}) \\ &= -(v_t' \otimes I_N)(H_t' \otimes I_N)(H_t'^{-1} \otimes H_t^{-1}) \\ &= -(v_t' \otimes I_N)(I_N \otimes H_t^{-1}).\end{aligned}$$

Thus,

$$\begin{aligned}\frac{\partial \log g(v_t)}{\partial \theta'} &= -\frac{\partial \log g(v_t)}{\partial v_t'} (v_t' \otimes I_N)(I_N \otimes H_t^{-1}) \frac{\partial \operatorname{vec}(H_t)}{\partial \theta'} \\ &= -\operatorname{vec}\left(\frac{\partial \log g(v_t)}{\partial v_t} v_t'\right)' (I_N \otimes H_t^{-1}) \frac{\partial \operatorname{vec}(H_t)}{\partial \theta'} \\ &= -\operatorname{vec}\left(\frac{\partial \log g(v_t)}{\partial v_t} v_t'\right)' W_t'\end{aligned}$$

with $W_t = (\partial \operatorname{vec}(H_t)'/\partial\theta)(I_N \otimes H_t'^{-1})$. Defining $\psi_t = -\operatorname{vec}(I_N + (\partial \log g(v_t)/\partial v_t)v_t')$, this can be written as

$$\frac{\partial \log g(v_t)}{\partial \theta_i} = \psi_t' W_t' + \operatorname{vec}(I_N)' W_t'.$$

However, $\operatorname{vec}(I_N)' W_t' = \operatorname{vec}(H_t'^{-1})'(\partial \operatorname{vec}(H_t)/\partial\theta')$. Thus, the score vector can be written as

$$\begin{aligned}\frac{\partial l_t(\theta)}{\partial \theta} &= \frac{\partial \operatorname{vec}(H_t)'}{\partial \theta} \left\{-\operatorname{vec}(H_t'^{-1}) - (I_N \otimes H_t'^{-1}) \operatorname{vec}\left(\frac{\partial \log g(v_t)}{\partial v_t} v_t'\right)\right\} \\ &= -\frac{\partial \operatorname{vec}(H_t)'}{\partial \theta} \operatorname{vec}(H_t'^{-1}) + W_t \psi_t + \frac{\partial \operatorname{vec}(H_t)'}{\partial \theta} \operatorname{vec}(H_t'^{-1}) \\ &= W_t \psi_t,\end{aligned}$$

as stated. ■

Proof of Proposition 2. The proof follows directly from Theorem 2.1 of Drost et al. (1997).

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

Proof of Proposition 3. First note that $P_t \in \mathcal{T}$ because $\mathrm{E}[P_t] = 0$, $\mathrm{E}[P_t F_t'] = 0$, and $\mathrm{E}[P_t P_t']$ is finite because $\mathrm{E}[\psi_t \psi_t'] < \infty$ by Assumption 2 and $\mathrm{E}[F_t F_t'] < \infty$ by Assumption 1 (finite fourth moments).

Next, we show that the orthogonal complement of the projection is orthogonal to $\mathcal{T}$. It can be written as

$$\dot{l}_t^*(\theta) = \dot{l}_t(\theta) - P_t = (W_t - \mathrm{E}[W_t])\psi_t + \mathrm{E}[W_t] M_{\psi F} M_{FF}^{-1} F_t. \tag{C.1}$$

The first term on the right-hand side of (C.1) is orthogonal to $\mathcal{T}$ because $(W_t - \mathrm{E}[W_t])$ has mean zero and is independent of v_t and, hence, independent of all elements in $\mathcal{T}$. The second term on the right-hand side of (C.1) is orthogonal to $\mathcal{T}$ because it consists of linear combinations of F_t and F_t is, by definition, orthogonal to $\mathcal{T}$. ■

Proof of Proposition 4.

(a) $J = \mathrm{E}[\dot{l}_t \dot{l}_t'] = \mathrm{E}[W_t \psi_t \psi_t' W_t'] = \mathrm{E}[W_t M_{\psi\psi} W_t']$.

(b) $J^* = \mathrm{E}[\dot{l}_t \dot{l}_t'] - \mathrm{E}[\dot{l}_t P_t'] - \mathrm{E}[P_t \dot{l}_t'] + \mathrm{E}[P_t P_t']$.
The second term is

$$\begin{aligned}\mathrm{E}[\dot{l}_t P_t'] &= \mathrm{E}[W_t \psi_t (\psi_t' - F_t' M_{FF}^{-1} M_{F\psi}) \mathrm{E}(W_t')] \\ &= \mathrm{E}[W_t] \mathrm{E}[\psi_t \psi_t' - \psi_t F_t' M_{FF}^{-1} M_{F\psi}] \mathrm{E}[W_t'] \\ &= \mathrm{E}[W_t](M_{\psi\psi} - M_{\psi F} M_{FF}^{-1} M_{F\psi}) \mathrm{E}[W_t'].\end{aligned}$$

Similar calculations show that $\mathrm{E}[\dot{l}_t P_t'] = \mathrm{E}[P_t \dot{l}_t'] = \mathrm{E}[P_t P_t']$, which then gives the stated result.

(c) Recall that $V_{qml} = \mathcal{J}^{-1} \mathcal{I} \mathcal{J}^{-1}$ and thus $V_{qml}^{-1} = \mathcal{J} \mathcal{I}^{-1} \mathcal{J}$ with

$$\mathcal{J} = -\mathrm{E}\left[\frac{\partial^2 l_t^{qml}}{\partial\theta \partial\theta'}\right], \qquad \mathcal{I} = \mathrm{E}\left[\frac{\partial l_t^{qml}}{\partial\theta} \frac{\partial l_t^{qml}}{\partial\theta'}\right]$$

and $\partial l_t^{qml}/\partial\theta = W_t K_N F_t$, where $K_N = [0_{N^2 \times N} \quad D_N]$. We have

$$\mathcal{I} = \mathrm{E}[W_t K_N F_t F_t' K_N' W_t'] = \mathrm{E}[W_t K_N M_{FF} K_N' W_t']$$

and

$$\mathcal{J} = -\int_{\mathbb{R}^N} g(x) \frac{\partial^2 l_t^{qml}}{\partial\theta \partial\theta'} dx = \int_{\mathbb{R}^N} \frac{\partial g(x)}{\partial\theta} \frac{\partial l_t^{qml}}{\partial\theta'} dx \tag{C.2}$$

$$= \int_{\mathbb{R}^N} g(x) \frac{\partial \log g(x)}{\partial\theta} \frac{\partial l_t^{qml}}{\partial\theta'} dx = \mathrm{E}\left[\frac{\partial \log g(x)}{\partial\theta} \frac{\partial l_t^{qml}}{\partial\theta'}\right] \tag{C.3}$$

$$= \mathrm{E}[W_t \psi_t F_t K_N' W_t'] = \mathrm{E}[W_t M_{\psi F} K_N' W_t']. \tag{C.4}$$

■

Proof of Proposition 5. (a) $M_{\psi F}$: Note first that ψ_t is orthogonal to v_t: $\mathrm{E}[\psi_t v_t'] = -\mathrm{E}[\mathrm{vec}\{(\partial \log g(v_t)/\partial v_t) v_t'\} v_t']$ and writing this expectation elementwise one obtains $-\int_{\mathbb{R}^N} (\partial g(v_t)/\partial v_{it}) v_{jt} v_{kt} \, dv_t$. Using integration by parts, for $i = j = k$ this is just equal to

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

$2\mathrm{E}[v_{it}] = 0$, and for $i = j \neq k$ it is equal to $\mathrm{E}[v_{kt}] = 0$, and the same holds for $i \neq j = k$ and $i \neq j = k$. Hence, the left block of $M_{\psi F}$ is equal to $0_{N^2 \times N}$.

The right block can be written as

$$-\mathrm{E}\left[\mathrm{vec}\left(\frac{\partial \log g(v_t)}{\partial v_t} v_t'\right)\mathrm{vech}(v_t v_t')'\right] - \mathrm{vec}(I_N)\mathrm{vech}(I_N)'. \tag{C.5}$$

Writing the expectation term in (C.5) elementwise, one obtains $-\int_{\mathbb{R}^N}(\partial g(v_t)/\partial v_{it})$ $v_{jt}v_{kt}v_{lt}\,dv_t$, which for $i = j = k = l$ is equal to 3, for $i = j \neq k = l$ is equal to 1, and zero otherwise. Therefore, we have the following symmetry relation:

$$\int_{\mathbb{R}^N} \frac{\partial g(v_t)}{\partial v_{it}} v_{jt}v_{kt}v_{lt}\,dv_t = \int_{\mathbb{R}^N} \frac{\partial g(v_t)}{\partial v_{jt}} v_{it}v_{kt}v_{lt}\,dv_t,$$

and, as a consequence, the first term in (C.5) can be written as $D_N J_t$ where

$$J_t = -\mathrm{E}\left[\mathrm{vech}\left(\frac{\partial \log g}{v_t} v_t'\right)\mathrm{vech}(v_t v_t')'\right]$$

is, because of the preceding elementwise calculations, an $(N^* \times N^*)$ matrix with 3 at the (i,i)th position, $i \in \mathcal{K}_{ii}^N$; a 1 at the positions (i,j), $i,j \in \mathcal{K}_{ii}^N, i \neq j$; a 1 at the positions (i,i), $i \in \mathcal{K}_{ij}^N$; and zeros elsewhere. Thus, using Lemmas 1 and 2, $J_t = \mathrm{vech}(I_N)$ $\mathrm{vech}(I_N)' + 2D_N^+ D_N^{+\prime}$. Rearranging and multiplying from the left by D_N we obtain $D_N J_t - \mathrm{vec}(I_N)\mathrm{vech}(I_N)' = 2D_N D_N^+ D_N^{+\prime}$, the stated expression for the right block of $M_{\psi F}$ in (C.5).

(b) $g \in \mathcal{D}_s$: To derive M_{FF} for spherical distributions, note that v_t is orthogonal to $\mathrm{vech}(v_t v_t' - I_N)$ because of Lemma 3 (moments containing odd orders are zero). Thus, M_{FF} is block-diagonal. The upper left block of M_{FF} is just $\mathrm{E}[v_t v_t'] = I_N$ by Assumption 1. The lower right block of M_{FF} can be written as

$$\mathrm{E}[\mathrm{vech}(v_t v_t')\mathrm{vech}(v_t v_t')'] - \mathrm{vech}(I_N)\mathrm{vech}(I_N)'. \tag{C.6}$$

Using Lemma 3, the (i,i)th element, $i \in \mathcal{K}_{ii}^N$, of the first term in (C.6) is equal to $3c$; the (i,j)th element, $i,j \in \mathcal{K}_{ii}^N$, $i \neq j$, is equal to c; the (i,i)th element, $i \in \mathcal{K}_{ij}^N$, is equal to c; and all other elements are zero. Together with Lemma 2, this implies that the (i,i)th element, $i \in \mathcal{K}_{ii}^N$ of (C.6), is $3c - 1$; the (i,j)th element, $i,j \in \mathcal{K}_{ii}^N$, $i \neq j$, is $c - 1$; and the (i,i)th element, $i \in \mathcal{K}_{ij}^N$, is c. The stated formula is then immediately obtained by applying Lemmas 1 and 2.

Next, we derive $M_{\psi\psi}$ for the case $g \in \mathcal{D}_s$. The idea is to show that $D_N^+ M_{\psi\psi} D_N^{+\prime}$ has the same structure as the lower right block of M_{FF}. We have

$$D_N^+ M_{\psi\psi} D_N^{+\prime} = \mathrm{E}\left[\mathrm{vech}\left(\frac{\partial \log g(x)}{\partial x} x'\right)\mathrm{vech}\left(\frac{\partial \log g(x)}{\partial x} x'\right)'\right] - \mathrm{vech}(I_N)\mathrm{vech}(I_N)'.$$

A typical element of the first term can be written as

$$\mathrm{E}\left[\frac{\partial \log g(x)}{\partial x_i}\frac{\partial \log g(x)}{\partial x_j} x_k x_l\right] = 4\mathrm{E}\left[f^{-2}(x'x)\left(\frac{\partial f(x'x)}{\partial x'x}\right)^2 x_i x_j x_k x_l\right]$$

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

because $g(x) = f(x'x)$, which is equal to

$$4\int_{\mathbb{R}^N} f^{-1}(x'x)\left(\frac{\partial f(x'x)}{\partial x'x}\right)^2 x_i x_j x_k x_l\, dx. \tag{C.7}$$

Now the function $h(x'x) = 4f^{-1}(x'x)(\partial f(x'x)/\partial x'x)^2$ depends on x only through $x'x$ and is positive and integrable by Assumption 1. Thus, it is itself a spherical density up to some scale, and (C.7) is just the fourth-order moment structure with respect to h. Therefore, Lemma 3 applies to h, and we obtain the same structure as for the lower right block of M_{FF}. That is, for $i \neq j$,

$$\mathrm{E}\left[\left(\frac{\partial \log g(x)}{\partial x_i}\right)^2 x_i^2\right] = 3\mathrm{E}\left[\left(\frac{\partial \log g(x)}{\partial x_i}\right)^2 x_j^2\right]$$

and

$$\mathrm{E}\left[\frac{\partial \log g(x)}{\partial x_i}\,\frac{\partial \log g(x)}{\partial x_j}\,x_i x_j\right] = \mathrm{E}\left[\left(\frac{\partial \log g(x)}{\partial x_i}\right)^2 x_j^2\right]. \tag{C.8}$$

Thus, $D_N^+ M_{\psi\psi} D_N^{+\prime} = 2\tau D_N^+ D_N^{+\prime} + (\tau - 1)\mathrm{vech}(I_N)\mathrm{vech}(I_N)'$, and, because $D_N D_N^+$ is symmetric and idempotent, we obtain $M_{\psi\psi} = 2\tau D_N D_N^+ + (\tau - 1)\mathrm{vec}(I_N)\mathrm{vec}(I_N)'$.

(c) $g \in \mathcal{D}_i$: Because of the independence and symmetry, M_{FF} is diagonal where the first N diagonal elements are 1 because $\mathrm{E}[v_{it}^2] = 1$, $i = 1, \ldots, N$ by Assumption 1. The next N^* diagonal elements are either equal to $\kappa - 1$, corresponding to $\mathrm{E}[v_{it}^4] - 1$, or equal to 1, corresponding to $\mathrm{E}[v_{it}^2 v_{jt}^2]$, $i \neq j$. The formula given in the statement is easily seen to fulfill this requirement.

Finally, the given formula for $M_{\psi\psi}$ can easily be checked by noting that a typical element can be written as

$$\mathrm{E}\left[\frac{\partial \log g(x)}{\partial x_i}\,\frac{\partial \log g(x)}{\partial x_j}\,x_k x_l\right] - I\{(i = k) \wedge (j = l)\},$$

where $I(\cdot)$ denotes the indicator function. For $i = j = k = l$, this is equal to $\tau_1 - 1$ by definition. For $i = k \neq j = l$, we have

$$\mathrm{E}\left[\frac{\partial \log g(x)}{\partial x_i}\,x_i\right]\mathrm{E}\left[\frac{\partial \log g(x)}{\partial x_j}\,x_j\right] - 1 = 1 - 1 = 0$$

and for $i = j \neq k = l$, because of independence and $\mathrm{E}[x_j^2] = 1$,

$$\mathrm{E}\left[\left(\frac{\partial \log g(x)}{\partial x_i}\right)^2 x_j^2\right] = \mathrm{E}\left[\left(\frac{\partial \log g(x)}{\partial x_i}\right)^2\right],$$

which is equal to τ_2 by definition. All off-diagonal elements of $M_{\psi\psi}$ are zero, and the provided formula is easily seen to hold. ■

Proof of Proposition 6. Full parametric efficiency is possible if and only if $P_t = 0$. As the case $\mathrm{E}[W_t] = 0$ can be excluded in general, this is equivalent to $\psi_t = AF_t$, with constant matrix $A = M_{\psi F} M_{FF}^{-1}$. This is a system of partial differential equations where

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

$(\partial g(x)/\partial x)x'/g(x)$ is affine in linear and quadratic terms in x. In the case $N \geq 2$, the unique solution is the multivariate normal distribution. ■

Proof of Proposition 7. It suffices to show that our estimator is asymptotically linear in the efficient influence function. Denote $J_n^* = J^*(\theta_n)$, $\tilde{J}_n^* = J^*(\tilde{\theta}_n)$, and $\hat{J}_n^* = \hat{J}^*(\tilde{\theta}_n)$. We have to show that

$$\sqrt{n}(\hat{\theta}_n - \theta_n) - n^{-1/2} J_n^{*-1} \sum_{t=1}^{n} l_t^*(\theta_n) = \sum_{i=1}^{6} T_i = o_p(1),$$

where, denoting by $\tilde{\mathrm{E}}$ expectation under $P_{\tilde{\theta}_n}$

$$T_1 = \sqrt{n}(\tilde{\theta}_n - \theta_n) + \tilde{J}_n^{*-1} n^{-1/2} \sum_{t=1}^{n} l_t^*(\tilde{\theta}_n) - J_n^{*-1} n^{-1/2} \sum_{t=1}^{n} l_t^*(\theta_n),$$

$$T_2 = n^{-1/2}(\hat{J}_n^{*-1} - \tilde{J}_n^{*-1}) \sum_{t=1}^{n} l_t^*(\tilde{\theta}_n),$$

$$T_3 = n^{-1/2} \hat{J}_n^{*-1} \sum_{t=1}^{\alpha_n} W_t(\tilde{\theta}_n)(\hat{\psi}_{2t} - \tilde{\psi}_t - \tilde{\mathrm{E}}[\hat{\psi}_{2t} | \tilde{v}_{\alpha_n+1}, \ldots, \tilde{v}_n]),$$

$$T_4 = -n^{-1/2} \hat{J}_n^{*-1} \left(\frac{1}{\alpha_n} \sum_{s=1}^{\alpha_n} W_s(\tilde{\theta}_n) \right) \sum_{t=1}^{\alpha_n} (\hat{\psi}_{2t} - \tilde{\psi}_t - \tilde{\mathrm{E}}[\hat{\psi}_{2t} | \tilde{v}_{\alpha_n+1}, \ldots, \tilde{v}_n]),$$

$$T_5 = n^{-1/2} \hat{J}_n^{*-1} \sum_{t=\alpha_n+1}^{n} W_t(\tilde{\theta}_n)(\hat{\psi}_{1t} - \tilde{\psi}_t - \tilde{\mathrm{E}}[\hat{\psi}_{1t} | \tilde{v}_1, \ldots, \tilde{v}_{\alpha_n}]),$$

$$T_6 = -n^{-1/2} \hat{J}_n^{*-1} \left(\frac{1}{n - \alpha_n} \sum_{s=\alpha_n+1}^{n} W_s(\tilde{\theta}_n) \right) \sum_{t=\alpha_n+1}^{n} (\hat{\psi}_{1t} - \tilde{\psi}_t - \tilde{\mathrm{E}}[\hat{\psi}_{1t} | \tilde{v}_1, \ldots, \tilde{v}_{\alpha_n}]).$$

The projection $P_t(\theta)$ in (13) is continuous in θ so that the smoothness of the influence function $J^{-1}(\theta) n^{-1} \sum_{t=1}^{n} \dot{l}_t(\theta)$ as implied by LAN (Proposition 2) carries over to the efficient influence function $J^{*-1}(\theta) n^{-1} \sum_{t=1}^{n} l_t^*(\theta)$. Hence, $T_1 \to_{P_{\theta_n}} 0$. The same holds for T_2 by Lemma 4. For the remaining terms to be $o_{P_{\theta_n}}(1)$ it suffices to show that condition (3.1) of Drost et al. (1997) holds, i.e.,

$$\mathrm{E}_{\theta_n}[\|\hat{\psi}_t - \psi_t\|^2 | v_1 \ldots, v_n] \to_{P_{\theta_n}} 0, \tag{C.9}$$

where $v_1, \ldots, v_n$ are i.i.d. with density g. The proof of (C.9) is similar to that of Theorem 5 of Linton (1993) except that his Lemmas 5.1 and 5.2 are replaced by our Lemma 5. ■

Proof of Proposition 8. The proof follows directly from Theorem 3.1 of Drost et al. (1997).

Proof of Proposition 9. First note that $P_t^s \in \mathcal{T}_s$ because $\mathrm{E}[P_t^s] = 0$, $\mathrm{E}[P_t^s F_t'] = 0$, and $\mathrm{E}[P_t^s P_t^{s\prime}]$ is finite because $\mathrm{E}[\psi_t^s \psi_t^{s\prime}] \leq \mathrm{E}[\mathrm{E}[\psi_t \psi_t' | w_t]] = \mathrm{E}[\psi_t \psi_t'] < \infty$ by Assumption 2 and $\mathrm{E}[F_t^s F_t^{s\prime}] \leq \mathrm{E}[\mathrm{E}[F_t F_t' | w_t]] = \mathrm{E}[F_t F_t'] < \infty$ by Assumption 1 (finite fourth moments).

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

Next, we show that the orthogonal complement of the projection is orthogonal to $\mathcal{T}_s$. It can be written as

$$\dot{l}_t^*(\theta) = \dot{l}_t(\theta) - P_t^s(\theta) = (W_t - \mathrm{E}[W_t])\psi_t^s + \mathrm{E}[W_t]M_{\psi^s F}M_{F^s F}^{-1}F_t^s. \tag{C.10}$$

The first term on the right-hand side of (C.10) is orthogonal to $\mathcal{T}_s$ because $(W_t - \mathrm{E}[W_t])$ has mean zero and is independent of w_t and, hence, independent of all elements in $\mathcal{T}_s$. The second term on the right-hand side of (C.10) is orthogonal to $\mathcal{T}_s$ because $\forall s(w_t) \in \mathcal{T}_s$, $\mathrm{E}[\mathrm{E}[F_t|w_t]s(w_t)] = \mathrm{E}[\mathrm{E}[F_t s(w_t)|w_t]] = \mathrm{E}[F_t s(w_t)]$ by the law of iterated expectations. However, by definition, F_t is orthogonal to $\mathcal{T}_s$. This proves that $\dot{l}_t^*(\theta)$ is orthogonal to $\mathcal{T}_s$. ■

Proof of Proposition 10. If x has a spherical density $g(x) = f(y)$, $y = x'x$, then the density of y is given by

$$h(y) = \frac{\pi^{N/2}}{\Gamma(N/2)} y^{N/2-1} f(y); \tag{C.11}$$

see Fang, Kotz, and Ng (1990, p. 36). Thus, a consistent estimator of $g(\cdot)$ is obtained by estimating $h(\cdot)$ and transforming according to (C.11).

The motivation for the estimator of ψ_t^s is

$$\mathrm{E}\left[\frac{\partial \log g(v_t)}{\partial v_t} v_t' \,\middle|\, w_t\right] = \mathrm{E}\left[2\frac{\partial \log f(w_t)}{\partial w_t} v_t v_t' \,\middle|\, w_t\right] = 2\frac{\partial \log f(w_t)}{\partial w_t} m(w_t),$$

where $m(w_t) = \mathrm{E}[v_t v_t'|w_t]$ can be calculated with arbitrary precision; see Brown and Hodgson (2004). Note that $|\mathrm{vec}(m(w_t))| = w_t$. The remainder of the proof is analogous to that of Proposition 7, except that the trimming conditions change. The gamma kernel estimator has variance $\mathrm{Var}(n^{-1}\sum_t G_{y/b+1,b}(w_t)) = O(n^{-1}b^{-1/2}y^{-1/2})$ for $y/b \to \infty$; see Chen (2000). Thus, the variance of $\hat{g}(x)$ is of order $\mathrm{Var}(\hat{g}(x)) = O((x'x)^{3/2-N}n^{-1}b^{-1/2})$. By Lemma 7, $\mathrm{Var}(n^{-1}\sum_t \partial G_{y/b+1,b}(w_t)/\partial y) = O(n^{-1}b^{1/2}y^{-3/2})$ for $y/b \to \infty$. Now, letting $G_t = G_{y/b+1,b}(w_t)$ and $C_N = \Gamma(N/2)/\pi^{N/2}$,

$$\frac{\partial \hat{g}_b(x)}{\partial x} = C_N(x'x)^{-N/2}\frac{1}{n}\sum_{t=1}^{n}\left(G_t + (x'x)\frac{\partial G_t}{\partial y}\right)x$$

with euclidean norm

$$\left\|\frac{\partial \hat{g}_b(x)}{\partial x}\right\| = C_N(x'x)^{1/2-N/2}\frac{1}{n}\sum_{t=1}^{n}\left|G_t + (x'x)\frac{\partial G_t}{\partial y}\right|$$

$$\leq C_N(x'x)^{1/2-N/2}\frac{1}{n}\sum_{t=1}^{n}\left(G_t + (x'x)\left|\frac{\partial G_t}{\partial y}\right|\right).$$

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press

Noting that $|\mathrm{Cov}(G_t, \partial G_t/\partial y)| \leq \max(\mathrm{Var}(G_t), \mathrm{Var}(\partial G_t/\partial y))$, we obtain

$$\mathrm{Var}\left\| \frac{\partial \hat{g}_b(x)}{\partial x} \right\| \leq C_N^2 y^{1-N} n^{-1} \{\mathrm{Var}(G_t) + y^2 \,\mathrm{Var}(\partial G_t/\partial y) + 2y \,\mathrm{Cov}(G_t, \partial G_t/\partial y)\}$$

$$= O\{n^{-1} y^{1/2-N} b^{-1/2} + n^{-1} y^{3/2-N} b^{1/2}\}, \quad y = x'x,$$

which is $o(1)$ under the given trimming conditions. This enables us to show that (C.9) holds, following the lines of argument of Bickel (1982), Kreiss (1987), and Linton (1993). ■

https://doi.org/10.1017/S0266466607070119 Published online by Cambridge University Press