Semiparametrically Efficient R-Estimation for Dynamic Location-Scale Models

Marc Hallin
ECARES, Université Libre de Bruxelles and ORFE, Princeton University

Davide La Vecchia
Monash University and University of Sankt Gallen

ECARES working paper 2014-45

ECARES
ULB - CP 114/04
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

Semiparametrically Efficient R-Estimation for Dynamic Location-Scale Models

Marc Hallin* Davide La Vecchia†

Abstract

We define rank-based estimators (R-estimators) for semiparametric time series models in which the conditional location and scale depend on a Euclidean parameter, while the innovation density is an infinite-dimensional nuisance. Applications include linear and nonlinear models, featuring either homo- or heteroskedasticity (e.g., AR-ARCH and discretely observed diffusions with jumps). We show how to construct easy-to-implement R-estimators, which achieve semiparametric efficiency at some predetermined reference density while preserving root-n consistency, irrespective of the actual density. Numerical examples illustrate the good performances of the proposed estimators. An empirical analysis of the log-return and log-transformed two-scale realized volatility concludes the paper.

MSC: 62G20, 62G35, 62M05, 62M10

Keywords: Conditional heteroskedasticity, Distribution-freeness, Forecasting, Lévy processes, One-step R-estimators

1 Introduction

Stochastic processes are essential tools of analysis in many scientific fields, such as biology, economics and finance; see, e.g., Kloeden and Platen (1994) and Cont and Tankov (2004). While their probabilistic properties have been studied in great details via stochastic calculus (see, e.g. Karlin and Taylor (1981), Davidson (1994)), their statistical analysis still present several challenges; see, e.g., Aït-Sahalia (2006), Zhao (2008), Bibby et al. (2010), and references therein.

The fact that underlying conditional densities, as a rule, either are unknown or cannot be specified in closed-form led to inference procedures relying on semiparametric methods, among

*ECARES, Université libre de Bruxelles, Belgium and ORFE, Princeton University, Princeton, N.J., USA.

†Dept. of Econometrics and Business Statistics, Monash University, Melbourne, Australia, and School of Economics and Political Science, University of Sankt Gallen, Switzerland.

which the Gaussian quasi-likelihood (henceforth, QL) procedures are the most popular. However, it is well-known that QL estimates for the analysis of real data may be quite inefficient (Linton (1993)), or highly unreliable (Mancini et al. (2005)), because of heavy tails and extreme observations.

To cope with this problem, other semiparametric methods can be applied, along the lines of Drost et al. (1997), to construct root-n consistent and *semiparametrically efficient* estimators. These methods achieve semiparametric efficiency. Efficiency must be understood à la Le Cam (see Le Cam and Yang (2000)), in a local and asymptotic sense, over some class $\mathcal{G}$ of densities.

Typical examples are Linton (1993) and Wefelmeyer (1996), who specify the first two conditional moments of the process under study, while the innovation density remains a nuisance parameter. Then, local asymptotic normality (LAN) leads to the definition of a *semiparametrically efficient score* or *central sequence*, which is obtained as the projection of the *LAN central sequence* along the *tangent space* generated by the nuisance parameter. However, derivation of tangent space projections generally entails complicated calculation. Moreover, in order to achieve root-n consistency, those methods require the estimation of the actual innovation density (thus very large samples are needed) and imposes such numerically cumbersome procedure a sample-splitting.

Building on the results by Hallin and Werker (2003), we propose an alternative estimation methodology, based on the *residual ranks*. Our procedure is similar in the spirit to the standard semiparametric approach, but its implementation avoids both the derivation of tangent space projections and the estimation of the actual innovation density. Thanks to this, we can construct semiparametrically efficient R-estimators in complex econometric models for which the construction of semiparametrically efficient estimators otherwise would be hardly possible. On the other hand, our R-estimators represent a substantial contribution to the literature on rank-based methods, which so far has been focused mainly on statistical procedures for conditionally homoskedastic (and quite often linear) models. For instance, our method, to the best of our knowledge, is the first one deriving R-estimators in the context of discretely observed diffusion processes.

Other attempts have been made to introduce R-estimation in the context of time-series models: see, among the others, Koul and Saleh (1993), Koul and Ossiander (1994), Terpstra et al. (2001), Mukherjee and Bai (2002), Mukherjee (2007), Andrews (2008, 2012). The estimators developed there, however, rely on an extension of the method introduced by Jaeckel (1972) for linear regression with independent observations. Contrary to the original Hodges-Lehmann definition, Jaeckel's R-estimators are based on somewhat hybrid objective functions which combine the residual ranks and residuals themselves. In the time-series settings considered in this paper, Jaeckel-type objec-

tive functions do not follow from any solid decision-theoretic argument and their equivalence to the Hodges-Lehmann approach has to be shown. Jaeckel-type R-estimators are not measurable (not even asymptotically) with respect to residual ranks, and their connection to invariance and semiparametric efficiency is lost.

Our procedure, essentially, is described in three steps: (i) we specify the form of the conditional location and scale of the stochastic process, (ii) we fix a reference density (f, say) which, however, does not need to coincide with the actual one, and (iii) we derive the semiparametrically efficient (at f) score by a suitable projection of the LAN central sequence associated with f onto the σ-algebra generated by the *ranks* of the innovations. Semiparametric efficiency then is attained at f while root-n consistency is preserved under densities $g \neq f$: R-estimators thus are resistant to the misspecification of the innovation density and, in that sense, feature global robustness. Moreover, considerable improvements in efficiency can be achieved over the QL method; see Hallin (1994).

Several examples illustrate the advantages of our estimation procedures. Numerical analysis in the context of normal variance-mean mixture models (see Barndorff-Nielsen et al. (1982)) for the return dynamics and a Monte Carlo experiment for the Value-at-Risk (VaR) forecasting in an AR(1)-ARCH(1) model provide evidence that rank-based procedures improve on the accuracy of estimates and forecasts obtained via traditional QL methods. Finally, an empirical analysis of the modeling and forecasting of the log-return and log-transformed realized volatility of the USD/CHF exchange rate confirms the increased accuracy of rank-based forecasts compared to the QL ones.

The paper is organized as follows. In Sections 2 and 3, we describe the setting and show how to derive rank-based, hence distribution-free, versions of *semiparametrically efficient central sequences*. Those *rank-based central sequences* constitute the main tool in the construction of our R-estimators, and in Section 4 we explain how they can be constructed in a variety of econometric models. In Section 5, we show how R-estimators can be implemented by applying to those *rank-based central sequences* the ideas of Le Cam's one-step estimation method and using a consistent estimator of the cross-information matrix. In Section 6, we present some numerical illustrations. In Section 7, we apply our method to real data. Finally, in Section 8 we conclude and mention some possible topics for future research. Proofs and technical details are concentrated in Appendix.

2 Model setting and main assumptions

2.1 General setting

Let $\mathbf{Y}^{(n)} := (Y_{-q+1}, \ldots, Y_0, Y_1, \ldots, Y_t, \ldots, Y_n)$ be the finite realization of some stationary real-valued discrete-time process $\{Y_t; t \in \mathbb{Z}\}$ satisfying

$$Y_t = m(\mathbf{Y}_{t-1}, \boldsymbol{\theta}) + v(\mathbf{Y}_{t-1}, \boldsymbol{\theta})\varepsilon_t \tag{2.1}$$

with $\mathbf{Y}_{t-1} := (Y_{t-1}, \ldots, Y_{t-q})$. The functions $\boldsymbol{y} \mapsto m(\boldsymbol{y}, \boldsymbol{\theta})$ and $\boldsymbol{y} \mapsto v(\boldsymbol{y}, \boldsymbol{\theta})$, $\boldsymbol{y} \in \mathbb{R}^q$, are specified and depend on some unknown p-dimensional real parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$; $\{\varepsilon_t; t \in \mathbb{Z}\}$ is an independent and identically distributed (i.i.d.) process with unspecified standardized density g, such that ε_t and $Y_{t'}$ are mutually independent for all $t' > t$.

The interpretation of $m(\boldsymbol{y}, \boldsymbol{\theta})$ and $v(\boldsymbol{y}, \boldsymbol{\theta})$ depends on the way g is standardized. If g is assumed to have mean zero and variance one, then $m(\boldsymbol{y}, \boldsymbol{\theta})$ is the mean, and $v(\boldsymbol{y}, \boldsymbol{\theta})$ the standard error, of Y_t conditional on $\mathbf{Y}_{t-1} = \boldsymbol{y}$; this is the traditional specification, in which g is required to have finite moments of order two. Moment assumptions can be avoided, however, if g, for instance, is assumed to have median zero and interquartile range one; then $m(\boldsymbol{y}, \boldsymbol{\theta})$ is the median, and $v(\boldsymbol{y}, \boldsymbol{\theta})$ the interquartile range, of Y_t conditional on $\mathbf{Y}_{t-1} = \boldsymbol{y}$. With obvious notation, model equation (2.1) generalizes into $Y_t = m(\mathbf{Y}_{t-1}, \mathbf{X}_t, \boldsymbol{\theta}) + v(\mathbf{Y}_{t-1}, \mathbf{X}_t, \boldsymbol{\theta})\varepsilon_t$ in order to accommodate the presence of exogenous covariates $\mathbf{X}_t$. For the sake of notational simplicity, we do not pursue with this, but all results below straightforwardly extend to that case.

For fixed g and $\boldsymbol{\theta}$ in (2.1), we denote by $\mathrm{P}_{\boldsymbol{\theta},g}$ the q-dimensional marginal of the process. Equation (2.1) yields as particular cases a number of classical models studied in the literature.

Example 1. *Discrete-time AR-ARCH models.* Setting

$$m(\boldsymbol{y}, \boldsymbol{\theta}) = \sum_{i=1}^{p} \theta_i y_{t-i} \quad \text{and} \quad v^2(\boldsymbol{y}, \boldsymbol{\theta}) = 1 + \sum_{i=i}^{q} \theta_i y_{t-i}^2$$

for $\boldsymbol{y} = (y_1, \ldots, y_{\max(p,q)}) \in \mathbb{R}^{\max(p,q)}$ yields the AR(p)-ARCH(q) processes; traditional AR(p) and ARCH(q) are particular cases.

Example 2. *Discretely sampled continuous-time Markov processes.* Assume that the dynamics of the continuous-time Markov process $\mathcal{Y}$ are described by the stochastic differential equation

$$dY_s = \mu(Y_s, \boldsymbol{\theta})ds + \sigma(Y_s, \boldsymbol{\theta})d\mathcal{W}_s$$

with drift $\mu(y, \boldsymbol{\theta})$ and diffusion coefficient $\sigma(y, \boldsymbol{\theta})$. Suppose we observe equispaced discrete-time values from that process. That is, we observe $(Y_0, Y_h, ..., Y_{jh}, ..., Y_{nh})$, with fixed h-frequency (e.g., daily), for some $h > 0$. Maximum likelihood estimation of $\boldsymbol{\theta}$ can be too rigid or even inapplicable, since for many diffusions the likelihood is not available in closed form. To cope with this issue, semiparametric versions of discretely sampled diffusions and Lévy processes, of the form (2.1), can be obtained by exploiting the knowledge, in closed form, of the first two conditional moments. This is the case, for instance, for the Ornstein-Uhlenbeck process considered in Aït-Sahalia (2002), the semiparametric version of which is studied in Hallin et al. (2000).

2.2 Assumptions

We propose a method for conducting inference about $\boldsymbol{\theta}$ while avoiding a complete parametric specification of the data-generating mechanism described in (2.1). Throughout, we assume that the functions m and v are specified, but not the parameter $\boldsymbol{\theta}$ nor the density g; the resulting statistical models thus are semiparametric ones, with Euclidean parameter of interest $\boldsymbol{\theta}$, and infinite-dimensional nuisance g. Denoting by $\mathrm{P}^{(n)}_{\boldsymbol{\theta},g}$ the joint distribution, under (2.1), of $\mathbf{Y}^{(n)}$, consider the semiparametric models $\mathcal{P}^{(n)} = \{\mathrm{P}^{(n)}_{\boldsymbol{\theta},g} : \boldsymbol{\theta} \in \boldsymbol{\Theta}, g \in \mathcal{G}\}$, $n \in \mathbb{N}$, where $\boldsymbol{\Theta}$ and $\mathcal{G}$ are such that the following assumptions (Assumptions (A) and (B), but also Assumptions (C)-(D), see Section 3.1) hold for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $g \in \mathcal{G}$.

Assumption (A). The functions $\boldsymbol{\theta} \mapsto m(\boldsymbol{y}, \boldsymbol{\theta})$ and $\boldsymbol{\theta} \mapsto v(\boldsymbol{y}, \boldsymbol{\theta})$ are differentiable for all $\boldsymbol{y}$, with gradients $\dot{\boldsymbol{m}}(\boldsymbol{y}, \boldsymbol{\theta}) := \mathrm{grad}_{\boldsymbol{\theta}}\, m(\boldsymbol{y}, \boldsymbol{\theta})$ and $\dot{\boldsymbol{v}}(\boldsymbol{y}, \boldsymbol{\theta}) := \mathrm{grad}_{\boldsymbol{\theta}}\, v(\boldsymbol{y}, \boldsymbol{\theta})$. Moreover, denoting by $\mathrm{E}_{\boldsymbol{\theta},g}$ expectations under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},g}$, both $\mathrm{E}_{\boldsymbol{\theta},g}[\dot{\boldsymbol{m}}(\mathbf{Y}_{t-1}, \boldsymbol{\theta})]$ and $\mathrm{E}_{\boldsymbol{\theta},g}[\dot{\boldsymbol{v}}(\mathbf{Y}_{t-1}, \boldsymbol{\theta})]$ exist and are finite.

Assumption (B). (B1) For all $x \in \mathbb{R}$, the density $g(x)$ is strictly positive, with location zero and scale one (any choice of location and scale can be adopted here).

(B2) The mapping $x \mapsto g(x)$ is *absolutely continuous on finite intervals*, i.e. there exists an *a.e. derivative* $\dot{g}$ such that, for all $-\infty < a < b < \infty$, $g(a) - g(b) = \int_a^b \dot{g}(x)dx$.

(B3) Letting $\phi_g(x) := -\dot{g}(x)/g(x)$ and $\psi_g(x) := x\phi_g(x) - 1$, the *Fisher information for location* $I_1(g) := \int_{\mathbb{R}} \phi_g^2(x)g(x)dx$ and the *Fisher information for scale* $I_2(g) := \int_{\mathbb{R}} \psi_g^2(x)g(x)dx$ exist and are finite. Cauchy-Schwarz then implies that $I_{12}(g) = I_{21}(g) := \int x\phi_g^2(x)g(x)dx$ exists and is finite.

For given $g \in \mathcal{G}$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, let $\mathcal{H}^{(n)}_{\boldsymbol{\theta},g}$ denote the simple hypothesis $\{\mathrm{P}^{(n)}_{\boldsymbol{\theta},g}\}$ and write $\mathcal{H}^{(n)}_{\boldsymbol{\theta}}$ for the nonparametric collection $\{\mathrm{P}^{(n)}_{\boldsymbol{\theta},g} : g \in \mathcal{G}\}$. Denote by

$$Z_t(\boldsymbol{\theta}) := (Y_t - m(\mathbf{Y}_{t-1}, \boldsymbol{\theta}))/v(\mathbf{Y}_{t-1}, \boldsymbol{\theta}) \tag{2.2}$$

the residuals associated with the parameter value $\boldsymbol{\theta}$. Clearly, the hypotheses $\mathcal{H}_{\boldsymbol{\theta}}^{(n)}$ and $\mathcal{H}_{\boldsymbol{\theta},g}^{(n)}$ hold true iff the residuals $Z_t(\boldsymbol{\theta})$ are i.i.d. and iff they are i.i.d. with density g, respectively.

3 Uniform local asymptotic normality and ranks

In this section, we introduce the main methodological tools to be used in the sequel. First, we establish the uniform local asymptotic normality (ULAN), with central sequence $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g)$, of the parametric fixed-g submodels $\mathcal{P}_g^{(n)} := \{\mathrm{P}_{\boldsymbol{\theta},g}^{(n)} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$. Then, following Hallin and Werker (2003), we project $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g)$ onto the σ-algebra generated by the ranks of the residuals $Z_t(\boldsymbol{\theta})$.

3.1 Uniform local asymptotic normality (ULAN)

Defining

$$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g) := n^{-1/2} \sum_{t=1}^{n} \dot{\boldsymbol{l}}(Z_t, \mathbf{Z}_{t-1}, \boldsymbol{\theta}, g) \quad \text{and} \quad \boldsymbol{\Gamma}(\boldsymbol{\theta}, g) := \mathrm{E}_{\boldsymbol{\theta},g}\Big[\dot{\boldsymbol{l}}(Z_t, \mathbf{Z}_{t-1}, \boldsymbol{\theta}, g)\dot{\boldsymbol{l}}'(Z_t, \mathbf{Z}_{t-1}, \boldsymbol{\theta}, g)\Big], \tag{3.1}$$

where

$$\dot{\boldsymbol{l}}(Z_t, \mathbf{Z}_{t-1}, \boldsymbol{\theta}, g) := \frac{\dot{\boldsymbol{v}}(\mathbf{Y}_{t-1}, \boldsymbol{\theta})}{v(\mathbf{Y}_{t-1}, \boldsymbol{\theta})}\psi_g(Z_t(\boldsymbol{\theta})) - \frac{\dot{\boldsymbol{m}}(\mathbf{Y}_{t-1}, \boldsymbol{\theta})}{v(\mathbf{Y}_{t-1}, \boldsymbol{\theta})}\phi_g(Z_t(\boldsymbol{\theta})), \tag{3.2}$$

we make the additional assumption

Assumption (C). For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $g \in \mathcal{G}$, (i) the matrix $\boldsymbol{\Gamma}(\boldsymbol{\theta}, g)$ exists, is finite and has full rank, and (ii) the mapping $\boldsymbol{\theta} \mapsto \boldsymbol{\Gamma}(\boldsymbol{\theta}, g)$ is continuous.

The following ULAN property then follows.

Proposition 3. *(ULAN) Let Assumptions (A)-(C) hold. For all $g \in \mathcal{G}$, the parametric model $\mathcal{P}_g^{(n)}$ is ULAN with central sequence $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g)$ and information matrix $\boldsymbol{\Gamma}(\boldsymbol{\theta}, g)$. More precisely, we have, for all $g \in \mathcal{G}$, all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, all $\boldsymbol{\theta}^{(n)}$ such that $\boldsymbol{\theta}^{(n)} - \boldsymbol{\theta} = O(n^{-1/2})$, and all bounded sequence $\boldsymbol{\tau}_n \in \mathbb{R}^p$,*

$$\Lambda_n := \log \frac{\mathrm{dP}_{\boldsymbol{\theta}^{(n)}+n^{-1/2}\boldsymbol{\tau}^{(n)},g}^{(n)}}{\mathrm{dP}_{\boldsymbol{\theta}^{(n)},g}^{(n)}} = \boldsymbol{\tau}_n'\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}^{(n)}, g) - \frac{1}{2}\boldsymbol{\tau}_n'\boldsymbol{\Gamma}(\boldsymbol{\theta}, g)\boldsymbol{\tau}_n + o_{\mathrm{P}}(1), \tag{3.3}$$

and $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}^{(n)}, g) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}; \boldsymbol{\Gamma}(\boldsymbol{\theta}, g))$, under $\mathrm{P}_{\boldsymbol{\theta}^{(n)},g}^{(n)}$ as $n \to \infty$.

3.2 Semiparametrically efficient central sequences

In the classical semiparametric approach, the *semiparametrically efficient central sequence* is the tool that one needs to conduct inference and reach semiparametric efficiency bounds. That semiparametrically efficient central sequence, denoted as $\boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, g)$, is obtained by projecting the central sequence $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g)$ along the *tangent space*; see the monograph by Bickel et al. (1993) for the i.i.d. framework, or Hallin and Werker (2003) for time series models.

Typically, the actual computation of semiparametrically efficient central sequences for models of the form (2.1) is a painful case-by-case task. Moreover, once $\boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, f)$ has been obtained for all $f \in \mathcal{G}$, the actual density g still has to be estimated by some adequate $\hat{g}^{(n)}$, to be plugged into $\boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, f)$, yielding $\boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, \hat{g}^{(n)})$. Indeed, the asymptotic distribution of $\boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, f)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},g}$ is unknown whenever $f \neq g$, and, typically $\mathrm{E}_{\boldsymbol{\theta},g}[\boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, f)] \neq 0$. As a consequence, the estimators based on $\boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, f)$ reach the semiparametric efficiency bounds under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},f}$, but are no longer root-n consistent under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},g}$ for $f \neq g$. Finally, additional technical complications like sample-splitting are required in order for $\boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, \hat{g}^{(n)})$ to be asymptotically equivalent, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},g}$, to $\boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, g)$.

It has been shown in Hallin and Werker (2003) that, for a very broad class of models (including most time series models), the invariance properties of residual ranks offer an attractive alternative way to achieve semiparametric efficiency at f. More precisely, projecting $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)$ onto the σ-field generated by the ranks of the residuals $Z_1(\boldsymbol{\theta}), \ldots, Z_n(\boldsymbol{\theta})$ yields a rank-based, hence distribution-free, version $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$, say, of the semiparametrically efficient central sequence $\boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, f)$. Namely, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},f}$,

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f) - \boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, f) = o_{\mathrm{P}}(1), \tag{3.4}$$

so that estimators (tests) based on $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$ reach the semiparametric efficiency bounds under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},f}$. However, the distribution-freeness of ranks ensures that those estimators remain root-n consistent (keep their nominal asymptotic size) under any $\mathrm{P}^{(n)}_{\boldsymbol{\theta},g}$. Thus, they remain valid under the unknown actual density g, the estimation of which is not required.

3.2.1 Theoretical derivation of the rank-based central sequence

Let us provide some details on the rank-based approach just described. Let f be some reference density. Denote by $\boldsymbol{R}^{(n)}(\boldsymbol{\theta})$ the vector $(R^{(n)}_1(\boldsymbol{\theta}), \ldots, R^{(n)}_n(\boldsymbol{\theta}))$ of *residual ranks*, where $R^{(n)}_t(\boldsymbol{\theta})$ is the rank of $Z^{(n)}_t(\boldsymbol{\theta})$ among $Z^{(n)}_1(\boldsymbol{\theta}), \ldots, Z^{(n)}_n(\boldsymbol{\theta})$. For notational convenience, we write $\boldsymbol{R}^{(n)}$, $R^{(n)}_t$

and $Z_t^{(n)}$, or $\boldsymbol{R}(\boldsymbol{\theta})$, $R_t(\boldsymbol{\theta})$ and $Z_t(\boldsymbol{\theta})$, dropping the dependence on $\boldsymbol{\theta}$ and n when no confusion is possible.

To perform the rank-based construction, we first rewrite the central sequence $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)$ as a function of the present and past residuals $Z_t(\boldsymbol{\theta})$ only. To this end, let $\mathbf{Z}_{t-1} := (Z_{t-1}, \ldots, Z_1, \varepsilon_0, \varepsilon_{-1}, \ldots)$. The structure of the dynamic location and scale model (2.1) implies that, for any fixed $\boldsymbol{\theta}$, Y_t is a measurable function of $\mathbf{Z}_t$, with $\mathbf{Z}_{t-1}$-measurable conditional location and scale. This yields, with a slight abuse of notation,

$$\dot{\boldsymbol{l}}(Z_t, \mathbf{Z}_{t-1}, \boldsymbol{\theta}, f) = \frac{\dot{\boldsymbol{v}}(\mathbf{Z}_{t-1}, \boldsymbol{\theta})}{v(\mathbf{Z}_{t-1}, \boldsymbol{\theta})}\psi_f(Z_t(\boldsymbol{\theta})) - \frac{\dot{\boldsymbol{m}}(\mathbf{Z}_{t-1}, \boldsymbol{\theta})}{v(\mathbf{Z}_{t-1}, \boldsymbol{\theta})}\phi_f(Z_t(\boldsymbol{\theta})). \tag{3.5}$$

It may happen that (3.5) only requires a finite number s_n of lagged residuals Z_{t-i} but, quite often, an infinity of them are involved; this is the case, for instance, in AR models. Considering the approximation $\dot{\boldsymbol{l}}^{s_n}(Z_t, \mathbf{Z}_{t-1}^{s_n}, \boldsymbol{\theta}, f)$ of $\dot{\boldsymbol{l}}(Z_t, \mathbf{Z}_t, \boldsymbol{\theta}, f)$ obtained by replacing, in (3.5), $\mathbf{Z}_{t-1}$ with the truncation $\mathbf{Z}_{t-1}^{s_n} := (Z_{t-1}, \ldots, Z_{t-\min(t-1,s_n)}, 0, 0, \ldots)$, we make the following assumption (for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $g \in \mathcal{G}$):

Assumption (D). There exists a sequence s_n of integers such that $n > s_n \uparrow \infty$ as $n \to \infty$, and

$$n^{-1/2}\sum_{t=1}^{n}\left(\dot{\boldsymbol{l}}^{s_n}(Z_t, \mathbf{Z}_{t-1}^{s_n}, \boldsymbol{\theta}, g) - \dot{\boldsymbol{l}}(Z_t, \mathbf{Z}_{t-1}, \boldsymbol{\theta}, g)\right) = o_{\mathrm{qm}}(1) \quad \text{as } n \to \infty, \quad \text{under } \mathrm{P}_{\boldsymbol{\theta},g}^{(n)}, \tag{3.6}$$

where $o_{\mathrm{qm}}(1)$ stands for a sequence that tends to zero in quadratic mean. Assumption (D) is satisfied by many stationary Markov processes of order q and by q-dependent processes; examples are provided in Section 4.

Assumption (D) implies that substituting the truncated scores $\dot{\boldsymbol{l}}^{s_n}(Z_t, \mathbf{Z}_{t-1}^{s_n}, \boldsymbol{\theta}, f)$ for the exact ones $\dot{\boldsymbol{l}}(Z_t, \mathbf{Z}_t, \boldsymbol{\theta}, f)$ in the definition of $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)$ still yields a central sequence (central sequences are only defined up to $o_{\mathrm{P}}(1)$'s), which, for simplicity, we still denote as $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)$. Note that (3.6) implies that the variance (under $\mathrm{P}_{\boldsymbol{\theta},f}^{(n)}$) of $\dot{\boldsymbol{l}}^{s_n}(Z_t, \mathbf{Z}_{t-1}^{s_n}, \boldsymbol{\theta}, f)$ is $O(1)$, hence, in view of the independence between Z_t and $\mathbf{Z}_{t-1}^{s_n}$, that the expectation (under $\mathrm{P}_{\boldsymbol{\theta},f}^{(n)}$) of $\dot{\boldsymbol{l}}^{s_n}(Z_t, \mathbf{Z}_{t-1}^{s_n}, \boldsymbol{\theta}, f)$ remains zero.

Finally, in order to define a rank-based version of $\boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, f)$, we further make the following very mild assumption on the truncated score function $\dot{\boldsymbol{l}}^{s_n}$ associated with the reference density f.

Assumption (E). The mapping $(Z_t, \mathbf{Z}_{t-1}^{s_n}) \mapsto \dot{\boldsymbol{l}}^{s_n}(Z_t, \mathbf{Z}_{t-1}^{s_n}, \boldsymbol{\theta}, f)$ is componentwise monotone in all its arguments, or a linear combinations of such componentwise monotone functions.

Denoting by $\mathcal{B}^{(n)}(\boldsymbol{\theta})$ the sigma-field generated by $\boldsymbol{R}^{(n)}(\boldsymbol{\theta})$, define

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f) := \mathrm{E}_{\boldsymbol{\theta},f}\big[\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)\big|\mathcal{B}^{(n)}(\boldsymbol{\theta})\big] = \sum_{t=1}^{n} \mathrm{E}_{\boldsymbol{\theta},f}\big[\dot{\boldsymbol{l}}^{s_n}(Z_t, \mathbf{Z}_{t-1}^{s_n}, \boldsymbol{\theta}, f)\big|\mathcal{B}^{(n)}(\boldsymbol{\theta})\big]. \tag{3.7}$$

Being $\boldsymbol{R}^{(n)}(\boldsymbol{\theta})$-measurable, $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$ is distribution-free: let $\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}(\boldsymbol{\theta}, f)$ be its covariance matrix under $\mathcal{H}_{\boldsymbol{\theta}}^{(n)}$ (since that matrix only depends on $\boldsymbol{\theta}$). The following result then follows, with minor changes, from Proposition 3.1, Corollary 3.2, and Proposition 3.3 in Hallin and Werker (2003); details are left to the reader.

Proposition 4. *Let Assumptions (A)-(E) be satisfied. Denote by $\boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, f)$ a semiparametrically efficient central sequence for $\mathcal{P}_f^{(n)}$, and by $\boldsymbol{\Gamma}^{*(n)}(\boldsymbol{\theta}, f)$ its covariance matrix under $\mathrm{P}_{\boldsymbol{\theta},f}^{(n)}$. Then,*

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f) - \boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, f) = o_{\mathrm{P}}(1) \quad \textit{under } \mathrm{P}_{\boldsymbol{\theta},f}^{(n)}, \quad \textit{and} \quad \lim_{n\to\infty} \underset{\sim}{\boldsymbol{\Gamma}}^{(n)}(\boldsymbol{\theta}, f) = \lim_{n\to\infty} \boldsymbol{\Gamma}^{*(n)}(\boldsymbol{\theta}, f) =: \boldsymbol{\Gamma}^{*}(\boldsymbol{\theta}, f),$$

where $\boldsymbol{\Gamma}^{*}(\boldsymbol{\theta}, f)$ is the semiparametric information matrix.

The asymptotic equivalence, under $\mathrm{P}_{\boldsymbol{\theta},f}^{(n)}$, of $\boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, f)$ and $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$ implies that the latter can be considered a rank-based version of the same semiparametrically efficient (at f) central sequence.

3.2.2 Computation of the rank-based central sequence

The rank-based score in (3.7) is obtained as the projection of $\dot{\boldsymbol{l}}^{s_n}(Z_t, \mathbf{Z}_{t-1}^{s_n}, \boldsymbol{\theta}, f)$ onto $\mathcal{B}^{(n)}(\boldsymbol{\theta})$. Defining $\boldsymbol{R}_t^{s_n} := (R_t^{(n)}, \ldots, R_{t-s}^{(n)})$, let (recall that the score function is square-integrable)

$$\boldsymbol{a}_f(\boldsymbol{R}_t^{s_n}; \boldsymbol{\theta}) := \mathrm{E}_{\boldsymbol{\theta},f}\left[\dot{\boldsymbol{l}}^{s_n}(Z_t, \mathbf{Z}_{t-1}^{s_n}, \boldsymbol{\theta}, f)|R_t^{(n)}, \ldots, R_{t-s}^{(n)}\right] \tag{3.8}$$

(the *exact scores*). Those exact scores in general do not admit a closed form. However, they can be replaced by the so-called *approximate scores*

$$a_f^{(n)}(\boldsymbol{R}_t^{s_n}; \boldsymbol{\theta}) := \dot{\boldsymbol{l}}^{s_n}\left(F^{-1}\left(\frac{R_t^{(n)}}{n+1}\right), \ldots, F^{-1}\left(\frac{R_{t-s}^{(n)}}{n+1}\right), \boldsymbol{\theta}, f\right), \tag{3.9}$$

where F is the distribution function associated with f; see Hallin and Werker (2003) and reference therein. An intuitive and heuristic justification of Eq. (3.9) is that, when $Z_1, \ldots, Z_n$ are i.i.d., with nonvanishing density f, hence strictly monotone distribution function F, then $R_t^{(n)}/(n+1) \approx F(Z_t)$,

so that $Z_t \approx F^{-1}(R_t^{(n)}/(n+1))$, where $F(Z_1), \ldots, F(Z_n)$ are i.i.d., uniform over $[0,1]$. With approximate scores, the rank-based central sequence takes the form

$$\underset{\sim}{\Delta}^{(n)}(\boldsymbol{\theta}, f) = \frac{1}{\sqrt{n-s}} \sum_{t=s+1}^{n} a_f^{(n)}(\boldsymbol{R}_t^s; \boldsymbol{\theta}) - \boldsymbol{m}_f^{(n)} \tag{3.10}$$

with the re-centering $\boldsymbol{m}_f^{(n)} := \sqrt{n-s}\mathrm{E}\left[\sum_{t=s+1}^{n} a_f^{(n)}(\boldsymbol{R}_t^s; \boldsymbol{\theta})\right]$. By standard U-statistics results (see, e.g., Chapter 12 of van der Vaart (1998)),

$$\boldsymbol{m}_f^{(n)} = \frac{\sqrt{n-s}}{n(n-1)\cdots(n-s)} \sum_{1 \le i_1 \ne \cdots \ne i_s \le n} \cdots \sum a_f^{(n)}\left(\frac{i_1}{n+1}, ..., \frac{i_s}{n+1}; \boldsymbol{\theta}\right) + o_P(1), \tag{3.11}$$

so that $\boldsymbol{m}_f^{(n)}$ indeed qualifies as a centering.

In the sequel, for the sake of simplicity, we call $\underset{\sim}{\Delta}^{(n)}(\boldsymbol{\theta}, f)$ a *rank-based central sequence* instead of a *rank-based semiparametrically efficient central sequence*. Unlike its traditional counterpart, the rank-based central sequence does not require (i) the nontrivial exercise of deriving the tangent spaces and corresponding projections, (ii) being computed at an estimator $\hat{g}^{(n)}$ of the actual density.

4 Examples

4.1 Discrete-time models

4.1.1 Conditional heteroskedasticity models

(a) ARCH(q) Consider the class of models with dynamics of the form

$$Y_t = \left(1 + \sum_{j=1}^{q} \theta_j Y_{t-1}^2\right)^{1/2} \epsilon_t, \tag{4.1}$$

where the ϵ_t's are i.i.d. with standardized (with mean zero and variance one) density g, $\theta_j > 0$ for $j = 1, ..., p$, and $\sum_{j=1}^{q} \theta_j \le \rho$ for some $\rho < 1$. This model is ULAN, with central sequence

$$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\psi_g\left(Z_t(\boldsymbol{\theta})\right)}{1 + \sum_{j=1}^{q} \theta_j Y_{t-j}^2} \begin{pmatrix} Y_{t-1}^2 \\ \vdots \\ Y_{t-q}^2 \end{pmatrix}, \tag{4.2}$$

where $\boldsymbol{\theta} := (\theta_1, \ldots, \theta_q)$ and $Z_t(\boldsymbol{\theta}) := Y_t / \big(1 + \sum_{j=1}^q \theta_j Y_{t-1}^2\big)^{1/2}$. The definition of a rank-based central sequence requires, for every t, $(Y_{t-1}, \ldots, Y_{t-q})$ in (4.2) to be expressed in terms of a finite number of past shocks. In Appendix A.4.1, we show that this is possible via a Volterra series expansion. We here provide an alternative heuristic argument. For $t = 1$ let $Z_1^{(n)}(\boldsymbol{\theta}) = Y_1$, that is, assume (arbitrarily, but this has no impact asymptotically) the unobserved initial values $Y_{-q}, ..., Y_0$ to be equal to zero. This provides the n-tuple $Z_1^{(n)}(\boldsymbol{\theta}), ..., Z_n^{(n)}(\boldsymbol{\theta})$, with ranks $R_1^{(n)}(\boldsymbol{\theta}), ..., R_n^{(n)}(\boldsymbol{\theta})$. Since, by definition, $Y_1 = Z_1^{(n)}(\boldsymbol{\theta})$, we set $\underset{\sim}{Y}_1 := G^{-1}(R_1^{(n)}(\boldsymbol{\theta})/(n+1))$, and start the recurrence

$$\underset{\sim}{Y}_t := \Big(1 + \sum_{j=1}^q \theta_j \underset{\sim}{Y}_{t-j}^2\Big)^{1/2} G^{-1}\left(\frac{R_t^{(n)}(\boldsymbol{\theta})}{n+1}\right), \quad t \geq 2. \tag{4.3}$$

This procedure yields (for reference density f) a central sequence which depends on a finite number of past shocks, and we define

$$\underset{\sim}{\boldsymbol{\Delta}}^{(\boldsymbol{n})}(\boldsymbol{\theta}, f) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\psi_f\Big(F^{-1}\Big(\frac{R_t^{(n)}(\boldsymbol{\theta})}{n+1}\Big)\Big) - m_{(2),f}^{(n)}}{1 + \sum_{j=1}^q \theta_j \underset{\sim}{Y}_{t-j}^2} \begin{pmatrix} \underset{\sim}{Y}_{t-1}^2 \\ \vdots \\ \underset{\sim}{Y}_{t-q}^2 \end{pmatrix} \tag{4.4}$$

where (see example 4.4 of Hallin and Werker (2003))

$$m_{f,(2)}^{(n)} := \frac{1}{n} \sum_{i=1}^n \psi_f\Big(F^{-1}\Big(\frac{i}{n+1}\Big)\Big). \tag{4.5}$$

From the re-centering Lemma in Appendix, it follows that $m_{f,(2)}^{(n)}$ is $o\big(n^{-1/2}\big)$. The asymptotic covariance $\boldsymbol{\Gamma}^*(\boldsymbol{\theta}, f)$ of $\underset{\sim}{\boldsymbol{\Delta}}^{(\boldsymbol{n})}(\boldsymbol{\theta}, f)$ under $\mathcal{H}_{\boldsymbol{\theta}}^{(n)}$ (which is also the semiparametric information matrix under $\mathcal{H}_{\boldsymbol{\theta},f}^{(n)}$) is of the form $I_2(f)\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\theta})$, where the $q \times q$-matrix $\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\theta})$ is the Gaussian information matrix given in Theorem 2.1 of Kristensen and Rahbek (2005).

(b) AR(p)-LARCH(q) Consider the discrete-time bilinear process with dynamics

$$Y_t = \sum_{j=1}^p \vartheta_j Y_{t-j} + \left(1 + \sum_{l=1}^q \beta_l Y_{t-j}\right) \epsilon_t, \tag{4.6}$$

where the ϵ_t's are i.i.d. with standardized density g, and $\boldsymbol{\theta} = (\vartheta_1, ..., \vartheta_q, \beta_1, ..., \beta_q)$. If the conditions of Theorem 2.1 in Francq and Zakoïan (2010) are satisfied (see Appendix A.4.2), the ULAN central

sequence for $\boldsymbol{\theta}$ reads

$$\boldsymbol{\Delta}^{(\boldsymbol{n})}(\boldsymbol{\theta}, g) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(1 + \sum_{l=1}^{q} \beta_l Y_{t-l}\right)^{-1} \begin{pmatrix} \phi_g\left(Z_t(\boldsymbol{\theta})\right) \begin{pmatrix} Y_{t-1} \\ \vdots \\ Y_{t-p} \end{pmatrix} \\ \psi_g\left(Z_t(\boldsymbol{\theta})\right) \begin{pmatrix} Y_{t-1} \\ \vdots \\ Y_{t-q} \end{pmatrix} \end{pmatrix}. \tag{4.7}$$

A rank-based central sequence for reference density f is obtained by replacing, in (4.7), the residual Z_t by $F^{-1}(R_t^{(n)}/(n+1))$, for every t. We illustrate the construction in the AR(1)-LARCH(1) case, with

$$Y_t = \vartheta Y_{t-1} + (1 + \beta Y_{t-1})\epsilon_t, \tag{4.8}$$

which is ULAN with central sequence

$$\boldsymbol{\Delta}^{(\boldsymbol{n})}(\boldsymbol{\theta}, g) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{pmatrix} \phi_g\left(Z_t\right) \\ \psi_g\left(Z_t\right) \end{pmatrix} \frac{Y_{t-1}}{1 + \beta Y_{t-1}}$$

where $Z_t = Z_t(\boldsymbol{\theta}) := (Y_t - \vartheta Y_{t-1})/(1 + \beta Y_{t-1})$. In A.4.2, we show how to derive the rank-based central sequence in two steps: (i) an application of the Volterra series expansion provides a version of the central sequence which depends on a finite number of past shocks, in which (ii) the replacement of the shocks by their ranks yields $\underset{\sim}{\boldsymbol{\Delta}}^{(\boldsymbol{n})}(\boldsymbol{\theta}, f)$. We here provide an alternative heuristic argument similar to the one we gave for the ARCH(q) case. For $t = 1$, arbitrarily put $Z_1^{(n)}(\boldsymbol{\theta}) = Y_1$. This provides the n-tuple $Z_1^{(n)}(\boldsymbol{\theta}), ..., Z_n^{(n)}(\boldsymbol{\theta})$, whose ranks are $R_1^{(n)}(\boldsymbol{\theta}), ..., R_n^{(n)}(\boldsymbol{\theta})$. Since $Y_1 = Z_1^{(n)}(\boldsymbol{\theta})$, define $\underset{\sim}{Y}_1 := F^{-1}(R_1^{(n)}(\boldsymbol{\theta})/n + 1)$, and start the recurrence

$$\underset{\sim}{Y}_t := \vartheta \underset{\sim}{Y}_{t-1} + (1 + \beta \underset{\sim}{Y}_{t-1}^2) F^{-1}\left(\frac{R_t^{(n)}(\boldsymbol{\theta})}{n+1}\right), \quad t \geq 2. \tag{4.9}$$

Finally, the rank-based central sequence (for reference density f) is

$$\underset{\sim}{\boldsymbol{\Delta}}^{(\boldsymbol{n})}(\boldsymbol{\theta}, f) \;=\; \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{pmatrix} \phi_f\left(F^{-1}\left(\frac{R_t^{(n)}(\boldsymbol{\theta})}{n+1}\right)\right) - m_{f,(1)}^{(n)} \\ \psi_f\left(F^{-1}\left(\frac{R_t^{(n)}(\boldsymbol{\theta})}{n+1}\right)\right) - m_{f,(2)}^{(n)} \end{pmatrix} \frac{\underset{\sim}{Y}_{t-1}}{1 + \beta \underset{\sim}{Y}_{t-1}},$$

where $m^{(n)}_{f,(2)}$ is as in (4.5) and

$$m^{(n)}_{f,(1)} = \frac{1}{n}\sum_{i=1}^{n} \phi_f\left(F^{-1}\left(\frac{i}{n+1}\right)\right). \tag{4.10}$$

From the re-centering Lemma in Appendix, it follows that both $m^{(n)}_{f,(1)}$ and $m^{(n)}_{f,(2)}$ are $o\big(n^{-1/2}\big)$. The asymptotic covariance $\mathbf{\Gamma}^*(\boldsymbol{\theta}, f)$ of $\underset{\sim}{\boldsymbol{\Delta}}^{(\boldsymbol{n})}(\boldsymbol{\theta}, f)$ under $\mathcal{H}^{(n)}_{\boldsymbol{\theta}}$, which is also the semiparametric information matrix under $\mathcal{H}^{(n)}_{\boldsymbol{\theta},f}$, is of the block-diagonal form

$$\left(\begin{array}{c|c} I_1(f)\boldsymbol{I}_{p_1\times p_1} & \mathbf{0} \\ \hline \mathbf{0} & I_2(f)\boldsymbol{I}_{p_2\times p_2} \end{array}\right) \boldsymbol{\Upsilon^{-1}(\theta)} \tag{4.11}$$

where $\boldsymbol{\Upsilon^{-1}(\theta)}$ is the Gaussian information matrix obtained as in Chebana and Laïb (2010).

(c) AR(p)-ARCH(q) Consider the process with dynamics

$$Y_t = \sum_{j=1}^{p} \vartheta_j Y_{t-j} + \left(1 + \sum_{l=1}^{q} \beta_l Y^2_{t-j}\right)^{1/2} \epsilon_t, \tag{4.12}$$

where the ϵ_t's are i.i.d. with standardized density g, $\boldsymbol{\theta} = (\vartheta_1, ..., \vartheta_q, \beta_1, ..., \beta_q)$, and the parameters satisfy the assumptions for stationarity in Pantula (1988). Because of the combination of AR with ARCH process, the ULAN central sequence for $\boldsymbol{\vartheta}$ features a location and a scale component:

$$\boldsymbol{\Delta}^{(\boldsymbol{n})}(\boldsymbol{\theta}, g) \;=\; \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{\phi_g\left(Z_t(\boldsymbol{\theta})\right)}{\left(1+\sum_{l=1}^{q}\beta_l Y^2_{t-l}\right)^{1/2}}\begin{pmatrix} Y_{t-1} \\ \vdots \\ Y_{t-p}\end{pmatrix}' \frac{\psi_g\left(Z_t(\boldsymbol{\theta})\right)}{1+\sum_{l=1}^{q}\beta_l Y^2_{t-l}}\begin{pmatrix} Y^2_{t-1} \\ \vdots \\ Y^2_{t-q}\end{pmatrix}'\right)' \tag{4.13}$$

Using the results in Hansen (1991), one can show that there exists an asymptotically equivalent version of (4.13) which only depends on a finite number of past shocks. Then, the definition of a rank-based central sequence (associated with reference density f) is obtained using the approximate scores. However, one can derive heuristically a rank-based central sequence by: (i) using a recurrence similar to (4.3) and (4.9), starting from $Y_0 = 0$; (ii) replacing Z_t in (4.13) by $F^{-1}(R^{(n)}_t/(n+1))$. The resulting rank-based statistic must be re-centered by means of $m^{(n)}_{f,(1)}$ and $m^{(n)}_{f,(2)}$. Closed-form expression of $m^{(n)}_{f,(1)}$ when f is the Gaussian, the Logistic, and the Laplace density follow from

the values provided by Hallin and Mélard (1988), while $m^{(n)}_{f,(2)}$ is easily computed as in (4.5). The re-centering Lemma in Appendix shows that both $m^{(n)}_{f,(1)}$ and $m^{(n)}_{f,(2)}$ are $o\big(n^{-1/2}\big)$. The asymptotic covariance $\mathbf{\Gamma}^*(\boldsymbol{\theta}, f)$ of $\underset{\sim}{\boldsymbol{\Delta}}^{(\boldsymbol{n})}(\boldsymbol{\theta}, f)$ under $\mathcal{H}^{(n)}_{\boldsymbol{\theta}}$ (which coincides with the semiparametric information matrix under $\mathcal{H}^{(n)}_{\boldsymbol{\theta},f}$) is of the form (4.11), with $\boldsymbol{\Upsilon^{-1}(\theta)}$ derived as in Theorem 3.1 of Pantula (1988).

4.2 Discretely observed continuous-time models

(d) Discretely observed Lévy processes with known conditional moments of order 1 and 2 Lévy processes are flexible and popular models, which are able to capture many features of financial time series such as fat tails and jumps. Inference on this class of processes seldom can be conducted using classical likelihood methods, and alternative approaches need to be explored.

Jump diffusion. Consider the Poisson-Gaussian process $\mathcal{Y}$ solution to

$$dY_s = -\varkappa Y_s ds + d\mathcal{W}_s + d\mathcal{Z}_s, \tag{4.14}$$

where $d\mathcal{W}_s$ is the standard Brownian motion and $d\mathcal{Z}_s = J_s d\pi(s)$, with π a Poisson process with intensity 1, and i.i.d. $\mathcal{N}(\alpha, \eta^2)$ jump sizes J_s. The exact first and second conditional moments of $\mathcal{Y}$ are available in closed form, yielding

$$\mathrm{E}(Y_{th}|Y_{(t-1)h}) = \frac{\alpha h}{\varkappa}\left(1 - \exp(-\varkappa h)\right) + Y_{(t-1)h}\exp(-\varkappa h) \tag{4.15}$$

and

$$\mathrm{Var}(Y_{th}|Y_{(t-1)h}) = \frac{1+\eta^2}{2\varkappa}\left(1 - \exp(-2\varkappa h)\right). \tag{4.16}$$

That class of models has been considered by Das (2002) in the dynamic analysis of bond markets, with special focus on the series of Fed funds rates; he points out that the bond market often overreacts, i.e., exhibits large moves in the interest rate followed by speedy reversals. The parameter $\varkappa$ measures the the speed of mean reversion and it plays the main role. Das' estimation of $\varkappa$ is essentially based on a dynamic location and scale model, as obtained by a first-order (Euler) discretization of the stochastic differential equation (4.14) — Das' method ignores the discretization bias; see example *(e)* for two possible solutions to this problem.

We are proposing here a semiparametric rank-based alternative to Das' method. Assume that the discrete-time process $\{Y_{th}; t \in \mathbb{Z}\}$ is observed over $n+1$ periods, yielding $(Y_0, Y_h, Y_{2h}, \ldots, Y_{nh})$.

We base our analysis on the autoregressive model

$$Y_{th} = \frac{\alpha h}{\varkappa}(1-\exp(-\varkappa h)) + Y_{(t-1)h}\exp(-\varkappa h) + \left[\frac{1+\eta^2}{2\varkappa}\left(1-\exp(-2\varkappa h)\right)\right]^{1/2}\epsilon_{th}, \tag{4.17}$$

where the innovation density g remains unspecified. That AR(1) model matches the first- and second-order moments (4.15) and (4.16) of the original model (4.14); see Hallin et al. (2000) for a related discussion in the case of Ornstein-Uhlenbeck process. For fixed g, model (4.17) is ULAN with respect to $\boldsymbol{\theta} := (\varkappa, \alpha, \eta)$, with central sequence (see (3.2))

$$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix} \psi_g(Z_t)\beta_1(\boldsymbol{\theta}) + \phi_g(Z_t)\varkappa^2 Y_{(t-1)h}\beta_2(\boldsymbol{\theta}) - \phi_g(Z_t)\beta_2(\boldsymbol{\theta})\beta_3(\boldsymbol{\theta}) \\ \phi_g(Z_t)\beta_4(\boldsymbol{\theta}) \\ \psi_g(Z_t)\beta_5(\boldsymbol{\theta}) \end{pmatrix} \tag{4.18}$$

where

$$Z_t = Z_t(\boldsymbol{\theta}) = \frac{Y_{th} - \frac{\alpha h}{\varkappa}(1-\exp(-\varkappa h)) - \exp(-\varkappa h)Y_{(t-1)h}}{\left[\frac{1+\eta^2}{2\varkappa}\left(1-\exp(-2\varkappa h)\right)\right]^{1/2}}, \tag{4.19}$$

with

$$\begin{aligned}
\beta_1(\boldsymbol{\theta}) &= \tfrac{1}{2}\left[h(\mathrm{Coth}(h\varkappa)-1) - \tfrac{\eta^2}{\varkappa(2\varkappa+\eta^2)}\right], \\
\beta_2(\boldsymbol{\theta}) &= \alpha(1-\exp(h\varkappa)+h\varkappa), \\
\beta_3(\boldsymbol{\theta}) &= \varkappa^{-3/2}h\exp(-h\varkappa/2)\left[(2\varkappa+\eta^2)\,\mathrm{Sinh}(h\varkappa)\right]^{-1/2}, \\
\beta_4(\boldsymbol{\theta}) &= h\left(1-\exp(-h\varkappa)\right)^{1/2}/\varkappa\left(1+\tfrac{\eta^2}{2\varkappa}\right)^{1/2},
\end{aligned}$$

and $\beta_5(\boldsymbol{\theta}) = \eta/\left(2\varkappa+\eta^2\right)$; $\mathrm{Coth}(x)$ and $\mathrm{Sinh}(x)$ as usual stand for the hyperbolic cotangent and hyperbolic sinus of x, respectively. Canceling $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g)$ yields M-estimators for $\boldsymbol{\theta}$—among them, the Gaussian QL estimator. Due to the highly non-linear form of the estimating equations, however, numerical implementation is likely to be problematic, and so is the derivation of fully semiparametric estimators in the style of Bickel et al. (1993).

The R-estimation methods developed here thus naturally enter into the picture. Projection of $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)$ (where f is some chosen reference density) onto the space of the ranks cancels the second and third components; as for the first one, the terms with hyperbolic functions disappear, and only $\phi_g(Z_t)\varkappa^2 Y_{(t-1)h}\beta_2(\boldsymbol{\theta})$ has a nondegenerate projection. This means that neither α nor η are identifiable unless the actual density g of ϵ_{th} in (4.17) can be correctly specified—which is impossible, due to the approximate nature of the model. Going back to (4.17), this is intuitively clear, as α and η

only appear in the innovation's unconditional location and scale, while the ranks are invariant to location and scale perturbations. For (standardized) reference density f, the projection onto the σ-algebra of residual ranks of the component of the central sequence associated with $\varkappa$ thus coincides (up to a multiplicative constant that will play no role) with that of $n^{-1/2}\sum_{t=1}^{n}\phi_f(Z_t)Y_{(t-1)h}$.

More formally, let $R_t^{(n)}(\boldsymbol{\theta})$ denote the rank of $Z_t(\boldsymbol{\theta})$ in (4.19). That rank is the same as the rank $R_t^{(n)}(\varkappa)$ of $Z_t^{\dagger}(\varkappa) := Y_{th} - \exp(-\varkappa h)Y_{(t-1)h}$. Then, along the same lines as in example 4.3 of Hallin and Werker (2003) a (univariate) rank-based central sequence emerges, of the form (up to a multiplicative constant and $o_{\mathrm{P}}(1)$ terms)

$$\underset{\sim}{\Delta}^{(n)}(\varkappa, f) := n^{1/2}\sum_{i=0}^{s_n}\exp(-i\varkappa h)(n-i)^{-1}\sum_{t=i+1}^{n}\Big(\phi_f\Big(F^{-1}\Big(\frac{R_t^{(n)}(\varkappa)}{n+1}\Big)\Big)F^{-1}\Big(\frac{R_{t-i}^{(n)}}{n+1}\Big) - m_f^{(n)}\Big)$$

where $m_f^{(n)} := [n(n-1)]^{-1}\sum_{1\le i_1\neq i_2\le n}\phi_f(F^{-1}(i_1/n+1))F^{-1}(i_2/n+1)$, with, under $\mathcal{H}_{\boldsymbol{\theta}}^{(n)}$, asymptotic variance $\Gamma^*(\boldsymbol{\theta}, f) = I_1(f)/\big(1-\exp(-2\varkappa h)\big)$; semiparametric efficiency arguments refer to the approximate model (4.17).

Building on a relatively simple AR(1) technology, the projection onto the space of the ranks defines (see Section 5 for details) root-n consistent R-estimators for $\varkappa$ in the rather sophisticated context of a discretized jump diffusion process, where the jump parameters are treated as nuisance —hence, to some extent, the R-estimators are robust to jumps misspecification.

Discretely observed Cox-Ingersoll-Ross (CIR) processes The CIR process is the solution to

$$dY_s = k(1-Y_s)dt + \sigma\sqrt{Y_s}d\mathcal{W}_s, \tag{4.20}$$

where $2k > \sigma^2$, with conditional mean and variance (see, e.g., Bibby et al. (2010))

$$\begin{aligned}\mathrm{E}\big(Y_{th}|Y_{(t-1)h}\big) &= (1-\exp(-kh)) + Y_{(t-1)h}\exp(-kh),\\ \mathrm{Var}\big(Y_{th}|Y_{(t-1)h}\big) &= Y_{(t-1)h}\frac{\sigma^2}{k}\left(\exp(-kh)-\exp(-2kh)\right) + \frac{\sigma^2}{2k}\left(1-\exp(-kh)\right)^2.\end{aligned}$$

Assume that the discrete-time process $\{Y_{th}; t\in\mathbb{Z}\}$ is observed. The exact transition density of the process is known in closed form, but the derivation of the maximum likelihood estimator (MLE)

for $\boldsymbol{\theta} := (k, \sigma^2)$ is numerically cumbersome. As a result, most empirical studies rely on the model

$$\begin{aligned} Y_{th} &= (1 - \exp(-kh)) + Y_{(t-1)h} \exp(-kh) \\ &\quad + \left[Y_{(t-1)h} \frac{\sigma^2}{k} (\exp(-kh) - \exp(-2kh)) + \frac{\sigma^2}{2k} (1 - \exp(-kh))^2 \right]^{1/2} \epsilon_{th}, \end{aligned} \tag{4.21}$$

with i.i.d. standard normal ϵ_{th}'s. The resulting QL estimator typically exhibits quite large standard errors, but improvements can be expected from using R-estimators in the context of the semiparametric model (2.1). To this end, notice that the discrete-time model in (4.21) is an AR(1)-ARCH(1) model; this brings us back, thus, to example (c).

(e) Discretely observed diffusions with unknown conditional moments Consider the general diffusion model $\mathcal{Y}$ solution to

$$dY_s = \mu(Y_s; \boldsymbol{\theta}) dt + \sigma(Y_s; \boldsymbol{\theta}) d\mathcal{W}_s, \tag{4.22}$$

with parameter $\boldsymbol{\theta} \in \mathbb{R}^p$, and assume that the drift and diffusion coefficient satisfy the standard conditions for the existence of a solution (see, e.g., Karlin and Taylor (1981)). Even though the conditional moments of $\mathcal{Y}$ are not known in closed form, approximate models can be constructed for the discretized process $\{Y_{th}\}$ of the form $(Y_0, Y_h, Y_{2h}, \ldots, Y_{nh})$. We briefly sketch here two ways of obtaining such approximate models.

(i) One can derive accurate analytical approximations to the conditional mean $m(\boldsymbol{y}, \boldsymbol{\theta})$ and conditional variance $v^2(\boldsymbol{y}, \boldsymbol{\theta})$ using the infinitesimal generator $\mathcal{L}_{\boldsymbol{\theta}} := \mu(y, \boldsymbol{\theta}) d/dy + 0.5\sigma^2(y, \boldsymbol{\theta}) d^2/dy^2$ of the process. For a smooth function φ of class $C^{\varsigma+1}$ (typically, a polynomial), indeed, we have

$$\mathrm{E}_{\boldsymbol{\theta}} \left(\varphi \left(Y_{th} \right) \middle| Y_{(t-1)h} \right) = \sum_{i=0}^{\varsigma} \frac{h^i}{i!} \mathcal{L}^i_{\boldsymbol{\theta}} \varphi(Y_{(t-1)h}) + O\left(h^{\varsigma+1} \right), \quad t = 1, \ldots, n \quad \varsigma \in \mathbb{N}; \tag{4.23}$$

see Bibby et al. (2010). The formula in (4.23) yields approximations of $m(\boldsymbol{y}, \boldsymbol{\theta})$ and $v^2(\boldsymbol{y}, \boldsymbol{\theta})$ setting $\varphi(y) = y$ and $\varphi(y) = y^2$, respectively. Those approximate conditional moments then can be matched, in the spirit of examples (e) and (f), by simple (conditional) location-scale models. We refer to Bibby et al. (2010) for details, and for numerical evidence that (4.23) in practice provides surprisingly precise approximations for daily ($h = 1/250$) and weekly ($h = 1/52$) data; moreover, for any given h, arbitrarily accurate approximations can be obtained using further powers of $\mathcal{L}_{\boldsymbol{\theta}}$.

(ii) One also can define a *first-order Euler discretization* of (4.22), namely,

$$Y_{th} = Y_{(t-1)h} + \mu(Y_{(t-1)h}; \boldsymbol{\theta})h + \sigma(Y_{(t-1)h}; \boldsymbol{\theta})\epsilon_{th} \tag{4.24}$$

and use it as an auxiliary model in the context of indirect inference, see Gourieroux et al. (1993).

Both strategies lead to a semiparametric model for the discrete-time observation of the diffusion $\mathcal{Y}$ which, under fixed density g and suitable conditions, can be shown to be ULAN with central sequence $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g)$, say. Then, the projection argument of the central sequence onto the σ-algebra generated by the ranks of the residuals leads to the definition of a rank-based central sequence.

As an illustration, consider the class of diffusion equations (see Morozan (1986))

$$dY_s = \lambda(1 - Y_s)ds + (1 + \sigma Y_s)d\mathcal{W}_s \tag{4.25}$$

where $d\mathcal{W}_s$ is standard Brownian motion, and set $\boldsymbol{\theta} := (\lambda, \sigma)$. Assume that $(Y_0, Y_h, Y_{2h}, ..., Y_{jh}, ..., Y_{nh})$ are observed. Under strategy (i), we approximate the first two conditional moments of $\mathcal{Y}$ by (4.23):

$$\mathrm{E}(Y_{th}|Y_{(t-1)h}) = Y_{(t-1)h} + h\lambda(1 - Y_{(t-1)h}) + O(h^2) \quad \text{and} \quad \mathrm{Var}(Y_{th}|Y_{(t-1)h}) = h(1 + \sigma Y_{(t-1)h})^2 + O(h^2).$$

Those approximate conditional moments are matched by the discrete-time AR(1)-LARCH(1) semiparametric model

$$Y_{th} = h\lambda + (1 - h\lambda)Y_{(t-1)h} + \sqrt{h}(1 + \sigma Y_{(t-1)h})\epsilon_{th}, \tag{4.26}$$

where the ϵ_{th}'s are i.i.d. with unspecified standardized density g.

Turning to strategy (ii), in can be seen that the first step of the indirect inference procedure yields a first-order (Euler) discretization of (4.25) which, in this particular model, coincides with (4.26); the two strategies thus lead to the same approximate semiparametric model.

Once this approximate model (4.26) is obtained, the literature typically proposes a pseudo-Gaussian approach, and derives the QL estimator of $\boldsymbol{\theta}$. The QL approach here is only one among many possible inferential methods, though, and Gaussian densities usually constitute a poor fit to the error distributions in (4.26). Following Hallin et al. (2000), we abandon the Gaussian specification of g. For $g \in \mathcal{G}$, the ULAN central sequence for (4.26) reads

$$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g) \;=\; \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{pmatrix} \phi_g(Z_{th}) h (1 - Y_{(t-1)h}) \\ \psi_g(Z_{th}) \sqrt{h} Y_{(t-1)h} \end{pmatrix} \frac{1}{1 + \sigma Y_{(t-1)h}}, \tag{4.27}$$

with $Z_{th} = (Y_{th} - (1 - h\lambda) Y_{(t-1)h}) / \sqrt{h}(1 + \sigma Y_{(t-1)h})$. A rank-based central sequence is obtained along the same lines as in example (b).

The resulting R-estimator of λ and σ^2 achieve root-n consistency and semiparametric efficiency at the reference density in the approximate model (4.26), not in the original one (4.25), where they are likely to be biased. For "small" values of h (e.g., daily or weekly data), the bias is "small", and in many practical applications, one could simply ignore it. However, a careful treatment of that bias is worth a theoretical investigation. In particular, we conjecture that the bias, in both strategies, depends on h and can be controlled by well-known arguments. In strategy (i), assuming that $n \to \infty$ and $h \to 0$ in such a way that $nh^{\varrho} \to 0$ for some $\varrho \in \mathbb{N}$ (which depends on ς in (4.23)) should yield a root-n consistent estimator for the parameters in (4.25); see Bibby et al. (2010) for a discussion. In strategy (ii), we suppose that one is able to correct the bias by standard simulation procedures as in the indirect inference method; see Gourieroux et al. (1993) for a discussion.

5 R-estimation

We now explain how the rank-based central sequences $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$ obtained in the previous section can be used in the construction of R-estimators. As a test statistic, the quadratic form

$$\underset{\sim}{Q}_{\mathrm{HL}}(\boldsymbol{\theta}_0, f) := \underset{\sim}{\boldsymbol{\Delta}}^{(n)\prime}(\boldsymbol{\theta_0}, f) \boldsymbol{\Gamma}^{*-1}(\boldsymbol{\theta}_0, f) \underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}_0, f)$$

provides an optimal rank test for the hypothesis $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Therefore, it would be natural to define (for reference density f) the R-estimator of $\boldsymbol{\theta}$ as the minimizer, with respect to $\boldsymbol{\theta}$, of the rank-based quadratic form $\underset{\sim}{Q}_{\mathrm{HL}}(\boldsymbol{\theta}, f) := \underset{\sim}{\boldsymbol{\Delta}}^{(n)\prime}(\boldsymbol{\theta}, f) \boldsymbol{\Gamma}^{*-1}(\boldsymbol{\theta}, f) \underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$. Despite its simplicity and intuitive appeal, this definition, which in a much simpler context goes back to Hodges and Lehmann, runs into serious difficulties. The non-convex form of $\underset{\sim}{Q}_{\mathrm{HL}}(\boldsymbol{\theta}, f)$ indeed results in practical implementation problems (e.g., multiple solutions and local minima).

Those difficulties have been solved, in the context of linear models with independent observations, by Jaeckel (1972) who observed that the minimizer of $\underset{\sim}{Q}_{\mathrm{HL}}(\boldsymbol{\theta}, f)$ is asymptotically the same as that of another quadratic form, $\underset{\sim}{Q}_{\mathrm{J}}(\boldsymbol{\theta}, f)$, which involves residual ranks but also the residuals themselves. Jaeckel's method, unfortunately, does not extend readily to the present context, since

its statistical justification does not hold anymore. Instead, we consider here a one-step version of the minimization of $\underset{\sim}{Q}_{\mathrm{HL}}(\boldsymbol{\theta}, f)$, inspired from Le Cam's one-step estimation method.

5.1 One-step R-estimators

Let $\hat{\boldsymbol{\theta}}^{(n)}$ and $\widehat{\boldsymbol{\Gamma}}_f^{(n)}$ denote an arbitrary root-n consistent (under $\mathrm{P}_{\boldsymbol{\theta},f}^{(n)}$) estimator of $\boldsymbol{\theta}$ and a consistent estimator of $\boldsymbol{\Gamma}^*(\boldsymbol{\theta}, f)$, respectively. Assume moreover that $\hat{\boldsymbol{\theta}}^{(n)}$ is *asymptotically discrete*, that is, only takes a finite number of values in balls of radius $cn^{-1/2}$ ($c > 0$) centered at $\boldsymbol{\theta}$. Our one-step R-estimation method is based on the following result, the proof of which readily follows from standard results (see, e..g, Chapter 6 of Le Cam and Yang (2000)) and (since part (iii) of Assumption (F), under $g = f$, follows from ULAN) can be considered a particular case of Proposition 6.

Proposition 5. *Let Assumptions (A)-(E) hold. The one-step R-estimator*

$$\underset{\sim}{\boldsymbol{\theta}}_f^{(n)} := \hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\big(\widehat{\boldsymbol{\Gamma}}_f^{(n)}\big)^{-1}\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\hat{\boldsymbol{\theta}}^{(n)}, f) \tag{5.1}$$

under $\mathrm{P}_{\boldsymbol{\theta},f}^{(n)}$ *is root-n consistent and asymptotically normal, with*

$$n^{1/2}(\underset{\sim}{\boldsymbol{\theta}}_f^{(n)} - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}\big(\mathbf{0}, \boldsymbol{\Gamma}^{*-1}(\boldsymbol{\theta}, f)\big).$$

Since (Proposition 4) the rank-based central sequence $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$ is a version of the semiparametrically efficient central sequence at $(\boldsymbol{\theta}, f)$, with asymptotic covariance the semiparametric information matrix $\boldsymbol{\Gamma}^*(\boldsymbol{\theta}, f)$, the R-estimator $\underset{\sim}{\boldsymbol{\theta}}^{(n)}$ in (5.1) is semiparametrically efficient under density f.

Next, in the general case where the reference density f does not necessarily match the actual one g, consider the following assumption.

Assumption (F). Under $\mathrm{P}_{\boldsymbol{\theta},g}^{(n)}$, (i) $\hat{\boldsymbol{\theta}}^{(n)}$is a root-$n$ consistent and asymptotically discrete estimator of $\boldsymbol{\theta}$,

(ii) $\widehat{\boldsymbol{\Gamma}}_f^{(n)}$ is a consistent estimator of the *cross-information matrix*

$$\boldsymbol{\Gamma}(\boldsymbol{\theta}, f, g) := \lim_{n\to\infty} \mathrm{E}_{\boldsymbol{\theta},g}\left[\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)\Big(\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, g)\Big)'\right], \tag{5.2}$$

(in section 5.2, we explain how to construct such estimators), and

(iii) g is such that (asymptotic linearity) $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}, f) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f) = -\boldsymbol{\Gamma}(\boldsymbol{\theta}, f, g)\boldsymbol{\tau} + o_{\mathrm{P}}(1)$,

(note that, for $f = g$, $\boldsymbol{\Gamma}(\boldsymbol{\theta}, f, f) = \boldsymbol{\Gamma}^*(\boldsymbol{\theta}, f)$).

We then have, for the one-step R-estimator (5.1), the following result (see the Appendix for a proof):

Proposition 6. *Let Assumptions (A)-(F) hold. Then, under* $\mathrm{P}_{\boldsymbol{\theta},g}^{(n)}$,

$$n^{1/2}\big(\underset{\sim}{\boldsymbol{\theta}}_f^{(n)} - \boldsymbol{\theta}\big) \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta}, f, g)\boldsymbol{\Gamma}^{*}(\boldsymbol{\theta}, f)\boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta}, f, g)\right).$$

Proposition 6 implies that $\underset{\sim}{\boldsymbol{\theta}}_f^{(n)}$ remains root-n consistent for any $(f, g) \in \mathcal{G}$ such that part (iii) of Assumption (F) holds; contrary to many M- and L-estimators, thus, our R-estimators are robust to model misspecification.

5.2 Cross-information quantities

5.2.1 Estimation

An important issue in the implementation of our R-estimation methodology is related to the need, in part (ii) of Assumption (F), for a consistent estimator of the cross-information matrix $\boldsymbol{\Gamma}(\boldsymbol{\theta}, f, g)$ in (5.2). Constructing such an estimator is typically a complicated task, since $\boldsymbol{\Gamma}(\boldsymbol{\theta}, f, g)$ involves the expectation, under the actual density g, which is unknown, of quantities that themselves depend on g and f. Estimation procedures have been proposed in Cassart et al. (2010). A fully general method is developed in Hallin and Paindaveine (2014).

Very often, though, the matrix $\boldsymbol{\Gamma}(\boldsymbol{\theta}, f, g)$ has a special structure that can be exploited in order to simplify that estimation. For instance, some models (e.g., the AR or ARCH ones) yield the factorization $\boldsymbol{\Gamma}(\boldsymbol{\theta}, f, g) = \mathcal{J}(f, g)\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\theta})$, where $\mathcal{J}(f, g)$ is a scalar quantity depending on f and g only, while $\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\theta})$ only depends on $\boldsymbol{\theta}$. In some others, $\boldsymbol{\Gamma}(\boldsymbol{\theta}, f, g)$ is block-diagonal, with $\jmath$ blocks, each of which is enjoying a similar factorization. This is the case for most models considered here.

A precise formulation of that simplifying assumption is as follows.

Assumption (G). For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $f, g \in \mathcal{G}$, the cross-information matrix $\boldsymbol{\Gamma}(\boldsymbol{\theta}, f, g)$

(G1) is block-diagonal, with $\jmath$ full-rank blocks of the form $\mathcal{J}_1(f, g)\boldsymbol{\Upsilon}_1^{-1}(\boldsymbol{\theta}), \ldots, \mathcal{J}_\jmath(f, g)\boldsymbol{\Upsilon}_\jmath^{-1}(\boldsymbol{\theta})$ where the scalar *cross-information quantities* $\mathcal{J}_j(f, g)$ only depend on f and g, while the matrices $\boldsymbol{\Upsilon}_j(\boldsymbol{\theta})$ only depend on $\boldsymbol{\theta}$, $j = 1, \ldots, \jmath$;

(G2) is such that the mapping $\boldsymbol{\theta} \mapsto \boldsymbol{\Gamma}(\boldsymbol{\theta}, f, g)$ is continuous on $\boldsymbol{\Theta}$.

When Assumption (G) holds, the procedure developed in Cassart et al. (2010), applies; the same procedure was also used by Hallin et al. (2013) in the context of linear models for independent observations with symmetric α-stable innovation density.

In our location-scale models, Assumption (G), when it holds, takes the even simpler form

$$\mathbf{\Gamma}(\boldsymbol{\theta}, f, g) = \left(\begin{array}{c|c} I_1(f,g)\boldsymbol{I}_{p_1\times p_1} & \mathbf{0} \\ \hline \mathbf{0} & I_2(f,g)\boldsymbol{I}_{p_2\times p_2} \end{array} \right) \boldsymbol{\Upsilon}^{-1}(\boldsymbol{\theta}) \tag{5.3}$$

with $\jmath = 2$, $\boldsymbol{I}_{p_1\times p_1}$ and $\boldsymbol{I}_{p_1\times p_1}$ unit matrices of adequate dimension p_1 and p_2, and $\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\theta})$ is the asymptotic covariance matrix of the Gaussian quasi-likelihood estimator. In particular, Assumption (G) holds with $\mathbf{\Gamma}(\boldsymbol{\theta}, f, g)$ of the form (5.3) as soon as $\mathcal{G}$ is restricted to symmetric (with respect to 0) densities – see Section 5.2.2 for details – an assumption which is quite common in the literature, see, e.g., Gourieroux et al. (1984), Linton (1993), and Hallin et al. (2013). In that setting, if consistent estimators $\hat{I}_1(f,g)$ and $\hat{I}_2(f,g)$ for the scalars $I_1(f,g)$ and $I_2(f,g)$ are available, the one-step R-estimator $\underset{\sim}{\boldsymbol{\theta}}{}_f^{(n)}$ is defined as

$$\underset{\sim}{\boldsymbol{\theta}}{}_f^{(n)} := \hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\boldsymbol{\Upsilon}(\boldsymbol{\theta}) \left(\begin{array}{c|c} \hat{I}_1^{-1}(f,g)\boldsymbol{I}_{p_1\times p_1} & \mathbf{0} \\ \hline \mathbf{0} & \hat{I}_2^{-1}(f,g)\boldsymbol{I}_{p_2\times p_2} \end{array} \right) \underset{\sim}{\boldsymbol{\Delta}}{}^{(n)}(\hat{\boldsymbol{\theta}}^{(n)}, f). \tag{5.4}$$

Cassart et al. (2010) propose the following consistent estimators. For any $(\lambda_1, \lambda_2) \in \mathbb{R}^2$, let

$$\tilde{\boldsymbol{\theta}}^{(n)}(\lambda_1, \lambda_2) := \hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\boldsymbol{\Upsilon}(\boldsymbol{\theta}) \left(\begin{array}{c|c} \lambda_1\boldsymbol{I}_{p_1\times p_1} & \mathbf{0} \\ \hline \mathbf{0} & \lambda_2\boldsymbol{I}_{p_2\times p_2} \end{array} \right) \underset{\sim}{\boldsymbol{\Delta}}{}^{(n)}(\hat{\boldsymbol{\theta}}^{(n)}, f);$$

the desired estimators of $I_1(f,g)$ and $I_2(f,g)$ then are $(\hat{I}_1(f,g), \hat{I}_2(f,g)) := \big((\lambda_{*1}^{(n)})^{-1}, (\lambda_{*2}^{(n)})^{-1}\big)$, where

$$(\lambda_{*1}^{(n)}, \lambda_{*2}^{(n)}) := \inf_{(\lambda_1,\lambda_2)\in\mathbb{R}^+\times\mathbb{R}^+} \left\{ \lambda_1, \lambda_2 | \underset{\sim}{\boldsymbol{\Delta}}{}^{(n)}(\hat{\boldsymbol{\theta}}^{(n)})' \boldsymbol{\Upsilon}(\hat{\boldsymbol{\theta}}^{(n)}) \boldsymbol{\Upsilon}(\tilde{\boldsymbol{\theta}}^{(n)}(\lambda_1, \lambda_2)) \underset{\sim}{\boldsymbol{\Delta}}{}^{(n)}(\tilde{\boldsymbol{\theta}}^{(n)}(\lambda_1, \lambda_2)) < 0 \right\}. \tag{5.5}$$

5.2.2 Examples

Let us show that, under suitable assumptions on innovation densities, Assumption (G) holds for all models considered in Section 4. To illustrate this point, let us define

$$\begin{aligned} I_1(f,g) &:= \int_0^1 \phi_f\left(G^{-1}(u)\right)\phi_g\left(F^{-1}(u)\right)du, & I_2(f,g) &:= \int_0^1 \psi_f\left(G^{-1}(u)\right)\psi_g\left(F^{-1}(u)\right)du, \\ I_{12}(f,g) &:= \int_0^1 \phi_f\left(G^{-1}(u)\right)\psi_g\left(F^{-1}(u)\right)du, \quad \text{and} & I_{21}(f,g) &:= \int_0^1 \psi_f\left(G^{-1}(u)\right)\phi_g\left(F^{-1}(u)\right)du. \end{aligned}$$

Those four (cross-)information quantities enter the definition of $\mathbf{\Gamma}(\boldsymbol{\theta}, f, g)$. Assumption (G1) clearly holds when $I_{12}(f,g)$ and $I_{21}(f,g)$ (which appear in the off-diagonal blocks, if any, of $\mathbf{\Gamma}(\boldsymbol{\theta}, f, g)$) both

vanish --a condition which is clearly satisfied when f and g both are symmetric. Here are a few examples from Section 4.1.

ARCH(q). The central sequence in Eq. (4.2) implies that the information matrix satisfies Assumption (G1) with $\jmath = 1$, scalar cross-information quantity $\mathcal{J}(f,g) = I_2(f,g)$ and the $q \times q$ matrix $\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\theta})$ given in Kristensen and Rahbek (2005) (page 951). Symmetry of the innovation density here is not required; that matrix is continuous in $\boldsymbol{\theta}$, so that (G2) also holds.

AR(p)-LARCH(q). If g is symmetric, Assumption (G1) holds with two blocks ($\jmath = 2$), $\mathcal{J}_j(f,g) = I_j(f,g)$, $j = 1, 2$, and $\boldsymbol{\Upsilon}(\boldsymbol{\theta})$ the asymptotic variance matrix of the Gaussian quasi-likelihood estimator derived in Corollary 1 of Chebana and Laïb (2010). In case g is not symmetric, the general method of Hallin and Paindaveine (2014) applies, though; (G2) clearly holds.

AR(p)-ARCH(q). Assuming again that g is symmetric, Assumption (G1) similarly holds, now with the asymptotic variance derived as in Theorem 3.1 of Pantula (1988), which also satisfies (G2). In case g is not symmetric, the general method of Hallin and Paindaveine (2014) applies.

6 Numerical examples and simulation study

6.1 Return and realized volatility

In this section, we study the asymptotic and finite-sample performance of several R-estimators in the model

$$r_t = \varsigma_t \epsilon_t \qquad \text{with} \qquad \log \varsigma_t = \theta_1 \log \varsigma_{t-1} + \theta_2 \log \varsigma_{t-2} + \theta_3 \log \varsigma_{t-3} + v_t, \tag{6.1}$$

where ς_t is a random variable taking values in $\mathbb{R}^+$, $\{\epsilon_t\}$ is independent standard normal white noise, the v_t's are i.i.d. with standardized density g, and ϵ_t is independent of v_s for all (s,t).

This model is related to the normal variance mean mixture models (Barndorff-Nielsen et al. (1982)), which are applied in modeling and forecasting the realized volatility of assets; see, e.g., Corsi (2009), Corsi et al. (2013), and references therein.

We study the efficiencies of the QL and R-estimators based on the van der Waerden, Wilcoxon, and Laplace scores. Centering and scaling constants $m_f^{(n)}$ and $s_f^{(n)}$ are provided explicitly in Hallin and Mélard (1988); see also Lemma in Appendix. Thus, we consider

(a) a Gaussian reference density f, yielding a rank-based central sequence $\underset{\sim}{\boldsymbol{\Delta}}{}^{(n)}_{\text{vdW}}$ which is a linear

combination of the van der Waerden correlation coefficients

$$\underset{\sim}{r}^{(n)}_{\mathrm{vdW};i} = \left(s^{(n)}_{vdW}\right)^{-1} \left\{(n-i)^{-1} \sum_{t=i+1}^{(n)} \Phi^{-1}\left(\frac{R^{(n)}_t}{n+1}\right) \Phi^{-1}\left(\frac{R^{(n)}_{t-i}}{n+1}\right) - m^{(n)}_{vdW}\right\}, \tag{6.2}$$

where Φ as usual stands for the standard normal distribution function;

(b) a logistic reference density f, yielding a rank-based central sequence $\underset{\sim}{\Delta}^{(n)}_W$ which is a linear combination of the Wilcoxon correlation coefficients

$$\underset{\sim}{r}^{(n)}_{W;i} = \left(s^{(n)}_W\right)^{-1} \left\{(n-i)^{-1} \sum_{t=i+1}^{(n)} \left(\frac{R^{(n)}_t}{n+1} - \frac{1}{2}\right) \log\left(\frac{R^{(n)}_{t-u}}{n+1-R^{(n)}_{t-i}}\right) - m^{(n)}_W\right\}; \tag{6.3}$$

(c) a double-exponential reference density f, yielding a rank-based central sequence $\underset{\sim}{\Delta}^{(n)}_L$ which is a linear combination of the Laplace correlation coefficients ($I[\cdot]$ stands for the indicator function)

$$\begin{aligned}\underset{\sim}{r}^{(n)}_{L;i} = \left(S^{(n)}_L\right)^{-1} \Bigg\{(n-i)^{-1} \sum_{t=i+1}^{(n)} \mathrm{sign}\left(\frac{R^{(n)}_t}{n+1} - \frac{1}{2}\right) \left[\log\left(2\frac{R^{(n)}_{t-u}}{n+1}\right) I\left[\frac{R^{(n)}_{t-u}}{n+1} \le \frac{1}{2}\right]\right] \\ - \left[\log\left(2 - 2\frac{R^{(n)}_{t-u}}{n+1}\right) I\left[\frac{R^{(n)}_{t-u}}{n+1} > \frac{1}{2}\right]\right] - m^{(n)}_L\Bigg\}.\end{aligned} \tag{6.4}$$

Under (6.1), Assumption (G) is satisfied with $\jmath = 1$, without any symmetry assumption on g.

Asymptotic performance (AREs). The performance of traditional methods as QL estimation is typically sensitive to skewness and kurtosis. In order to study the impact of skewness and leptokurtosis on R- and QL estimators, we are considering here densities g in the four-parameter family of *Johnson's densities* with unbounded support; see Jones and Pewsey (2009) and Ghysels and Wang (2011). In the sequel, we refer to the general density in this class by $\mathrm{J}_{\mathrm{SU}}(\gamma, \delta, \mu, \sigma)$, where γ and δ are skewness and kurtosis parameters, respectively, while μ and σ, as usual, stand for location and scale. In Figure 1 we plot the excess of kurtosis and the skewness corresponding to various choices for γ and δ. Remark that $\delta \le 1$ and $\gamma \ge 3$ yield high kurtosis excess and quite negative skewness.

The asymptotic relative efficiency (ARE) under g of the R-estimator associated with reference density f with respect to the QL estimator is easily obtained as $I_1^2(f,g)/I_1(g)$, with $I_1(f,g)$ defined in (5.6), and $I_1(g) := I_1(g,g)$. Results are provided in Table 1. They indicate that, under leptokurtic innovation density g (e.g., $\delta = 0.85$ or 1), all R-estimators considered here quite significantly

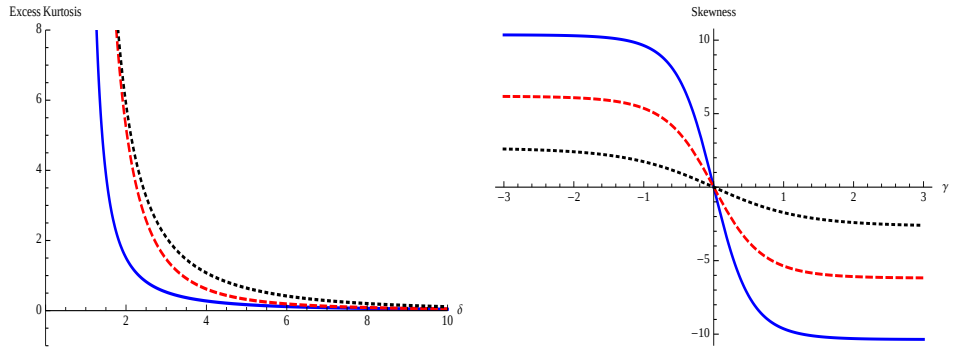

Figure 1: Johnson's densities $\mathrm{J_{SU}}(\gamma, \delta, 0, 1)$. Left panel: excess of kurtosis (y-axis) for different values of δ (x-axis) and $\gamma = 0$ (continuous line), $\gamma = 3$ (dotted line), $\gamma = 10$ (dashed line). Right panel: skewness (y-axis) for different values of γ (x-axis) and $\delta = 0.85$ (continuous line), $\delta = 1$ (dotted line), $\delta = 1.5$ (dashed line).

outperform their QL competitors, whether g is asymmetric or not. Skewness alone, however, has a somewhat limited impact on AREs. Under skewed and leptokurtic g's (e.g., $\gamma = 4$, $\delta = 1$), the relative performance of R-estimators is particularly impressive: an ARE value of 12.341 is reached under $(\gamma, \delta) = (4, 1)$.

6.2 Return and time-varying volatility

The next Monte Carlo simulation illustrates how semiparametrically efficient estimates improve on the efficiency of out-of-sample Value-at-Risk (VaR) prediction based on QL estimates.

Let $\mathcal{Y} := \{Y_t\}_{t\in\mathbb{Z}}$ be a strictly stationary stochastic process, modeling the daily rate of return of a financial asset with price P_t at time t; that is $Y_t = \log P_t - \log P_{t-1}$. We set for $\mathcal{Y}$ the AR(1)-ARCH(1) model

$$Y_t = \vartheta Y_{t-1} + \left(1 + \beta Y_{t-1}^2\right)^{1/2} \epsilon_t, \tag{6.5}$$

where the ϵ_t's are i.i.d., with location zero, scale one, and featuring a double-exponential density.

In principle, one should compute the MLE of $\boldsymbol{\theta} = (\vartheta, \beta)$. However, implementation of the MLE requires a reparameterization and is numerically quite heavy, see Peng and Yao (2003). As a consequence, the QL estimation represents an easy-to-implement alternative estimation procedure; see, e.g., Mancini et al. (2005). The QL simplicity, however, comes at the price of efficiency loss, while the R-estimator based on Laplace scores is a natural candidate in the context.

The aim of the next Monte Carlo exercise is twofold. We illustrate that the Laplace R-estimator:

	$J_{SU}(\gamma, \delta, \mu, \sigma)$					
	Leptokurtic		Skewed		Leptokurtic and Skewed	
ARE	$\gamma = 0$, $\delta = 0.85$	$\gamma = 0$, $\delta = 1$	$\gamma = 3$, $\delta = 10$	$\gamma = 10$, $\delta = 10$	$\gamma = 3$, $\delta = 1.5$	$\gamma = 4$, $\delta = 1$
vdW/QL	2.567	1.755	1.002	1.014	2.657	12.341
W/QL	3.245	2.124	0.960	0.968	2.207	7.319
L/QL	3.433	2.033	0.643	0.644	1.234	2.972

Table 1: AREs, under various values of γ and δ, of R-estimators (van der Waerden, Wilcoxon, Laplace) with respect to the QL estimator. The underlying process in an AR(3), with $J_{SU}(\gamma, \delta, \mu, \sigma)$ innovation density.

(i) is an easy-to-implement alternative to MLE; (ii) improves on the QL efficiency —by construction, it achieves semiparametric efficiency at the double exponential density. Our Monte Carlo exercise is conducted as follows: (Step 1) we generate from (6.5) a trajectory of $n = 100, 500, 750$ observations, using $\vartheta = 0.05$, and $\beta = 0.5$; (Step 2) from each of those trajectories, we compute the QL estimator and the Laplace one-step R-estimator; the latter is based on the projection of the central sequence (4.13) onto the space of residual ranks, with the score $\phi_L(x) = \sqrt{2}(x - 1/2)$ and the QL as preliminary estimator; (Step 3) from the resulting estimated conditional means and variances, we compute the out-of-sample one-step-ahead predicted conditional (to Y_t^*) VaR at confidence level α; that VaR is such that $\alpha = P_{\boldsymbol{\theta},g}\left(Y_{t+1} \leq \text{VaR}_{t,t+1}^{\alpha} \| Y_t^*\right)$, for $\alpha = 5\%$; (Step 4) we repeat steps (1)-(3) 5000 times for each sample size —we use the same Y_t^* in all replications.

We display the results in Table 2. Both the QL and the Laplace R-estimator exhibit median predicted VaRs which essentially coincide with the actual VaR, and only entail small bias for all sample sizes. However, the rank-based VaR predictions for the R-estimator display smaller MSEs (about one half) than the QL-based ones. The gain is due to the fact that the R-estimator achieves higher efficiency than the QL estimator. Since we use QL estimates as preliminary estimates, the one-step R-estimation procedure represents an efficiency tuning, which nevertheless comes with very valuable benefits in terms of robustness, since the R-estimator $\underset{\sim}{\boldsymbol{\theta}}^{(n)}$ is such that $n^{1/2}(\underset{\sim}{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta})$ is asymptotically equivalent to a rank-based random variable, hence asymptotically measurable with respect to the σ-algebra generated by the ranks. Such combinations of efficiency and robustness

are rare, and worth being mentioned.

	$n = 100$	$n = 500$	$n = 750$
	L/QL	L/QL	L/QL
MSE	56.27%	57.63%	60.53%
|Bias|	73.65%	72.82%	72.90%

Table 2: One-day-ahead conditional VaR estimates at 5% level for different sample sizes: ratios, with respect to the QL, of the MSE and the absolute value of the bias (|Bias|) associated with Laplace (L) R-estimator. The MC size is 5000 and $Y_t^* = 0.0072$ (from the limiting distribution).

7 Empirical analysis: USD/CHF exchange rate

One of the major problems in the analysis of financial time series is the relatively frequent occurrence of extreme values —a phenomenon rank-based methods, including R-estimation, are less sensitive to than traditional parametric and semiparametric methods. We illustrate this point with an empirical analysis of the series of USD/CHF exchange rate daily log-returns and its realized volatility, as measured by the Two Scales Realized Volatility (TSRV) series, see Aït-Sahalia et al. (2005). Our analysis builds on the empirical findings of Andersen et al. (2000, 2003).

7.1 Data

Our dataset consists of tick-by-tick *log mid prices* over 24 hours of USD/CHF FX rates provided by Olsen&Associates; log mid prices are computed as averages of the logarithmic bid and ask quotes, obtained from the Reuters FXFX screen. In order to avoid modeling the seasonal behavior of trading activity induced by week-ends, we exclude all trades taking place from Friday 21:00 GMT to Sunday 22:00 GMT. From the high-frequency quotes, we compute (as in Corsi (2009)) TSRVs by summing the high-frequency squared log-returns with slow scale of ten ticks, and daily log-returns as $r_t = \log P_t - \log P_{t-1}$, where P_t is the daily USD/CHF exchange rate provided by Reuters. We conduct our analysis on the 1993 and 1997 data. In each year, we use the first 200 observations (from January to end September) as training data for model estimation and diagnosis (see Sections 7.2.1 and 7.2.2 below), and the last 50 ones (from October to December) to evaluate forecasting performances. In Figure 2 we plot the log-return and TSRV trajectories.

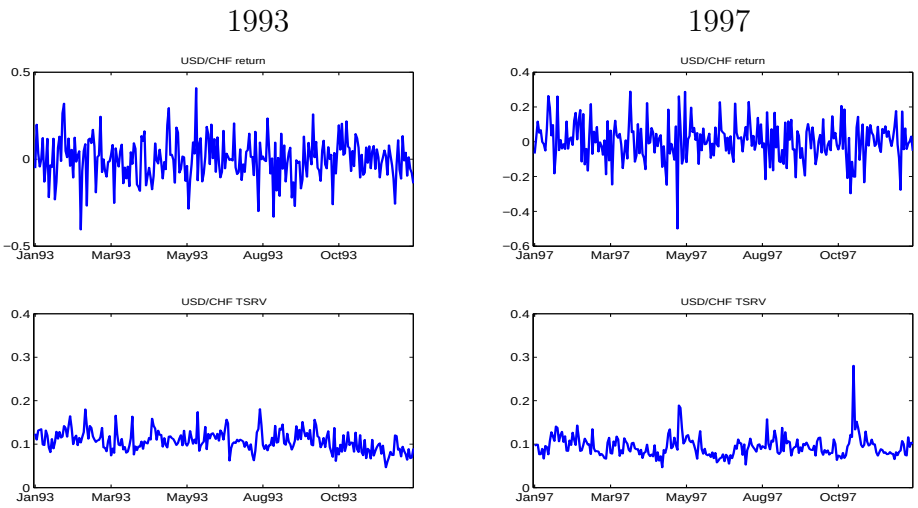

Figure 2: USD/CHF FX rate return: log-return (top panels) and annualized TSRV (bottom panels) series for the years 1993 and 1997.

7.2 Modeling and forecasting

7.2.1 Exploratory analysis and modeling

Log-return. As in Andersen et al. (2000), we consider the dynamics of the process of log-returns r_t standardized by the TSRV, namely $r_t/TSRV_t$. The resulting series has approximately mean zero, variance close to one, and a sample partial correlation analysis with robust standard errors (unreported) does not detect any predictability. The Shapiro-Wilks test for $r_t/TSRV_t$ are 0.896 and 0.208 for the 1997 and 1993 data, respectively. Thus, we conclude that the a standard normal approximation for the $r_t/TSRV_t$ is supported by the data.

Two Scales Realized Volatilities (TSRV and $\log$*(TSRV)).* Turning to volatilities, we consider the TSRV process and its log-transformation. Table 3 displays some summary statistics for their unconditional distribution. The years 1993 and 1997 illustrate different aspects of the data: (i) the 1993 training period (January-September) exhibits 9 extreme values; we label it as "standard", and believe it expresses the typical dynamics of the TSRV; (ii) in 1997, the training period (January-September) shows 7 extreme values, while the Asian crisis is causing 4 extreme values between October and December (the TSRV strikes 0.3). We label this period as "non-standard", since it contains several unfrequent negative volatility shocks related to a well-identified financial crisis. The log-transformation of the TSRV slightly reduces the number of extreme observations; however, similar considerations still hold. We model the series of logged TSRVs by (2.1), assuming an AR

specification for the conditional mean and $\epsilon_t \sim g(0,1)$. The autocorrelation analysis (unreported) of the training data suggests that an AR process with no more than three lags is a suitable model. Thus, we set that the conditional mean of the log(TSRV) is of the form $\sum_{j=1}^{3} \theta_j \log(\text{TSRV}_{t-j})$.

	1993		1997	
		TSRV		
	Jan-Sept	Oct-Dec	Jan-Sept	Oct-Dec
Mean	0.112	0.088	0.094	0.096
SD	0.022	0.018	0.021	0.033
Kurtosis	3.532	2.770	5.662	20.296
$q_{.75} - q_{.25}$	0.028	0.024	0.027	0.026
obs $\leq q_{.50} - 3\times$MAD	3	0	0	0
obs $\geq q_{.50} + 3\times$MAD	7	0	7	4
		log(TSRV)		
	Jan-Sept	Oct-Dec	Jan-Sept	Oct-Dec
Mean	-2.206	-2.451	-2.388	-2.375
SD	0.192	0.208	0.217	0.256
Kurtosis	3.421	3.212	3.634	7.982
$q_{.75} - q_{.25}$	0.256	0.278	0.298	0.288
obs $\leq q_{.50} - 3.5\times$MAD	4	1	2	0
obs $\geq q_{.50} + 3.5\times$MAD	3	0	3	2

Table 3: USD/CHF FX rates: descriptive statistics (empirical means, standard errors, kurtoses and interquartile ranges; numbers of extreme values) for the TSRV (top panel) and log(TSRV) (bottom panel) series, 1993 and 1997. We define as "extreme" any observation lying outside the region defined by the median plus or minus c_1 times the median absolute deviation (MAD) over the period considered; we set $c_1 = 3$ for the TSRV and $c_1 = 3.5$ for the log(TSRV).

7.2.2 Estimation and diagnostics

Building on the previous considerations, we set up a normal mean-variance mixture model, of the form (6.1), with TSRV_t playing the role of ς_t. We estimate the model parameters from the data in each training period, and assess the quality of the various estimates—QL, and the van der Warden (vdW), Wilcoxon (W), and Laplace (L) R-estimators—via their standard errors; the latter are obtained by estimating the cross-information quantities in the variance matrix as in Section 5.2. Results are displayed in Table 4.

In the 1993 training data, all estimation methods considered suggest an AR(1) model for the

	1993				1997			
	QL	vdW	W	L	QL	vdW	W	L
θ_1	0.2762	0.3204	0.3525	0.4014	0.3719	0.3517	0.3677	0.3921
	(0.072)	(0.051)	(0.070)	(0.045)	(0.071)	(0.063)	(0.080)	(0.077)
θ_2	0.0969	0.0988	0.0768	0.0190	0.1323	0.1586	0.1408	0.1761
	(0.074)	(0.061)	(0.075)	(0.048)	(0.076)	(0.066)	(0.085)	(0.081)
θ_3	-0.0371	-0.0396	-0.0316	0.0008	0.0911	0.0669	0.0606	0.033
	(0.073)	(0.051)	(0.070)	(0.045)	(0.071)	(0.063)	(0.080)	(0.077)

Table 4: USD/CHF: inference on log(TSRV). QL and R-estimates (and standard errors).

$\log(\text{TSRV}_t)$ series, while in 1997 the same estimation methods agree on an AR(2) model. Table 4 indicates that the standard errors of R-estimators in general are smaller than those of QL estimators.

To assess the validity of the fitted models, we consider some standard diagnostics. In Figure 3, we plot the sample partial autocorrelation of the standardized and squared-standardized residuals, as implied by the Laplace R-estimator for the training period January-September 1993. None of the plots provide any evidence of autocorrelation outside Bartlett's two-standard-error bands for white noise. Similar plots (unreported) are available for the QL estimator and the other R-estimators.

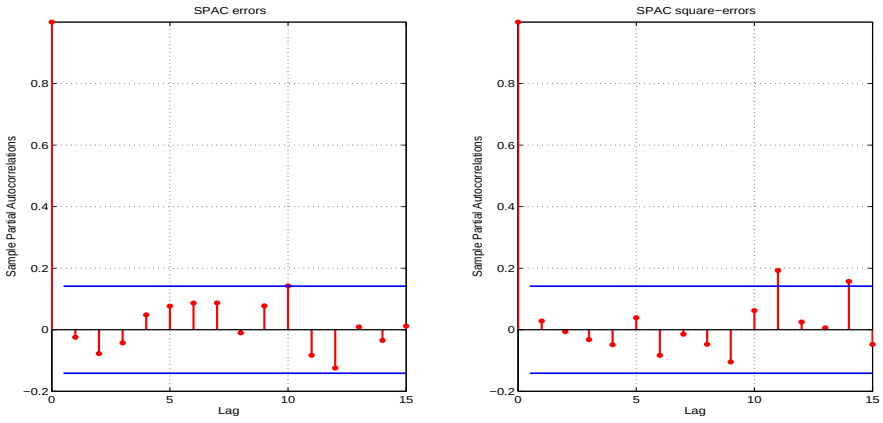

Figure 3: USD/CHF: diagnostics for Laplace R-estimator in 1993 data. Sample partial autocorrelation of residuals (left) and of squared-residuals (right).

7.2.3 Forecasting

We computed, for each day in the October-December period (still 1993 and 1997), the squared one-day-ahead prediction error for each estimator obtained from the corresponding training period.

Table 5 provides some classical (mean and standard deviation) and robust (median and mean absolute deviation) evaluations of the squared bias and dispersion of the prediction errors. R-estimators (particularly the Laplace ones) appear to provide more accurate forecasts than the QL estimators, but the improvements, in terms of location and dispersion, are smaller in the "crisis year" 1997 than in 1993. This is probably due to the extreme values related to the Asian crisis. Such large values, which are not representative of the actual dynamics, badly affect prediction errors —less so, however, when using rank-based methods than with the traditional QL ones.

	1993				1997			
	QL	vdW/QL	W/QL	L/QL	QL	vdW/QL	W/QL	L/QL
Mean	0.24	96%	94%	91%	0.17	100%	99%	99%
Median	0.13	97%	99%	88%	0.05	94%	97%	84%
SD	0.38	98%	97%	96%	0.56	100%	100%	98%
MAD	0.11	103%	108%	99%	0.04	96%	98%	90%

Table 5: USD/CHF: mean, median, standard deviation, and mean absolute deviation for the squared one-day-ahead prediction errors based on QL estimator and ratios, with respect to the latter, of the corresponding quantities for R-estimators.

8 Discussion

The class of econometric models that can be handled by our methodology is large. Moreover, our ideas could be extended to indirect inference, e.g., using the models in Section 4 as auxiliary models, and our R-estimation can be applied to construct indirect inference estimators. To this end, note that the choice of the reference density considered in the construction of the rank-based central sequence can be made data-driven, in order to (i) capture some features (such as skewness, kurtosis) of the actual density, and (ii) get closer to semiparametric efficiency at the actual density. Finally, R-estimation of the mean reversion parameter described in Section 4.2 (e) could be applied to derive robust statistical arbitrages.

References

Aït-Sahalia, Y., 2002. Telling from discrete data whether the underlying continuous–time model is a diffusion. The Journal of Finance 57 (5), 2075–2112.

Aït-Sahalia, Y., 2006. Likelihood inference for diffusions: a survey. Frontiers in Statistics: in Honor of Peter J. Bickel′s 65th Birthday, 369–407.

Aït-Sahalia, Y., Mykland, P., Zhang, L., 2005. A tale of two time scales: determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association 100, 1394–1411.

Andersen, T., Bollerslev, T., Diebold, F., Labys, P., 2000. Exchange rate returns standardized by realized volatility are (nearly) gaussian. Multinational Finance Journal 4, 159–179.

Andersen, T., Bollerslev, T., Diebold, F., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625.

Andrews, B., 2008. Rank-based estimation for autoregressive moving average time series models. Journal of Time Series Analysis 29 (1), 51–73.

Andrews, B., 2012. Rank-based estimation for GARCH processes. Econometric Theory 28 (05), 1037–1064.

Barndorff-Nielsen, O., Kent, J., Sørensen, M., 1982. Normal variance-mean mixtures and z-distribution. International Statistical Review 50, 145–159.

Bibby, B., Jacobsen, M., Sørensen, M., 2010. Estimating functions for discretely sampled diffusion-type models. Handbook of Financial Econometrics, Editor: Y. Aït-Sahalia, and L.P. Hansen, North Holland.

Bickel, P. J., Klaassen, C. A., Ritov, Y., Wellner, J., 1993. Efficient and adaptive estimation for semi-parametric models. Johns Hopkins Univ. Press, Baltimore.

Cassart, D., Hallin, M., Paindaveine, D., 2010. On the estimation of cross-information quantities in rank-based inference. Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková, 35-45.

Chebana, F., Laïb, N., 2010. An efficient locally asymptotic parametric test in nonlinear heteroscedastic time series models. Afrika Statistika 5 (1), 197-209.

Cont, R., Tankov, P., 2004. Financial modelling with jump processes. Vol. 133. Chapman & Hall/CRC, Boca Raton.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7, 174–196.

Corsi, F., Fusari, N., La Vecchia, D., 2013. Realizing smiles: Options pricing with realized volatility. Journal of Financial Economics 107 (2), 284–304.

Das, S., 2002. The surprise element: jumps in interest rates. Journal of Econometrics 106 (1), 27–65.

Davidson, J., 1994. Stochastic Limit Theory: An Introduction for Econometricians. Oxford university press.

Drost, F., Klaassen, C., Werker, B., 1997. Adaptive estimation in time-series models. The Annals of Statistics 25, 786–817.

Francq, C., Zakoïan, J.-M., 2010. Inconsistency of the MLE and inference based on weighted LS for LARCH models. Journal of Econometrics 159 (1), 151–165.

Ghysels, E., Wang, F., 2011. Some useful densities for risk management and their properties. Preprint, forthcoming in Econometric Reviews.

Giraitis, L., Kokoszka, P., Leipus, R., 2000a. Stationary ARCH models: dependence structure and central limit theorem. Econometric Theory 16 (1), 3–22.

Giraitis, L., Surgailis, D., 2002. ARCH-type bilinear models with double long memory. Stochastic Processes and their Applications 100 (1), 275–300.

Gourieroux, C., Monfort, A., Renault, E., 1993. Indirect inference. Journal of Applied Econometrics 8, S85–S118.

Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo maximum likelihood methods: theory. Econometrica 52, 681–700.

Hallin, M., 1994. On the nonadmissibility of correlogram-based time series methods. Journal of Time Series Analysis 16, 607–612.

Hallin, M., Koell, C., Werker, B., 2000. Optimal inference for discretely observed semiparametric Ornstein-Uhlenbeck processes. Journal of Statistical Planning and Inference 91, 323–340.

Hallin, M., Mélard, G., 1988. Rank-based tests for randomness against first-order serial dependence. Journal of the American Statistical Association 83 (404), 1117–1128.

Hallin, M., Paindaveine, D., 2014. Semiparametric efficient one-step R-estimators, working paper, ECARES, Université Libre de Bruxelles.

Hallin, M., Swan, Y., Verdebout, T., Veredas, D., 2013. One-step R-estimation in linear models with stable errors. Journal of Econometrics 172, 195–204.

Hallin, M., Werker, B., 2003. Semi-parametric efficiency, distribution-freeness and invariance. Bernoulli 9, 137–165.

Hansen, B. E., 1991. GARCH (1, 1) processes are near epoch dependent. Economics Letters 36 (2), 181–186.

Jaeckel, L. A., 1972. Estimating regression coefficients by minimizing the dispersion of the residuals. The Annals of Mathematical Statistics, 1449–1458.

Jones, M., Pewsey, A., 2009. Sinh-arcsinh distributions. Biometrika 96 (4), 761–780.

Karlin, S., Taylor, H., 1981. A second course in stochastic processes. Academic Press, New York.

Kloeden, P. E., Platen, E., 1994. Numerical solution of SDE through computer experiments. Springer, Berlin.

Koul, H., Ossiander, M., 1994. Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. The Annals of Statistics 22, 540–562.

Koul, H. L., Saleh, A. M. E., 1993. R-estimation of the parameters of autoregressive [AR(p)] models. The Annals of Statistics, 534–551.

Kristensen, D., Rahbek, A., 2005. Asymptotics of the QMLE for a class of ARCH(q) models. Econometric Theory 21 (05), 946–961.

Le Cam, L., Yang, G. L., 2000. Asymptotics in statistics: some basic concepts. Springer, New York.

Linton, O., 1993. Adaptive estimation in ARCH models. Econometric Theory 9, 539–569.

Mancini, L., Ronchetti, E., Trojani, F., 2005. Optimal conditionally unbiased bounded-influence inference in dynamic location and scale models. Journal of the American Statistical Association 100(470), 628–641.

Morozan, T., 1986. Periodic solutions of affine stochastic differential equations. Stochastic Analysis and Applications 4 (1), 87–110.

Mukherjee, K., 2007. Generalized R-estimators under conditional heteroscedasticity. Journal of Econometrics 141 (2), 383–415.

Mukherjee, K., Bai, Z., 2002. R-estimation in autoregression with square-integrable score function. Journal of Multivariate Analysis 81 (1), 167–186.

Pantula, S. G., 1988. Estimation of autoregressive models with ARCH errors. Sankhyā: The Indian Journal of Statistics, Series B, 119–138.

Peng, L., Yao, Q., 2003. Least absolute deviations estimation for ARCH and GARCH models. Biometrika 90 (4), 967–975.

Terpstra, J. T., McKean, J. W., Naranjo, J. D., 2001. GR-estimates for an autoregressive time series. Statistics & probability letters 51 (2), 165–172.

van der Vaart, A., 1998. Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics.

Wefelmeyer, W., 1996. Quasi-likelihood models and optimal inference. The Annals of Statistics 24, 405–422.

Zhao, Z., 2008. Parametric and nonparametric models and methods in financial econometrics. Statistics Surveys 2, 1–42.

A TECHNICAL APPENDIX

A.1 Proof of Proposition 3

Proposition 3 follows from Theorem 2.1 in Drost et al. (1997): we just need to check that their Assumptions A-E[1] are satisfied. Specifically,

(a) their Assumption A follows from our Assumption (E);

(b) their Assumption B is a consequence of the location-scale form of (2.1);

(c) setting, for $\tilde{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}^{(n)} = O(n^{-1/2})$,

$$W'_{nt}(\tilde{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}^{(n)}) = \frac{1}{v(\mathbf{Y}_{t-1}, \boldsymbol{\theta}^{(n)})} \Big(m(\mathbf{Y}_{t-1}, \tilde{\boldsymbol{\theta}}^{(n)}) - m(\mathbf{Y}_{t-1}, \boldsymbol{\theta}^{(n)}), v(\mathbf{Y}_{t-1}, \tilde{\boldsymbol{\theta}}^{(n)}) - v(\mathbf{Y}_{t-1}, \boldsymbol{\theta}^{(n)}) \Big)'$$

we have, as $n \to \infty$, $W_{nt} \to W_t$ in the sense of (2.4) in Drost et al., where

$$\begin{aligned} W_t = W_t(\boldsymbol{\theta}) &:= \frac{1}{v(\mathbf{Y}_{t-1}, \boldsymbol{\vartheta})} \frac{\partial}{\partial_{\boldsymbol{\vartheta}'}} (m(\mathbf{Y}_{t-1}, \boldsymbol{\vartheta}), v(\mathbf{Y}_{t-1}, \boldsymbol{\vartheta})) \Big|_{\boldsymbol{\vartheta}=\boldsymbol{\theta}} \\ &= \frac{1}{v(\mathbf{Y}_{t-1}, \boldsymbol{\theta})} (\dot{\boldsymbol{m}}(\mathbf{Y}_{t-1}, \boldsymbol{\theta}), \dot{\boldsymbol{v}}(\mathbf{Y}_{t-1}, \boldsymbol{\theta})) \end{aligned}$$

is $\mathcal{F}_{t-1}$-measurable (see section 4.1 in Drost et al. (1997)); Assumptions C and D thus are satisfied;

(d) $\dot{l}(Z_t, \mathbf{Z}_{t-1}, \boldsymbol{\theta}, g) = W_t(\boldsymbol{\theta})(-\phi_g(Z_t(\boldsymbol{\theta})), \psi_q(Z_t(\boldsymbol{\theta})))'$, as defined in Eq. (3.2), satisfies Assumption E.

Theorem 2.1 in Drost et al. (1997) thus applies, which concludes the proof. □

A.2 Proof of Proposition 6

From the definition of $\underset{\sim}{\boldsymbol{\theta}}{}_f^{(n)}$, the asymptotic linearity of $\underset{\sim}{\boldsymbol{\Delta}}{}^{(n)}$, the consistency of $\widehat{\boldsymbol{\Gamma}}_f^{(n)}$, and the asymptotic discreteness of $\hat{\boldsymbol{\theta}}^{(n)}$ (all following from Assumption (F)), we have, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},g}$,

$$\begin{aligned} n^{1/2}(\underset{\sim}{\boldsymbol{\theta}}{}_f^{(n)} - \boldsymbol{\theta}) &= n^{1/2}\Big(\hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}(\widehat{\boldsymbol{\Gamma}}_f^{(n)})^{-1} \underset{\sim}{\boldsymbol{\Delta}}{}^{(n)}(\boldsymbol{\theta} + n^{-1/2} n^{1/2}(\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}), f) - \boldsymbol{\theta}\Big) \\ &= n^{1/2}\Big(\hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta}, f, g)\Big(\underset{\sim}{\boldsymbol{\Delta}}{}^{(n)}(\boldsymbol{\theta}, f) - \boldsymbol{\Gamma}(\boldsymbol{\theta}, f, g) n^{1/2}(\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta})\Big) - \boldsymbol{\theta}\Big) + o_{\mathrm{P}}(1) \\ &= \boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta}, f, g) \underset{\sim}{\boldsymbol{\Delta}}{}^{(n)}(\boldsymbol{\theta}, f) + o_{\mathrm{P}}(1). \end{aligned}$$

The result then readily follows from the asymptotic normality of $\underset{\sim}{\boldsymbol{\Delta}}{}^{(n)}(\boldsymbol{\theta}, f)$. □

A.3 A re-centering Lemma

We start with a general result on square-integrable of monotone functions.

[1] In this proof, labels A, B, ... refer to the assumptions in Drost et al., labels (A), (B), ... to ours.

Lemma 7. *Let h be a square-integrable monotone non-decreasing function from $(0,1)$ to $\mathbb{R}$. Then,*

$$\frac{1}{n}\sum_{i=1}^{n} h\Big(\frac{i}{n+1}\Big) - \int_0^1 h(u)\ du = o\big(n^{-1/2}\big) \quad \text{as} \ \ n\to\infty.$$

Proof. Without loss of generality, we may assume that $\int_0^1 h(u)\ du = 0$. Since h is monotone non-decreasing and square-integrable,

$$\frac{1}{n+1}h^2\Big(\frac{n}{n+1}\Big) \le \int_{n/(n+1)}^1 h^2(u)\ du = o(1).$$

Therefore, $h^2\big(\frac{n}{n+1}\big) = o(n)$, $h\big(\frac{n}{n+1}\big) = o(n^{1/2})$, and $\frac{1}{n}h\big(\frac{n}{n+1}\big) = o\big(n^{-1/2}\big)$. Similarly, $\frac{1}{n}h\big(\frac{1}{n+1}\big) = o\big(n^{-1/2}\big)$, and hence

$$\max_{1\le i\le n}\frac{1}{n}\big|h\big(\frac{i}{n+1}\big)\big| = o\big(n^{-1/2}\big). \tag{A.1}$$

Let u_0, u^- and u^+ be such that $h(u_0-0)\le 0$, $h(u_0+0)\ge 0$, and $\int_0^{u^-} h(u)\ du = -\int_{u^+}^1 h(u)\ du$, so that $\int_{u^-}^{u^+} h(u)\ du = 0$. Defininig $i^- := \lceil (n+1)u^-\rceil$, $i_0 = \lfloor (n+1)u_0\rfloor$, and $i^+ := \lfloor (n+1)u^+\rfloor$, decompose the sum $S_n := \frac{1}{n}\sum_{i=1}^n h\Big(\frac{i}{n+1}\Big)$ into

$$\begin{aligned} S_n &= S_n^{--} + S_n^{-} + S_n^{+} + S_n^{++} \\ &:= \frac{1}{n}\sum_{i=1}^{i^- -1} h\Big(\frac{i}{n+1}\Big) + \frac{1}{n}\sum_{i=i^-}^{i_0} h\Big(\frac{i}{n+1}\Big) + \frac{1}{n}\sum_{i=i_0+1}^{i^+} h\Big(\frac{i}{n+1}\Big) + \frac{1}{n}\sum_{i=i^++1}^{n} h\Big(\frac{i}{n+1}\Big). \end{aligned}$$

Clearly,

$$0 \le -\frac{n}{n+1}S_n^{--} \le -\int_0^{u^-} h(u)\ du \quad \text{and} \quad 0 \le \frac{n}{n+1}S_n^{++} \le \int_{u^+}^1 h(u)\ du, \tag{A.2}$$

as the corresponding rectangular areas $\frac{1}{n+1}\big|h\big(\frac{i}{n+1}\big)\big|$ lie between the axis and the curve $u\mapsto h(u)$. Also,

$$-\frac{n}{n+1}S_n^{-} = \begin{cases} \underline{D}_n^- & + & \Big(u^- - \dfrac{i^- -1}{n+1}\Big)\Big|h\Big(\dfrac{i^-}{n+1}\Big)\Big| \\ \bar{D}_n^- & + & \Big(\dfrac{i_0+1}{n+1} - u_0\Big)\Big|h\Big(\dfrac{i_0}{n+1}\Big)\Big| - \Big(\dfrac{i^-}{n+1} - u^-\Big)\Big|h\big(u^-\big)\Big| \end{cases} \tag{A.3}$$

and

$$\frac{n}{n+1}S_n^{+} = \begin{cases} \underline{D}_n^+ & + & \Big(\dfrac{i^+ +1}{n+1} - u^+\Big)h\Big(\dfrac{i^+}{n+1}\Big) \\ \bar{D}_n^+ & + & \Big(u_0 - \dfrac{i_0}{n+1}\Big)h\Big(\dfrac{i_0+1}{n+1}\Big) - \Big(u^+ - \dfrac{i^+}{n+1}\Big)h\big(u^+\big) \end{cases} \tag{A.4}$$

where $\underline{D}_n^-$, $\bar{D}_n^-$, $\underline{D}_n^+$ and $\bar{D}_n^+$ are lower and upper Darboux sums, for $\int_{u^-}^{u_0}|h(u)|\ du$ and $\int_{u_0}^{u^+} h(u)\ du$, respectively. Those Darboux sums are such that

$$\underline{D}_n^+ - \bar{D}_n^- \le \int_{u^-}^{u^+} h(u)\ du = 0 \le \bar{D}_n^+ - \underline{D}_n^-$$

It follows from (A.3), (A.4) and (A.2) that

$$\underline{D}_n^+ - \bar{D}_n^- + \int_0^{u^-} h(u)\,du + o\big(n^{-1/2}\big) \tag{A.5}$$

$$\begin{aligned}
&= \underline{D}_n^+ - \bar{D}_n^- + \Big(\frac{i^+ + 1}{n+1} - u^+\Big)h\Big(\frac{i^+}{n+1}\Big) + \Big(\frac{i_0+1}{n+1} - u_0\Big)h\Big(\frac{i_0}{n+1}\Big) - \Big(\frac{i^-}{n+1} - u^-\Big)h\big(u^-\big) + \int_0^{u^-} h(u)\,du \\
&\le \frac{n}{n+1}S_n \\
&\le \bar{D}_n^+ - \underline{D}_n^- + \Big(u_0 - \frac{i_0}{n+1}\Big)h\Big(\frac{i_0+1}{n+1}\Big) - \Big(u^+ - \frac{i^+}{n+1}\Big)h\big(u^+\big) + \Big(u^- - \frac{i^- - 1}{n+1}\Big)h\Big(\frac{i^-}{n+1}\Big) + \int_{u^+}^1 h(u)\,du \\
&= \bar{D}_n^+ - \underline{D}_n^- + \int_{u^+}^1 h(u)\,du + o\big(n^{-1/2}\big)
\end{aligned} \tag{A.6}$$

Now,

$$\bar{D}_n^+ - \underline{D}_n^+ = \Big(u_0 - \frac{i_0}{n+1}\Big)h\Big(\frac{i_0+1}{n+1}\Big) + \Big(\frac{i^+ + 1}{n+1} - u^+\Big)h\Big(\frac{i^+}{n+1}\Big) + \Big(u^+ - \frac{i^+}{n+1}\Big)h\big(u^+\big) = o\big(n^{-1/2}\big),$$

and

$$\bar{D}_n^- - \underline{D}_n^- = -\Big(\frac{i^-}{n+1} - u^-\Big)h\big(u^-\big) - \Big(u^- - \frac{i^- - 1}{n+1}\Big)h\Big(\frac{i^-}{n+1}\Big) + \Big(\frac{i_0+1}{n+1} - u_0\Big)h\Big(\frac{i_0}{n+1}\Big) = o\big(n^{-1/2}\big)$$

It follows that the lower and upper bounds in (A.5) and (A.6) reduce to $\int_0^{u^-} h(u)\,du + o\big(n^{-1/2}\big)$ and $\int_{u^+}^1 h(u)\,du + o\big(n^{-1/2}\big)$, respectively, and their diffference to

$$2\int_{u^+}^1 h(u)\,du + o\big(n^{-1/2}\big),$$

where the $o\big(n^{-1/2}\big)$ quantity is uniform in u^+. The desired result that S_n is $o\big(n^{-1/2}\big)$ follows by considering a sequence u_n^+ converging to 1 in such a way that $\int_{u_n^+}^1 h(u)\,du = o\big(n^{-1/2}\big)$. □

Under Assumption (E), the score functions associated with the reference density f are assumed to be the difference between to monotone increasing square-integrable functions to which Lemma 7 applies.

A.4 Analytical derivation of some rank-based central sequences

A.4.1 ARCH(q)

Consider the class of ARCH(q) models, with dynamics of the form

$$Y_t = \Big(1 + \sum_{j=1}^q \theta_j Y_{t-j}^2\Big)^{1/2}\epsilon_t, \tag{A.7}$$

where the ϵ_t's are i.i.d. with density g, mean zero and variance one, $\boldsymbol{\theta} = (\theta_1, ..., \theta_q)$ where $\theta_j > 0$ for $j = 1, .., q$ and such that $\sum_{j=1}^q \theta_j \le \rho < 1$. Set $Z_t(\boldsymbol{\theta}) = Y_t/\big(1 + \sum_{j=1}^q \theta_j Y_{t-1}^2\big)^{1/2}$. From Eq. (3.2), the ULAN central sequence for $\boldsymbol{\theta}$ is

$$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\psi_g(Z_t(\theta))}{1 + \sum_{j=1}^{q} \theta_j Y_{t-j}^2} \begin{pmatrix} Y_{t-1}^2 \\ \vdots \\ Y_{t-q}^2 \end{pmatrix}. \tag{A.8}$$

Expanding Y_t^2 into a Volterra series, we obtain $Y_t^2 = \epsilon_t^2 + \sum_{k\geq 1} w_t(k)$, where

$$w_t(k) := \sum_{j_1,\ldots,j_k \geq 1} \prod_{r=1}^{k} \theta_{j_r} \prod_{r'=0}^{k} \epsilon_{t-\sum_{h=0}^{r'} j_h}^2 \tag{A.9}$$

(with $j_0 = 0$) converges in probablity (this follows from stationarity of Y_t, see, e.g., Giraitis et al. (2000a) and reference therein) is positive for all k. The condition $\sum_{j=1}^{q} \theta_j \leq \rho < 1$ implies that there exists a sequence s_n such that Y_t^2 can be asymptotically reconstructed using only a finite number s_n of past shocks. Indeed,

$$Y_t^2 = \tilde{Y}_t^2 + \sum_{k \geq s_n} w_t(k), \quad \text{with} \quad \tilde{Y}_t^2 := \epsilon_t^2 + \sum_{k=1}^{s_n} w_t(k). \tag{A.10}$$

In what follows, for the sake of notational simplicity, we set $s = s_n$. Since $\mathrm{E}(\varepsilon_t^2) = 1$, we have $\mathrm{E}_\theta \sum_{k>s}^{\infty} w_t(k) < C\rho^s$ for some $C > 0$. For any $\delta > 0$, the Markov inequality implies that

$$\mathrm{P}_\theta \left(\sum_{k>s}^{\infty} w_t(k) > \delta \right) \leq C\rho^s \delta^{-1}, \qquad \text{hence} \qquad \sum_{s\geq 1} \mathrm{P}_\theta \left(\sum_{k>s}^{\infty} w_t(k) > \delta \right) < \infty$$

and, in view of the first Borel-Cantelli Lemma, $\mathrm{P}_\theta \left(\sum_{k>s}^{\infty} w_t(k) > \delta, \text{i.o.} \right) = 0$. As a result, $\sum_{k>s}^{\infty} w_t(k)$ converges to zero almost surely, hence in probability, as $s \to \infty$.

Combining (A.10) and (A.8), we get

$$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\psi_g\left(Z_t\right)}{1 + \sum_{j=1}^{q} \theta_j \left(Z_{t-j}^2 + \sum_{k\geq 1} w_{t-j}(k) \right)} \begin{pmatrix} Z_{t-1}^2 + \sum_{k\geq 1} w_{t-1}(k) \\ \vdots \\ Z_{t-q}^2 + \sum_{k\geq 1} w_{t-q}(k) \end{pmatrix} \tag{A.11}$$

where, with a slight abuse of notation, $w_{t-j}(k)$ is as in (A.9), with ϵ_j replaced by $\boldsymbol{Z}_{t-j}$, for all $j = 1, ..., q$. Defining

$$e_{t-j}(s) := \frac{Y_{t-j}^2}{1 + \sum_{j=1}^{q} \theta_j Y_{t-j}^2} - \frac{Z_{t-j}^2 + \sum_{k=1}^{s} w_{t-j}(k)}{1 + \sum_{j=1}^{q} \theta_j \left(Z_{t-j}^2 + \sum_{k=1}^{s} w_{t-j}(k) \right)},$$

we have $e_{t-j}(s) \leq Y_{t-j}^2 - Y_{t-j}^{2(s)} \leq \sum_{k\geq s} w_{t-j}(k)$, which is $o_{\mathrm{P}}(1)$. Thus, we approximate (A.8) by another central sequence depending on a finite number s of lags only, which we also denote as $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g)$. In view of Section 3.2.2

and (A.11), the corresponding rank-based central sequence takes the form

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta},g)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\psi_g\left(G^{-1}\left(\frac{R_t^{(n)}}{n+1}\right)\right)-m_{g,(2)}^{(n)}\right)\begin{pmatrix}\dfrac{\left(G^{-1}\left(\frac{R_{t-1}^{(n)}}{n+1}\right)\right)^2+\sum_{k=1}^{s}w_{t-1}(k)}{1+\sum_{j=1}^{q}\theta_j\left(\left(G^{-1}\left(\frac{R_{t-j}^{(n)}}{n+1}\right)\right)^2+\sum_{k=1}^{s}w_{t-j}(k)\right)}\\ \vdots\\ \dfrac{\left(G^{-1}\left(\frac{R_{t-q}^{(n)}}{n+1}\right)\right)^2+\sum_{k=1}^{s}w_{t-q}(k)}{1+\sum_{j=1}^{q}\theta_j\left(\left(G^{-1}\left(\frac{R_{t-j}^{(n)}}{n+1}\right)\right)^2+\sum_{k=1}^{s}w_{t-j}(k)\right)}\end{pmatrix},$$

where $w_{t-j}(k)$ is computed by replacing all Z_t's by $G^{-1}(R_t^{(n)}/(n+1))$, and $m_{g,(2)}^{(n)}$, given in (4.5), is such that the expected value of $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta},g)$ is exactly zero for every n.

Lemma 7 implies that $m_{g,(2)}^{(n)}=o\big(n^{-1/2}\big)$, hence can be omitted, yielding

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta},g)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_g\left(G^{-1}\left(\frac{R_t^{(n)}}{n+1}\right)\right)\begin{pmatrix}\dfrac{\left(G^{-1}\left(\frac{R_{t-1}^{(n)}}{n+1}\right)\right)^2+\sum_{k=1}^{s}w_{t-1}(k)}{1+\sum_{j=1}^{q}\theta_j\left(\left(G^{-1}\left(\frac{R_{t-j}^{(n)}}{n+1}\right)\right)^2+\sum_{k=1}^{s}w_{t-j}(k)\right)}\\ \vdots\\ \dfrac{\left(G^{-1}\left(\frac{R_{t-q}^{(n)}}{n+1}\right)\right)^2+\sum_{k=1}^{s}w_{t-q}(k)}{1+\sum_{j=1}^{q}\theta_j\left(\left(G^{-1}\left(\frac{R_{t-j}^{(n)}}{n+1}\right)\right)^2+\sum_{k=1}^{s}w_{t-j}(k)\right)}\end{pmatrix}.$$

A.4.2 AR(p)-LARCH(q)

Consider the discrete-time bilinear process with dynamics

$$Y_t=\sum_{j=1}^{p}\vartheta_jY_{t-j}+\left(1+\sum_{l=1}^{q}\beta_lY_{t-l}\right)\epsilon_t, \tag{A.12}$$

where $\epsilon_t\sim g$, with mean zero and unit variance, and $\boldsymbol{\theta}=(\vartheta_1,...,\vartheta_q,\beta_1,...,\beta_q)$ satisfies Assumptions A_1-A_3 in Giraitis and Surgailis (2002). Model (A.12) includes as a special case the AR(p) process (for $p>0$ and $q=0$) and (for $p=0$ and $q>0$) the LARCH(q) (namely, Linear ARCH) process. Following Francq and Zakoïan (2010), we assume here that $\inf_{\theta\in\Theta}(1+\sum_{l=1}^{q}\beta_lY_{t-l})$ is almost surely bounded away from zero— sufficient condition for this is a compactly supported innovation and compact parameter space Θ with suitable endpoints.

Because of the combination of AR with LARCH process, the ULAN central sequence for $\boldsymbol{\theta}$ features both a

location and a scale component:

$$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta},g) \quad = \quad \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(1+\sum_{l=1}^{q}\beta_l Y_{t-l}\right)^{-1}\left(\begin{array}{cc} \phi_g\left(Z_t(\boldsymbol{\theta})\right) & \left(\begin{array}{c} Y_{t-1} \\ \vdots \\ Y_{t-p} \end{array}\right) \\ \psi_g\left(Z_t(\boldsymbol{\theta})\right) & \left(\begin{array}{c} Y_{t-1} \\ \vdots \\ Y_{t-q} \end{array}\right) \end{array}\right). \tag{A.13}$$

Let $A(z) := \sum_{j=1}^{\infty}\vartheta_j z^j$, $B(z) := \sum_{l=1}^{\infty}\beta_l z^l$ be analytic on $|z|<1$, with $B(z)\neq 1$, and write

$$U(z) := (1-B(z))^{-1} = \sum_{j=0}^{\infty}u_j z^j, \quad \text{and} \quad W(z) := A(z)U(z) = \sum_{j=0}^{\infty}w_j z^j.$$

Giraitis and Surgailis (2002) show the invertibility of Y_t, expressing it as the convergent orthogonal Volterra series

$$Y_t = \sum_{k=1}^{\infty}\sum_{j_k<..<j_1<t} u_{t-j_1}w_{j_1-j_2}...w_{j_{k-1}-j_k}\epsilon_{j_1}...\epsilon_{j_k}, \tag{A.14}$$

which depends on a infinite number of lagged shocks.

From Theorem 2.2 in Giraitis and Surgailis (2002) it follows that $Y_t = Y_t^s + o_{\mathrm{P}}(1)$, where Y_t^s is obtained by truncating Eq. (A.14) to the s-th term, with $s = s(n)\to\infty$ as $n\to\infty$. Additionally, it follows from the continuous mapping theorem that $(1+\sum_{l=1}^{q}\beta_l Y_{t-l}^s)^{-1} = (1+\sum_{l=1}^{q}\beta_l Y_{t-l})^{-1} + o_{\mathrm{P}}(1)$, for every t as $n\to\infty$. As a result,

$$e_{t-j} = \frac{Y_{t-j}}{(1+\sum_{l=1}^{q}\beta_l Y_{t-l})} - \frac{Y_{t-j}^s}{(1+\sum_{l=1}^{q}\beta_l Y_{t-l}^s)} = o_{\mathrm{P}}(1), \tag{A.15}$$

so that, letting

$$\zeta_t := \frac{1}{1+\sum_{l=1}^{q}\beta_l\left(\sum_{k=1}^{s}\sum_{j_k<..<j_1<t-l} g_{t-j_1}w_{j_1-j_2}...w_{j_{k-1}-j_k}Z_{j_1}...Z_{j_k}\right)},$$

$$\boldsymbol{\Delta}_{(1)}^{(t,s)}(\boldsymbol{\theta},g) = \zeta_t\phi_g\left(Z_t(\boldsymbol{\theta})\right)\left(\begin{array}{c} \sum_{k=1}^{s}\sum_{j_k<..<j_1<t-1} g_{t-j_1}w_{j_1-j_2}...w_{j_{k-1}-j_k}Z_{j_1}...Z_{j_k} \\ \vdots \\ \sum_{k=1}^{s}\sum_{j_k<..<j_1<t-p} g_{t-j_1}w_{j_1-j_2}...w_{j_{k-1}-j_k}Z_{j_1}...Z_{j_k} \end{array}\right),$$

and

$$\boldsymbol{\Delta}_{(2)}^{(t,s)}(\boldsymbol{\theta},g) = \zeta_t\psi_g\left(Z_t\left(\boldsymbol{\theta}\right)\right)\left(\begin{array}{c} \sum_{k=1}^{s}\sum_{j_k<..<j_1<t-1} g_{t-j_1}w_{j_1-j_2}...w_{j_{k-1}-j_k}Z_{j_1}...Z_{j_k} \\ \vdots \\ \sum_{k=1}^{s}\sum_{j_k<..<j_1<t-q} g_{t-j_1}w_{j_1-j_2}...w_{j_{k-1}-j_k}Z_{j_1}...Z_{j_k} \end{array}\right),$$

we have that

$$\boldsymbol{\Delta}^{(\boldsymbol{n})}(\boldsymbol{\theta}, g) \quad = \quad \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{pmatrix} \boldsymbol{\Delta}_{(\mathbf{1})}^{(\boldsymbol{t},\boldsymbol{s})}(\boldsymbol{\theta}, g) \\ \boldsymbol{\Delta}_{(\mathbf{2})}^{(\boldsymbol{t},\boldsymbol{s})}(\boldsymbol{\theta}, g) \end{pmatrix} \tag{A.16}$$

is another version of the central sequence in (A.13), since it approximates $\boldsymbol{\Delta}^{(\boldsymbol{n})}(\boldsymbol{\theta}, g)$ up to $o_{\mathrm{P}}(1)$. The rank-based central sequence again is obtained as the approximate-score projection of this central sequence $\boldsymbol{\Delta}^{(\boldsymbol{n})}(\boldsymbol{\theta}, g)$ onto the invariant σ-algebra generated by the ranks, namely replacing, in (A.16), Z_t with $G^{-1}(R_t^{(n)}/(n+1))$ for every t and re-centering the resulting rank-based statistic. We illustrate this construction in the AR(1)-LARCH(1) example.

AR(1)-LARCH(1). Let us consider model (4.8). The truncated Volterra series here takes the form

$$Y_t^s = \sum_{k=1}^{s} (\beta_1/\vartheta_1)^k \sum_{j_k,\ldots,j_1<t} \vartheta_1^{t-j_k} \epsilon_{j_1}, ..., \epsilon_{j_k},,$$

which is such that $Y_t = Y_t^s + o_{\mathrm{P}}(1)$; an alternative representation for Y_t^s follows from iterating Eq. (4.8):

$$Y_t^s = \epsilon_t + \sum_{k=1}^{s} \epsilon_{t-k} \prod_{m=t-k+1}^{t} (\beta_1 \epsilon_m + \vartheta_1), \tag{A.17}$$

which is still implying that $Y_t = Y_t^s + o_{\mathrm{P}}(1)$, see Giraitis and Surgailis (2002, page 282). Setting

$$Z_t := (Y_t - \vartheta_1 Y_{t-1})/(1 + \beta_1 Y_{t-1}),$$

and $\boldsymbol{\theta} := (\vartheta_1, \beta)$, ULAN holds with central sequence

$$\begin{aligned} \boldsymbol{\Delta}^{(\boldsymbol{n})}(\boldsymbol{\theta}, g) \quad &= \quad \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{pmatrix} \phi_g(Z_t) \\ \phi_g(Z_t) Z_t - 1 \end{pmatrix} \frac{\sum_{k=1}^{s} (\beta_1/\vartheta_1)^k \sum_{j_k,\ldots,j_1<t-1} \vartheta_1^{t-j_k} Z_{j_1}, ..., Z_{j_k}}{1 + \beta_1 \sum_{k=1}^{s} (\beta_1/\vartheta_1)^k \sum_{j_k,\ldots,j_1<t-1} \vartheta_1^{t-j_k} Z_{j_1}, ..., Z_{j_k}} \\ &= \quad \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{pmatrix} \phi_g(Z_t) \\ \phi_g(Z_t) Z_t - 1 \end{pmatrix} \frac{Z_t + \sum_{k=1}^{s} Z_{t-k} \prod_{m=t-k+1}^{t} (\beta_1 Z_m + \vartheta_1)}{1 + \beta_1 (Z_t + \sum_{k=1}^{s} Z_{t-k} \prod_{m=t-k+1}^{t} (\beta_1 Z_m + \vartheta_1))}, \end{aligned} \tag{A.18}$$

where the last expression, which is easier to work with, follows from Eq. (A.17). To derive the rank-based central sequence, put

$$\underset{\sim}{\zeta}_t := \frac{1}{1 + \beta_1 \left(G^{-1}\left(\frac{R_t^{(n)}}{n+1}\right) + \sum_{k=1}^{s} G^{-1}\left(\frac{R_{t-k}^{(n)}}{n+1}\right) \prod_{m=t-k+1}^{t} \left(\beta_1 G^{-1}\left(\frac{R_m^{(n)}}{n+1}\right) + \vartheta_1\right)\right)}.$$

Then, letting

$$\begin{aligned}\underset{\sim}{\boldsymbol{\Delta}}_{\mathbf{(1)}}^{\boldsymbol{(t,s)}}(\boldsymbol{\theta},g) &:= \left(\underset{\sim}{\zeta}_t\phi_g\left(G^{-1}\left(\frac{R_t^{(n)}}{n+1}\right)\right)-m_{g,(1)}^{(n)}\right)\\ &\quad\times\left(G^{-1}\left(\frac{R_t^{(n)}}{(n+1)}\right)+\sum_{k=1}^{s}G^{-1}\left(\frac{R_{t-k}^{(n)}}{n+1}\right)\prod_{m=t-k+1}^{t}\left(\beta_1 G^{-1}\left(\frac{R_m^{(n)}}{n+1}\right)+\vartheta_1\right)\right),\end{aligned}$$

and

$$\begin{aligned}\underset{\sim}{\boldsymbol{\Delta}}_{\mathbf{(2)}}^{\boldsymbol{(t,s)}}(\boldsymbol{\theta},g) &:= \underset{\sim}{\zeta}_t\left(\psi_g\left(G^{-1}\left(\frac{R_t^{(n)}}{n+1}\right)\right)-m_{g,(2)}^{(n)}\right)\\ &\quad\times\left(G^{-1}\left(\frac{R_t^{(n)}}{(n+1)}\right)+\sum_{k=1}^{s}G^{-1}\left(\frac{R_{t-k}^{(n)}}{n+1}\right)\prod_{m=t-k+1}^{t}\left(\beta_1 G^{-1}\left(\frac{R_m^{(n)}}{n+1}\right)+\vartheta_1\right)\right),\end{aligned}$$

with $m_{g,(1)}^{(n)}$ and $m_{g,(2)}^{(n)}$ such that the expectations of $\underset{\sim}{\boldsymbol{\Delta}}_{\mathbf{(1)}}^{\boldsymbol{(t,s)}}(\boldsymbol{\theta},g)$ and $\underset{\sim}{\boldsymbol{\Delta}}_{\mathbf{(2)}}^{\boldsymbol{(t,s)}}(\boldsymbol{\theta},g)$ are exactly zero. We set

$$\underset{\sim}{\boldsymbol{\Delta}}^{\boldsymbol{(n)}}(\boldsymbol{\theta},g) := -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}\boldsymbol{\Delta}_{\mathbf{(1)}}^{\boldsymbol{(n)}}(\boldsymbol{\theta},g)\\ \boldsymbol{\Delta}_{\mathbf{(1)}}^{\boldsymbol{(n)}}(\boldsymbol{\theta},g)\end{pmatrix}. \tag{A.19}$$

Finally, note that as in the AR case of Hallin and Werker (2003) and the case of ARCH processes discussed in A.4.1, Lemma 7 implies that $m_{g,(1)}^{(n)}$ and $m_{g,(2)}^{(n)}$ are $o\big(n^{-1/2}\big)$, hence can be omitted.