R-Estimation in Semiparametric Dynamic Location-Scale Models

Marc Hallin[a], Davide La Vecchia[b]

[a] *ECARES, Université libre de Bruxelles, Belgium.*
[b] *Research Center for Statistics, Geneva School of Economics and Management, University of Geneva, Switzerland.*

Abstract

We propose rank-based estimation (R-estimators) as a substitute for Gaussian quasi-likelihood and standard semiparametric estimation in time series models where conditional location and/or scale depend on a Euclidean parameter of interest, while the unspecified innovation density is an infinite-dimensional nuisance. Applications include linear and nonlinear models, featuring either homo- or heteroskedastic conditional distributions (e.g. conditional duration models, AR-ARCH, discretely observed diffusions with jumps, etc.). We show how to construct R-estimators achieving semiparametric efficiency at some predetermined reference density while preserving root-n consistency and asymptotic normality irrespective of the actual density. Contrary to the standard semiparametric estimators (in the style of Bickel, Klaassen, Ritov, and Wellner), our R-estimators neither require tangent space calculations nor innovation density estimation. Numerical examples illustrate their good performances on simulated data. A real-data analysis of the log-return and log-transformed two-scale realized volatility of the USD/CHF exchange rate concludes the paper.

Key words: Autoregressive conditional duration models, Conditional heteroskedasticity, Distribution-freeness, Discretely observed Lévy processes, Forecasting, R-estimation, Realized volatility, Skew-t family.

1. Introduction

1.1. Gaussian dynamic location-scale models

Dynamic location-scale processes are essential tools in time series econometrics, and have motivated the study of increasingly diverse and sophisticated classes of discrete- and continuous-time models such as ARCH, AR-ARCH or AR-LARCH models, AR conditional duration models, or discretely observed diffusions with jumps. While the probabilistic properties of those models have been studied extensively and in great details, their statistical analysis is less exhaustive, and still presents several challenges; see, for instance, Drost et al. (1997), Aït-Sahalia (2006), Zhao (2008), Bibby et al. (2010), and references therein.

Among those challenges is the specification of underlying densities. All models considered in the literature involve some unobserved driving noise, the density of which is often specified to be Gaussian. Although Gaussian assumptions are unrealistic in most applications, an all too common belief is that violating them is essentially harmless, and that Gaussian Quasi-Likelihood (QL) inference methods remain safe and valid as soon as finite second-order moments exist.[1] In particular, QL estimators erroneously are surmised to be root-n consistent and asymptotically normal under very general conditions.

[1] Actually, second-order moments are not sufficient, and root-n consistency and asymptotic normality of QL estimators also require a (uniform) Law of Large Numbers property for dynamic location-scale models, that holds, essentially, under finite *fourth* order moments; see Gouriéroux et al. (1984), Bollerslev and Wooldridge (1992), Hall and Yao (2003), Peng and Yao (2003).

1.2. Semiparametric extensions of Gaussian dynamic location-scale models

The trouble is that dynamic location-scale models are used, mainly, in a financial context where heavy tails are quite common and innovation processes do not have finite fourth moment. As a result, Gaussian QL estimators fail to be root-n consistent and asymptotically normal; see e.g. Hall and Yao (2003). Moreover, even when standard asymptotics (root-n consistency and normality) hold, Gaussian QL estimators yield good performances only if the actual density is "nearly Gaussian", and their efficiency[2] rapidly deteriorates in the presence of skewness or excess kurtosis, two characteristics which are quite common in financial data; see e.g., Engle and Gonzalez-Rivera (1991). Finally, Gaussian QL estimators are highly nonrobust: as shown in Mancini et al. (2005), they can be severely distorted by a small number of outliers. Those pitfalls have been stressed by many authors—Linton (1993) for ARCH models, Drost and Klaassen (1997) for GARCH, Hall and Yao (2003) for heavy-tailed ARCH and GARCH, Drost and Werker (2004) for duration models, Francq and Zakoïan (2010) for LARCH models, Fan et al. (2014) for GARCH, to quote only a few.

Attempts have been made to replace the Gaussian reference density with some more appropriate pseudo densities (e.g. Student ones), defining non-Gaussian QL methods (e.g. Engle and Bollerslev (1986)) which can also feature robustness properties (Lucas (1997)). However, Student-based M-estimators are not Fisher-consistent under misspecified densities (e.g. non-Student ones), leading to root-n inconsistent estimates. Moreover, if the degrees of freedom are not kept fixed, the robustness of Student-based M-estimators is lost (Lucas (1997)).

To palliate the non-standard asymptotics in the absence of finite fourth moment for the innovation term in ARCH-type models, Hall and Yao (2003) propose a bootstrap approach which allows to recover, under certain conditions, the asymptotic distribution of the Gaussian QL estimator. Their solution, however, does not restore root-n consistency, hence does not remedy the lack of rate-optimality of the estimator.

A semiparametric approach, along the standard lines of Bickel et al. (1993)[3] under which the innovation density (call it g) remains unspecified, is therefore more realistic and seems highly advisable. Typical examples of that approach are Linton (1993), Wefelmeyer (1996), Drost et al. (1997), Drost and Klaassen (1997), and Drost and Werker (2004). Moreover, several examples in financial contexts and new theoretical developments are available in Fernandes et al. (2007). "Standard" as it may look, that approach, however, is not without its own difficulties. Not only because of a methodologically and computationally heavy implementation, but also because several distinct semiparametric extensions of the Gaussian model, in general, are possible: those distinct extensions induce distinct semiparametrically efficient estimators (and distinct semiparametric efficiency bounds) the validity of which depends on the semiparametric model adopted.

[2]Namely, their Asymptotic Relative Efficiency (ARE), under the actual density, with respect to the actual efficiency bound.
[3]The monograph by Bickel et al. (1993) actually is restricted to the case of independent observations; we refer to Drost et al. (1997) for a detailed exposition in the time series context.

The following example provides an illustration of the burdens and pitfalls of the standard semiparametric method.

1.3. A motivating example: discretely observed non-Gaussian Ornstein-Uhlenbeck processes

The Ornstein-Uhlenbeck process $\mathcal{Y}$ has dynamics

$$dY_s = -\theta Y_s ds + dL_s. \tag{1.1}$$

Instead of the usual assumption that L_s is Brownian motion (which leads to Gaussian AR-type discretely observed processes), let us assume, more generally, that L_s is some Lévy process. This includes Lévy processes with jumps, such as compound Poisson processes, which are typically considered in the analysis of the (realized) volatility of financial assets: see, e.g., Barndorff-Nielsen and Shephard (2001).

Suppose we are given equally spaced discrete-time observations $\{Y_0, Y_h, Y_{2h}, ..., Y_{nh}\}$ of $\mathcal{Y}$, where h is the time lag between two consecutive observations (e.g., $h = 1/250$ for daily observations). It can be shown that

$$Y_{th} = m(\theta)Y_{(t-1)h} + v(\theta)\epsilon_{th} \qquad t \in \mathbb{Z}, \tag{1.2}$$

where the ϵ_{th}'s are independently and identically distributed, with some probability density g,

$$m(\theta) = \exp\{-\theta h\}, \quad \text{and} \quad v^2(\theta) = (1 - \exp\{-2\theta h\})/2\theta. \tag{1.3}$$

In the classical case under which L_s is Brownian motion, ϵ_{th} is standard normal, $m(\theta)Y_{(t-1)h} = \mathrm{E}[Y_{th}|Y_{(t-1)h}]$ and $v^2(\theta) = \mathrm{Var}[Y_{th}|Y_{(t-1)h}]$ are the conditional mean and variance, respectively, of Y_{th}. Let us call this the *Gaussian case*.

In general, both the distribution of the Lévy process L_s and the value of θ enter, in a non trivial way, in the characterization of the discrete-time innovation density g, thereby generating a complex class of possible distributions. Several semiparametric extensions of the Gaussian case therefore have been considered in the literature: they all consider equation (1.2) with $m(\theta)$ and $v(\theta)$ as in (1.3), and independently and identically distributed ϵ_{th}'s having density g, where either

(i) g is in the family $\mathcal{G}_0$ of all nonvanishing densities ($g(z) > 0$ for all z),

(ii) g is in the family $\mathcal{G}_{\text{Wef}}$ of all densities (Wefelmeyer (1996)) with mean zero, variance one, and finite moments of order four,

(iii) g is in the family $\mathcal{G}_{\text{HKW1}}$ of all densities (Hallin et al. (2000)) with median zero and

$$\int_{-\infty}^{-1} g(z)dz = \int_{-1}^{0} g(z)dz = \int_{0}^{1} g(z)dz = \int_{1}^{\infty} g(z)dz = 1/4, \quad \text{or}$$

(iv) g is in the family $\mathcal{G}_{\text{HKW2}}$ of all densities (Hallin et al. (2000)) with median zero and

$$\int_{-\infty}^{0} g(z)dz = \int_{-1}^{1} g(z)dz = 1/2.$$

Let us call $\mathcal{E}_{(i)}, \ldots, \mathcal{E}_{(iv)}$, respectively, the resulting models. Note that model $\mathcal{E}_{(i)}$ contains the other three, and contains all discretized versions of the original process[4]. The standard Bickel et al. (1993) semiparametric approach moreover requires g to satisfy some regularity assumptions: g should have finite variance, and be absolutely continuous, with (almost everywhere) derivative $\dot{g}$, such that $\int_{-\infty}^{\infty}(\dot{g}(z)/g(z))^2 g(z)dz < \infty$ and $\int_{-\infty}^{\infty}(1+z\dot{g}(z)/g(z))^2 g(z)dz < \infty$—namely, g has finite Fisher information for location and for scale.

That standard semiparametric approach can be described in three steps. Step (a) consists in establishing the so-called ULAN (Uniform Local Asymptotic Normality) property for the fixed-g submodels. Under very general assumptions on the density g, this property indeed holds here, with a *central sequence* of the form

$$\Delta^{(n)}(\theta, g) = -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\frac{\partial_\theta m(\theta)}{v(\theta)}\frac{\dot{g}(\epsilon_{th}(\theta))}{g(\epsilon_{th}(\theta))}Y_{(t-1)h} + \frac{\partial_\theta v(\theta)}{v(\theta)}\left(1+\epsilon_{th}(\theta)\frac{\dot{g}(\epsilon_{th}(\theta))}{g(\epsilon_{th}(\theta))}\right)\right\}, \tag{1.4}$$

where $\epsilon_{th}(\theta) := (Y_{th} - m(\theta)Y_{(t-1)h})/v(\theta)$; see Proposition 3.1 for details.[5] Step (b) requires the theoretical derivation of the so-called tangent space projection of $\Delta^{(n)}(\theta, g)$, which yields $\Delta^{*(n)}(\theta, g)$ (the *semiparametrically efficient*, at g and θ, central sequence). Tangent space projections are model-specific, and their derivation in general is far from trivial. Finally, in step (c) those semiparametrically efficient central sequences are to be treated in the same way as ordinary central sequences—that is, in a point estimation context, essentially, as log-likelihood gradients, yielding estimating equations of the form $\Delta^{*(n)}(\theta, g) = 0$ or entering the construction of one-step solutions to the latter (see Section 4 for details).

Depending on the semiparametric model adopted, one obtains in step (b) the following results.

(i) For $\mathcal{E}_{(i)}$, the dependence on θ of the scale does not bring any information: the model is perfectly equivalent to an AR(1) model with autoregressive parameter $m(\theta)$ and unspecified innovation density. Those models are well known to be *adaptive*—that is, their semiparametrically efficient central sequences coincide (for all g and θ) with their "parametric" central sequences. As a result, we obtain here

$$\Delta^{*(n)}(\theta, g) = -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial_\theta m(\theta)}{v(\theta)}\frac{\dot{g}(\epsilon_{th}(\theta))}{g(\epsilon_{th}(\theta))}Y_{(t-1)h}. \tag{1.5}$$

[4] There is no guarantee, though, that for every density g in $\mathcal{G}_0$ (in $\mathcal{G}_{\text{Wef}}$, $\mathcal{G}_{\text{HKW1}}$, or $\mathcal{G}_{\text{HKW2}}$) there exists a Lévy process such that the discretized version (1.2) of $\mathcal{Y}$ has innovation density g.

[5] Note that the Gaussian QL is obtained as the solution of the Gaussian likelihood equation, here reducing to

$$\Delta^{(n)}(\theta, \phi) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\frac{\partial_\theta m(\theta)}{v(\theta)}\epsilon_{th}(\theta)Y_{(t-1)h} + \frac{\partial_\theta v(\theta)}{v(\theta)}\left(\epsilon_{th}^2(\theta)-1\right)\right\} = 0,$$

where ϕ as usual stands for the standard Gaussian density.

(ii) For $\mathcal{E}_{(ii)}$, we have

$$\Delta^{*(n)}(\theta, g) = -\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \frac{\partial_\theta m(\theta)}{v(\theta)} \frac{\dot{g}(\epsilon_{th}(\theta))}{g(\epsilon_{th}(\theta))} Y_{(t-1)h} + c_g^{-1}(\theta) \partial_\theta v(\theta) \big(v(\theta)(\epsilon_{th}^2(\theta) - v(\theta)\big) - \hat{\mu}_3 \epsilon_{th}(\theta) \right\} \tag{1.6}$$

with $c_g(\theta) := \big(\hat{\mu}_4 - v^2(\theta)\big) v(\theta) - \hat{\mu}_3^2$, $\hat{\mu}_3$ and $\hat{\mu}_4$ the empirical moments of order 3 and 4 of the $\epsilon_{th}(\theta)$'s.

(iii) For $\mathcal{E}_{(iii)}$, denoting by $\mathrm{E}_g(\cdot)$ expectation under g, we have

$$\begin{aligned} \Delta^{*(n)}(\theta, g) \;=\; & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \frac{\partial_\theta m(\theta)}{v(\theta)} \frac{\dot{g}(\epsilon_{th}(\theta))}{g(\epsilon_{th}(\theta))} [Y_{(t-1)h} - \mathrm{E}_g(Y_{(t-1)h})] \right. \\ & + 4 \frac{\partial_\theta v(\theta)}{v(\theta)} g(\operatorname{sgn}(\epsilon_{th}(\theta))) \operatorname{sgn}(\epsilon_{th}^2(\theta) - 1) \\ & + 2 \frac{\partial_\theta m(\theta)}{v(\theta)} g(0) \operatorname{sgn}(\epsilon_{th}(\theta)) \mathrm{E}_g[Y_{(t-1)h}] \\ & \left. + \frac{\partial_\theta m(\theta)}{v(\theta)} [4 g(\operatorname{sgn}(\epsilon_{th}(\theta))) - 2g(0)] \operatorname{sgn}(\epsilon_{th}^2(\theta) - 1) \operatorname{sgn}(\epsilon_{th}(\theta)) \mathrm{E}_g[Y_{(t-1)h}] \right\}. \end{aligned} \tag{1.7}$$

(iv) The result for $\mathcal{E}_{(iv)}$, with the same notation and $\delta := \int_{-1}^{0} g(z)dz - \int_{-\infty}^{-1} g(z)dz$, similarly follows:

$$\begin{aligned} \Delta^{*(n)}(\theta, g) \;=\; & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \frac{\partial_\theta m(\theta)}{v(\theta)} \frac{\dot{g}(\epsilon_{th}(\theta))}{g(\epsilon_{th}(\theta))} [Y_{(t-1)h} - \mathrm{E}_g(Y_{(t-1)h})] \right. \\ & + 4 \frac{\partial_\theta v(\theta)}{v(\theta)} \; \frac{\frac{1}{2}(g(1) + g(-1)) \operatorname{sgn}(\epsilon_{th}^2(\theta) - 1) - \delta(g(1) + g(-1)) \operatorname{sgn}(\epsilon_{th}(\theta))}{1 - 4\delta^2} \\ & + 2 \frac{\partial_\theta m(\theta)}{v(\theta)} \; \frac{g(0) - 2\delta(g(1) - g(-1))}{1 - 4\delta^2} \operatorname{sgn}(\epsilon_{th}(\theta)) \mathrm{E}_g(Y_{(t-1)h})] \\ & \left. + \frac{\partial_\theta m(\theta)}{v(\theta)} \; \frac{2(g(1) - g(-1)) - 4g(0)}{1 - 4\delta^2} \operatorname{sgn}(\epsilon_{th}^2(\theta) - 1) \mathrm{E}_g(Y_{(t-1)h})] \right\}. \end{aligned} \tag{1.8}$$

This calls for several immediate remarks. First, one directly sees that the semiparametrically efficient central sequences $\Delta^{*(n)}(\theta, g)$ in equations (1.6)-(1.8) are considerably more complicated than $\Delta^{(n)}(\theta, g)$. Their derivation, moreover, is model-specific, hence has to be performed on a case-by-case basis. Second, semiparametrically efficient central sequences depend on g and its derivative $\dot{g}$, which are both unknown. For $f \neq g$, typically, $\mathrm{E}_g[\Delta^{*(n)}(\theta, f)] \neq 0$ (violating the so-called Fisher consistency condition), which implies that estimators based on $\Delta^{*(n)}(\theta, f)$ are not root-n consistent. In order to restore root-n consistency, kernel estimates of both g and $\dot{g}$ have to be computed and plugged-in into $\Delta^{*(n)}(\theta, g)$, yielding $\Delta^{*(n)}(\theta, \hat{g}^{(n)})$, on which (step (c)) standard semiparametric estimators are based. This implies careful bandwidth selection and some additional niceties such as sample splitting. Moreover, kernel estimation of g and $\dot{g}$ is unlikely to produce good results in small and moderately large samples. Third, the semiparametric extensions considered in *(i)-(iv)* all are equally plausible, offering little guidelines for choosing any one of them rather than the other: $\mathcal{E}_{(i)}$ is quite general, but does not exploit the dependence on θ of the scale; $\mathcal{E}_{(ii)}$ requires

finite fourth-order moments; $\mathcal{E}_{(iii)}$ and $\mathcal{E}_{(iv)}$ only require second-order moments, but $m(\theta)$ and $v^2(\theta)$ are losing their interpretations in terms of conditional mean and variance. On top of that, if the actual model lies in $\mathcal{E}_{(j)}$ but not in $\mathcal{E}_{(j')}$ $(j, j' = (i), \ldots, (iv))$, the semiparametrically efficient central sequence associated with $\mathcal{E}_{(j')}$ again is losing Fisher consistency. The choice of the "right" semiparametric extension thus is both crucial and problematic, the only "riskless choice" being that of $\mathcal{E}_{(i)}$.

1.4. R-estimation: an alternative semiparametric approach

1.4.1. Overview

The objective of this paper is to show that another semiparametric approach, based on residual ranks (the ranks of the ϵ_{th}'s), is possible, which avoids the derivation of complicated tangent space projections, does not require estimating any density function g, and remains valid under minimal regularity assumptions (those guaranteeing finite Fisher information and ULAN). Moreover, simple data-driven scores (accounting, for instance, for actual skewness and kurtosis) can be used, allowing for much flexibility in the tuning of asymptotic performances and improving a lot over the Gaussian methods.

Essentially, our methodology proceeds along the same steps as in the standard semiparametric approach, with two fundamental differences at step (b). First, a reference density f (rather than the actual density g) is adopted to derive the central sequence $\mathbf{\Delta}^{(n)}(\theta, f)$. Second, $\mathbf{\Delta}^{(n)}(\theta, f)$ is projected onto the σ-field generated by the ranks of the ϵ_{th}'s (rather than projected along the tangent spaces). So, in a nutshell, our method consists in the following three steps: (a') establishing ULAN, with central sequence $\mathbf{\Delta}^{(n)}(\theta, g)$, for all $g \in \mathcal{G}$ (where $\mathcal{G} \subset \mathcal{G}_0$ contains all densities satisfying the regularity assumptions required for ULAN to hold); (b') choosing some reference $f \in \mathcal{G}$ and projecting $\mathbf{\Delta}^{(n)}(\theta, f)$ onto the σ-field generated by the ranks of the ϵ_{th}'s—thus obtaining the so-called rank-based central sequence $\underset{\sim}{\mathbf{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$; (c') based on $\underset{\sim}{\mathbf{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$ rather than $\Delta^{*(n)}(\theta, \hat{g}^{(n)})$, constructing a root-$n$ consistent and asymptotically normal one-step[6] R-estimator.

The resulting R-estimators offer a number of advantages. Essentially, and irrespective of the choice of f, their validity (root-n consistency and asymptotic normality) extends to any density g in $\mathcal{G}$. In particular, in sharp contrast with Gaussian QL estimators, no finite moments are required beyond those ensuring finite Fisher information (see e.g. Drost and Klaassen (1997) for ARCH-type processes under Cauchy innovations or, in a slightly different context, Hallin et al. (2013) for R-estimation of regression under arbitrary stable noise); moreover, unlike standard semiparametric estimators, no estimation of the actual innovation density g is necessary. Performances (under g), of course, depend on the selected reference density f: the "closer" to g, the better. The choice of f can be made by the econometrician according to her/his preferences or past experience; it also can be data-driven as soon as it only depends on the order statistic of

[6] For a general description of one-step estimators, see, for instance, Section 5.7 of van der Vaart (1998).

the ϵ_{th}'s. For instance, letting $f = \hat{g}^{(n)}$, where $\hat{g}^{(n)}$ is a kernel estimator of g, yields an R-estimator which is semiparametrically efficient under any $g \in \mathcal{G}$, thus matching the performance of standard semiparametric estimation in the Bickel et al. style; contrary to the latter, it does not require sample splitting, tough, thanks to the independence between the ranks and the order statistic. Computationally less demanding choices, avoiding kernel density estimation, are also possible; for instance, a data-driven reference density f belonging to the family of skew-t densities can be obtained by estimating (via skew-t maximum likelihood) a degree of freedom and a skewness coefficient. Although the resulting f does not attempt to recover the actual g, it does account for its skewness and kurtosis; see Section 4.3 for details.

Finally, in many cases (like e.g. the AR process applied in the motivating example), the R-estimator associated with a standard Gaussian reference density (that is $f = \phi$) uniformly outperfoms the Gaussian QL estimator (the so-called Chernoff-Savage property discussed in Hallin (1994)).

Other attempts have been made to introduce R-estimation in the context of time series models: see, among 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, mostly apply to ARMA models. Moreover, they rely on an extension of the method introduced by Jaeckel (1972) for linear regression with independent observations. Contrary to the original Hodges and Lehmann (1956) definition, Jaeckel's R-estimators are based on somewhat hybrid objective functions which combine the residual ranks and the residuals themselves. In the time series settings considered in this paper, Jaeckel-type objective functions do not follow from any solid decision-theoretic invariance argument, and their equivalence to the Hodges-Lehmann approach is unlikely to hold. In contrast to the latter, our R-estimators are genuinely rank-based (measurable with respect to the σ-field generated by the ranks)[7], and have a clear link with invariance, hence with semiparametric efficiency: see Hallin and Werker (2003).

1.4.2. A motivating example (continued)

We conclude the motivating example of Section 1.3 by showing how our rank-based procedures apply and yield a root-n consistent and asymptotically normal estimator of the parameter θ in (1.1), even in the presence of misspecified innovation density ($f \neq g$). The central sequence (1.4), at reference density f with cumulative distribution function F, takes the form (writing $\epsilon_t(\theta)$ for $\epsilon_{th}(\theta) := (Y_{th} - m(\theta)Y_{(t-1)h})/v(\theta)$)

$$\Delta^{(n)}(\theta, f) = \frac{1}{\sqrt{n}} \left[c_1(\theta) \sum_{i=1}^{n} \frac{\dot{f}(\epsilon_{th}(\theta))}{f(\epsilon_{th}(\theta))} Y_{(t-1)h} + c_2(\theta) \sum_{t=1}^{n} \left(1 + \epsilon_{th}(\theta) \frac{\dot{f}(\epsilon_{th}(\theta))}{f(\epsilon_{th}(\theta))} \right) \right],$$

[7] Allal et al. (2001) also propose genuinely rank-based estimators, but they overlook the cross-information problem. We are addressing this issue in C, which contains Assumption (G).

with

$$c_1(\theta) = \frac{\sqrt{2}h\exp\{-\theta h\}}{((1-\exp\{-2\theta h\})/2\theta)^{1/2}} \quad \text{and} \quad c_2(\theta) = \frac{1-\exp\{2\theta h\}+2\theta h}{2\theta(1-\exp\{2\theta h\})}.$$

Using the standard representation of $Y_{(t-1)h}$ as a sum of past shocks and projecting that central sequence onto the ranks, we obtain

$$\underset{\sim}{\Delta}^{(n)}(\theta,f) = \sqrt{n}\sum_{i=1}^{n-1}\exp\{-\theta h(i-1)\}\underset{\sim}{r}^{(n)}_{f,i} \quad \text{with} \quad \underset{\sim}{r}^{(n)}_{f,i} = \frac{1}{s^{(n)}}\left[\frac{1}{n-i}\sum_{t=i+1}^{n}\frac{\dot{f}\left(F^{-1}\left(\frac{R^{(n)}_t}{n+1}\right)\right)}{f\left(F^{-1}\left(\frac{R^{(n)}_t}{n+1}\right)\right)}F^{-1}\left(\frac{R^{(n)}_{t-i}}{n+1}\right) - m^{(n)}\right],$$

where $R^{(n)}_t$ denotes the rank of $\epsilon^{(n)}_{th}(\theta)$ among $\epsilon^{(n)}_h(\theta),\ldots,\epsilon^{(n)}_{nh}(\theta)$, and $m^{(n)}$ and $s^{(n)}$ are exact standardizing constants[8]; see Section 3.3 and Section 5.2 for details. In the particular case of a Gaussian reference density, $\underset{\sim}{r}^{(n)}_{f,i}$, for $f=\phi$, takes the form of a *van der Waerden autocorrelation coefficient*[9]

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

where $m^{(n)}_{\mathrm{vdW}}$, being $o(n^{-1})$ can be omitted and $s^{(n)}_{\mathrm{vdW}} = \frac{1}{n}\sum_{j=1}^{n}\left(\Phi^{-1}\left(\frac{j}{n+1}\right)\right)^2 + O(n^{-1})$ can be replaced with $\frac{1}{n}\sum_{j=1}^{n}\left(\Phi^{-1}\left(\frac{j}{n+1}\right)\right)^2$ as $\underset{\sim}{\Delta}^{(n)}$ only needs to be defined up to $o_{\mathrm{P}}(1)$ quantities.

The major advantage of $\underset{\sim}{\Delta}^{(n)}(\theta,f)$ over $\Delta^{(n)}(\theta,f)$ is that its Fisher consistency is robust to misspecification: while $\Delta^{(n)}(\theta,f)$ does not have expectation zero under density g unless $f=g$, the expectation of $\underset{\sim}{\Delta}^{(n)}(\theta,f)$ remains zero for $f\neq g$; hence estimators derived from $\underset{\sim}{\Delta}^{(n)}(\theta,f)$, contrary to those derived from $\Delta^{*(n)}(\theta,g)$, remain root-n consistent and asymptotically normal even if $f\neq g$.

1.5. Outline of the paper

In Sections 2 through 5, we provide a precise description of our method, and illustrate its implementation in a variety of widely-applied econometric models. Section 6 presents the outcomes of several numerical exercises. First, in the context of normal variance-mean mixture models for return and realized volatility dynamics, we conduct a comparative analysis of the asymptotic relative efficiencies, with respect to QL estimators, of various R-estimators. Then, we perform a Monte Carlo study of several estimators of an ARCH(1) model under skew and leptokurtic g's, providing numerical evidence that rank-based procedures do improve on the accuracy of traditional Gaussian QL methods. Finally, in Section 7, a real-data analysis about

[8] Since the vector of ranks, under parameter value θ and irrespective of the actual density (be it g or f), is uniformly distributed over the $n!$ permutations of $(1,\ldots,n)$, those quantities do not depend on g and are easily computed: see (2.8) and (2.9) in Hallin and Mélard (1988) for explicit values. Also note that the ranks of $\epsilon_{th}(\theta)$ coincide with those of $\epsilon^{\dagger}_{th}(\theta) = Y_{th} - m(\theta)Y_{(t-1)h}$; see Example (d) in Section 5.2 for a similar argument.

[9] As usual, Φ stands for the standard normal distribution function.

modeling and forecasting of log-return and log-transformed realized volatility in the USD/CHF exchange rate brings evidence that the forecasts obtained from our rank-based method outperform the classical ones, based on the routinely-applied Gaussian QL approach. Section 8 concludes. Proofs and technical details are concentrated in A, B and D.

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 $\mathbf{Y} := \{Y_t; t \in \mathbb{Z}\}$ 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 $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ of interest; $\{\varepsilon_t; t \in \mathbb{Z}\}$ is an independently and identically distributed (i.i.d.) process with unspecified (within some family $\mathcal{G}$) density g; ε_t and $Y_{t'}$ are mutually independent for all $t > t'$.

Equation (2.1), along with a family $\mathcal{G}$ of densities, characterizes a semiparametric extension of the Gaussian model where the ε_t's are i.i.d. standard normal. We have shown, in Section 1.3, how different (and mostly arbitrary) choices of $\mathcal{G}$ may lead to different tangent space projections and definitions of semiparametric efficiency; in the sequel, although semiparametric efficiency is not emphasized, we only consider for $\mathcal{G}$ the most general choice $\mathcal{G}_0$ of the family of all nonvanishing densities over the real line.

The interpretation of $m(\boldsymbol{y}, \boldsymbol{\theta})$ and $v(\boldsymbol{y}, \boldsymbol{\theta})$ depends on g: 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) also 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.

2.2. Assumptions and notation

Throughout, we assume that the functions m and v in (2.1) are specified, but not the parameter of interest $\boldsymbol{\theta}$ nor the density g. Denoting by $\mathrm{P}^{(n)}_{\boldsymbol{\theta},g}$ the joint distribution, under (2.1), of $\mathbf{Y}^{(n)}$, we thus consider the (sequence of) 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 $\dot{\boldsymbol{m}}(\boldsymbol{y}, \boldsymbol{\theta}) := \text{grad}_{\boldsymbol{\theta}}\, m(\boldsymbol{y}, \boldsymbol{\theta})$ and $\dot{\boldsymbol{v}}(\boldsymbol{y}, \boldsymbol{\theta}) := \text{grad}_{\boldsymbol{\theta}}\, v(\boldsymbol{y}, \boldsymbol{\theta})$. Moreover, denoting by $\text{E}_{\boldsymbol{\theta},g}$ expectations under $\text{P}^{(n)}_{\boldsymbol{\theta},g}$, both $\text{E}_{\boldsymbol{\theta},g}[\dot{\boldsymbol{m}}(\mathbf{Y}_{t-1}, \boldsymbol{\theta})]$ and $\text{E}_{\boldsymbol{\theta},g}[\dot{\boldsymbol{v}}(\mathbf{Y}_{t-1}, \boldsymbol{\theta})]$ exist and are finite.

Assumption (B). (B1) For all $x \in \mathbb{R}$, the density $g(x)$ is strictly positive.

(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 $I_{12}(g) = I_{21}(g) := \int x\phi_g^2(x)g(x)dx$ also 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 $\{\text{P}^{(n)}_{\boldsymbol{\theta},g}\}$ and write $\mathcal{H}^{(n)}_{\boldsymbol{\theta}}$ for the nonparametric collection $\{\text{P}^{(n)}_{\boldsymbol{\theta},g} : g \in \mathcal{G}\}$. 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 $\mathcal{H}^{(n)}_{\boldsymbol{\theta}}$ and $\mathcal{H}^{(n)}_{\boldsymbol{\theta},g}$ 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 $\mathcal{P}^{(n)}_g := \{\text{P}^{(n)}_{\boldsymbol{\theta},g} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$. Then, 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

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

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 3.1. *(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 $\boldsymbol{\tau}_n \in \mathbb{R}^p$,*

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

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

3.2. Parametric and semiparametric efficiency bounds

The inverse $\boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta}, g)$ of $\boldsymbol{\Gamma}(\boldsymbol{\theta}, g)$ settles the parametric efficiency bound at g—the "best asymptotically achievable" covariance for a regular estimator of $\boldsymbol{\theta}$ in the parametric model where g is specified: an estimator reaching that bound then can be based on $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g)$ either by solving the likelihood equation $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g) = 0$, or as a $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}^{(n)}, g)$-based one-step update of some preliminary root-n consistent estimator $\hat{\boldsymbol{\theta}}^{(n)}$ (see Section 4.2 for details).

Parametric efficiency, in general, cannot be reached anymore in the semiparametric model where g remains unspecified, and the best one can go for is the semiparametric efficiency bound $\boldsymbol{\Gamma}^{*-1}(\boldsymbol{\theta}, g)$. The *semiparametrically efficient central sequence*, denoted as $\boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, g)$, is the tool one needs to construct estimators that reach that semiparametric efficiency bound; see e.g. Newey (1990). The semiparametrically efficient central sequence is obtained by projecting the central sequence $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g)$ along the so-called *tangent space*. We refer to the monograph by Bickel et al. (1993) for a full exposition in the i.i.d. framework and to Drost et al. (1997) for time series models.

It has been shown in Hallin and Werker (2003) that, for a very broad class of models (including those considered here), projecting $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)$ onto the σ-field generated by the ranks of the residuals $Z_1(\boldsymbol{\theta}), \ldots, Z_n(\boldsymbol{\theta})$ (projection here is to be understood as conditional expectation) 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 $\mathrm{P}^{(n)}_{\boldsymbol{\theta}, f}$,

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

Obviously, $\mathrm{E}_f[\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)] = 0$ (expectation of a conditional expectation). The distribution-freeness of ranks then entails that also $\mathrm{E}_g[\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)] = 0$ (Fisher consistency) for any g: therefore, estimators based on $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$ unlike those based on $\boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, f)$, remain root-$n$ consistent under any $\mathrm{P}^{(n)}_{\boldsymbol{\theta}, g}$. Those estimators

are reaching the semiparametric efficiency bound associated with f if g and f coincide—we say that they are *semiparametrically efficient at* f.

3.3. Rank-based central sequence

Let us provide some details on the rank-based approach just described. The reader can find the technicalities (along with Assumptions (D) and (E)) in B, while several illustrations are available in Section 5.

Let $f \in \mathcal{G}$ be a chosen reference density, with distribution function F. Denote by $\boldsymbol{R}^{(n)}(\boldsymbol{\theta})$ the vector $(R_1^{(n)}(\boldsymbol{\theta}), \ldots, R_n^{(n)}(\boldsymbol{\theta}))$ of *residual ranks*, where $R_t^{(n)}(\boldsymbol{\theta})$ is the rank of $Z_t^{(n)}(\boldsymbol{\theta})$ among $Z_1^{(n)}(\boldsymbol{\theta}), \ldots, Z_n^{(n)}(\boldsymbol{\theta})$. For notational convenience, dropping the dependence on $\boldsymbol{\theta}$ or n, we also write $\boldsymbol{R}^{(n)}$, $R_t^{(n)}$ and $Z_t^{(n)}$, or $\boldsymbol{R}(\boldsymbol{\theta})$, $R_t(\boldsymbol{\theta})$ and $Z_t(\boldsymbol{\theta})$ when no confusion is possible.

The rank-based central sequences $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$ we eventually will work with are very simply obtained under the so-called *approximate score form* (see Hallin and Werker (2003) and reference therein). Approximate score forms are derived by substituting $F^{-1}(R_t/(n+1))$ for Z_t in $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)$. An intuitive and heuristic justification of this method is that, when $Z_1, \ldots, Z_n$ are i.i.d., with nonvanishing density f, hence strictly monotone increasing 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 $[0,1]$.

However, in order to derive $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$, we first need rewriting the central sequence $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)$ as a function of the present and a finite number s_n ($s_n \to \infty$ as $n \to \infty$) of past residuals. Then, we consider the approximation $\dot{\boldsymbol{\imath}}^{s_n}(Z_t, \mathbf{Z}_{t-1}^{s_n}, \boldsymbol{\theta}, f)$ of $\dot{\boldsymbol{\imath}}(Z_t, \mathbf{Z}_{t-1}, \boldsymbol{\theta}, f)$ which is obtained by replacing $\mathbf{Z}_{t-1}$ with the truncated version $\mathbf{Z}_{t-1}^{s_n} := (Z_{t-1}, \ldots, Z_{t-\min(t-1,s_n)}, 0, 0, \ldots)$. This approximation is possible in most stationary Markov processes of order p and q-dependent processes—see Section 5. Finally, denoting by $\mathcal{B}^{(n)}(\boldsymbol{\theta})$ the σ-field generated by $\boldsymbol{R}^{(n)}(\boldsymbol{\theta})$, we define

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f) := \mathrm{E}_{\boldsymbol{\theta},f}\left[\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)\middle|\mathcal{B}^{(n)}(\boldsymbol{\theta})\right] = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathrm{E}_{\boldsymbol{\theta},f}\left[\dot{\boldsymbol{\imath}}^{s_n}(Z_t, \mathbf{Z}_{t-1}^{s_n}, \boldsymbol{\theta}, f)\middle|\mathcal{B}^{(n)}(\boldsymbol{\theta})\right], \tag{3.5}$$

and

$$\mathrm{E}_{\boldsymbol{\theta},f}\left[\dot{\boldsymbol{\imath}}^{s_n}(Z_t, \mathbf{Z}_{t-1}^{s_n}, \boldsymbol{\theta}, f)\middle|\mathcal{B}^{(n)}(\boldsymbol{\theta})\right] \approx \dot{\boldsymbol{\imath}}^{s_n}\left(F^{-1}\left(\frac{R_t^{(n)}}{n+1}\right), \ldots, F^{-1}\left(\frac{R_{t-s}^{(n)}}{n+1}\right), \boldsymbol{\theta}, f\right)$$

Since $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$ is $\boldsymbol{R}^{(n)}(\boldsymbol{\theta})$-measurable, it is distribution-free: let $\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}(\boldsymbol{\theta}, f)$ be its covariance matrix under $\mathcal{H}_{\boldsymbol{\theta}}^{(n)}$ (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).

Proposition 3.2. *Let Assumptions (A)-(E) be satisfied. Denote by* $\boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, f)$ *a semiparametrically effi-*

cient central sequence for $\mathcal{P}_f^{(n)}$*, and by* $\mathbf{\Gamma}^{*(n)}(\boldsymbol{\theta}, f)$ *its covariance matrix under* $\mathrm{P}_{\boldsymbol{\theta},f}^{(n)}$*. Then,*

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

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

The asymptotic equivalence, under $\mathrm{P}_{\boldsymbol{\theta},f}^{(n)}$, of $\mathbf{\Delta}^{*(n)}(\boldsymbol{\theta}, f)$ and $\underset{\sim}{\mathbf{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$ implies that the latter can be considered a rank-based version of the same semiparametrically efficient (at f) central sequence. In the sequel, for the sake of simplicity, we call $\underset{\sim}{\mathbf{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$ a *rank-based central sequence* instead of a *rank-based semiparametrically efficient (at* f*) central sequence*. Unlike its traditional counterpart, the rank-based central sequence does not require the nontrivial exercise of deriving the tangent spaces and corresponding projections.

4. R-estimation

4.1. Estimating equations

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

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

provides a test of the null hypothesis $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ (with unspecified g); that test is is locally and asymptotically optimal against $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ alternatives with density f. Therefore, it would be natural to define an R-estimator of $\boldsymbol{\theta}$ as the minimizer, with respect to $\boldsymbol{\theta}$, of $\underset{\sim}{Q}_{\mathrm{HL}}^{(n)}(\boldsymbol{\theta}, f) := \underset{\sim}{\mathbf{\Delta}}^{(n)\prime}(\boldsymbol{\theta}, f)\mathbf{\Gamma}^{*-1}(\boldsymbol{\theta}, f)\underset{\sim}{\mathbf{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$. Despite its simplicity and intuitive appeal, this definition, which in a much simpler context goes back to Hodges and Lehmann (1956), often runs into serious numerical difficulties. The non-convex form of $\boldsymbol{\theta} \mapsto \underset{\sim}{Q}_{\mathrm{HL}}^{(n)}(\boldsymbol{\theta}, f)$ indeed results in practical implementation problems such as multiple solutions and local minima, especially when the dimension of the parameter θ gets large.

Those difficulties have been avoided, in the context of linear models with independent observations, by Jaeckel (1972) who observed that the minimizer of $\underset{\sim}{Q}_{\mathrm{HL}}^{(n)}(\boldsymbol{\theta}, f)$ is asymptotically the same as that of another quadratic form, $\underset{\sim}{Q}_{\mathrm{J}}^{(n)}(\boldsymbol{\theta}, f)$, 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 (asymptotic equivalence with Hodges and Lehmann's) does not hold anymore. Instead, we consider here a one-step version of the minimization of $\underset{\sim}{Q}_{\mathrm{HL}}^{(n)}(\boldsymbol{\theta}, f)$, inspired from Le Cam's one-step estimation method.

4.2. 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*[10], 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 LeCam and Yang (1990)) and (since part (iii) of Assumption (F), under $g=f$, follows from ULAN) can be considered a particular case of Proposition 4.2.

Proposition 4.1. *Let Assumptions (A)-(E) hold. The one-step R-estimator*

$$\underset{\sim}{\boldsymbol{\theta}}_f^{(n)} := \hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\big(\widehat{\boldsymbol{\Gamma}}_f^{(n)}\big)^{-1}\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\hat{\boldsymbol{\theta}}^{(n)},f) \tag{4.1}$$

under $\mathrm{P}_{\boldsymbol{\theta},f}^{(n)}$ *root-n consistent and asymptotically normal, with* $n^{1/2}(\underset{\sim}{\boldsymbol{\theta}}_f^{(n)}-\boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}\big(\mathbf{0},\boldsymbol{\Gamma}^{*-1}(\boldsymbol{\theta},f)\big)$.

It follows that the R-estimator $\underset{\sim}{\boldsymbol{\theta}}^{(n)}$ in (4.1) 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 $\mathrm{P}_{\boldsymbol{\theta},g}^{(n)}$, as $n\to\infty$,

(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*

$$\boldsymbol{\Gamma}(\boldsymbol{\theta},f,g) := \lim_{n\to\infty} \mathrm{E}_{\boldsymbol{\theta},g}\left[\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta},f)\Big(\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta},g)\Big)'\right], \tag{4.2}$$

(in Section C, we explain how to construct such estimators), and

(iii) g is such that (asymptotic linearity) $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau},f)-\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta},f) = -\boldsymbol{\Gamma}(\boldsymbol{\theta},f,g)\boldsymbol{\tau}+o_{\mathrm{P}}(1)$.[11]

We then have, for the one-step R-estimator (4.1), the following result (see A for a proof):

Proposition 4.2. *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 4.2 implies that $\underset{\sim}{\boldsymbol{\theta}}_f^{(n)}$ remains root-n consistent for any (f,g) such that part (iii) of Assumption (F) holds; in that sense, thus, unlike most M- and L-estimators, our R-estimators are robust to model misspecification.

[10] *Asymptotic discreteness* is easily obtained via discretization. It should be insisted, however, that such discretization is only required in the statement and proof of asymptotic properties. In applications, there is no point in implementing it: indeed, $\hat{\boldsymbol{\theta}}^{(n)}$ in practice always has a finite number of digits; see pages 125 and 188 of LeCam and Yang (1990).

[11] Note that, for $f=g$, $\boldsymbol{\Gamma}(\boldsymbol{\theta},f,f)=\boldsymbol{\Gamma}^*(\boldsymbol{\theta},f)=\boldsymbol{\Gamma}^*(\boldsymbol{\theta},g)$.

4.3. Implementation issues

4.3.1. Choosing a preliminary estimator

A possible candidate for $\hat{\boldsymbol{\theta}}^{(n)}$ in part (i) of Assumption (F)—provided that one is willing to assume finite fourth-order moments—is the Gaussian QL estimator. More robust alternatives are highly recommended, though, such as the LAD estimator of Peng and Yao (2003) for ARCH and GARCH-type models, the non-Gaussian QL estimator introduced in Fan et al. (2014) or, in the presence of outliers and data contamination, the bounded-influence estimators by Mancini et al. (2005).

Clearly, different preliminary estimators yield different R-estimators. That impact is limited, though. In practice, indeed, the one-step update of $\hat{\boldsymbol{\theta}}^{(n)}$ is to be iterated ($\underset{\sim}{\boldsymbol{\theta}}_f^{(n)}$being used as the preliminary estimator in a further one-step update) until it stabilizes numerically. Such iterations do not modify the asymptotic behavior of the R-estimator, but they do improve on its finite-sample performances: this is in accordance with traditional Newton-Raphson practice. Numerically, the iterations are roughly equivalent to solving–locally rather than globally–the Hodges-Lehmann estimating equation (see Section 4.1); the role of $\hat{\boldsymbol{\theta}}^{(n)}$ thus essentially consists in selecting an initial value in a root-n neighborhood of the actual value of $\boldsymbol{\theta}$.

4.3.2. Choosing the score (the reference density): a data-driven approach

While the choice of the reference density f has no impact on the consistency properties of the corresponding R-estimator, it has a direct influence on its performances, both for finite n as for $n \to \infty$; the "closer" f is to the actual density g, the better the performance for $\underset{\sim}{\boldsymbol{\theta}}_f^{(n)}$. The efficiency loss due to a misspecified reference density f is revealed though an inspection of the (f, g)-cross-information quantities and their comparison with the corresponding g-information ones.

An important advantage of R-estimation over all other methods is that the selection of f can be data-driven as long as it is based on the order statistic of the residuals. The ranks and the order statistic indeed are mutually independent, which implies that the asymptotic results on the behavior of rank-based statistics remain conditionally (on the order statistic) valid: order statistic-driven scores therefore can be treated as if they were deterministic and not bearing any relation to the observations.

Asymptotically optimal choices of f, in that respect, are the many possible (order statistic-based) kernel estimators of g—which moreover do not require any sample-splitting precautions. As already explained, such choice is of theoretical rather than practical interest, and the fact that estimating g is compulsory in the standard semiparametric approach of Bickel et al. (1993) is one of its main drawbacks. A distinctive feature of R-estimators is the possibility of a much more flexible selection of f. For instance, in the spirit of Dodge and Jurečková (2000), we propose selecting f by fitting a parametric density to the (order statistic of the) residuals associated with the preliminary estimator. If skewness and kurtosis are to be accounted for, a convenient family of densities is the family of skew-t distribution (Azzalini and Capitanio, 2003) with

densities of the form

$$h_{\boldsymbol{\omega}}(x) = \frac{2}{\sigma} t_\nu(z) T_{\nu+1}\Big(\alpha z\Big(\frac{\nu+1}{\nu+z^2}\Big)^{1/2}\Big) \qquad \text{for } x \in \mathbb{R} \text{ and } z := \sigma^{-1}(x-\mu), \tag{4.3}$$

indexed by $\boldsymbol{\omega} := (\mu, \sigma, \alpha, \nu)$, where $\mu \in \mathbb{R}$ is a location, $\sigma \in \mathbb{R}_0^+$ a scale, $\alpha \in \mathbb{R}$ a skewness parameter, and $\nu > 0$ the number of degrees of freedom governing the tails; $t_\nu(z)$ and $T_\nu(z)$ are the density and cumulative distribution functions, respectively, of the traditional Student distribution with ν degrees of freedom; see Hallin and Mehta (2015).[12] Other parametric families of course can be considered, very much in the same way, such as the stable family (as in Hallin et al. (2013)), or the so-called *skew generalized error* family (see e.g. Hansen et al. (2007)).

4.3.3. Estimation of cross-information quantities

Part (ii) of Assumption (F), dealing with the estimation of the cross-information matrix $\boldsymbol{\Gamma}(\boldsymbol{\theta}, f, g)$ in (4.2), is discussed in C, where an easy-to-implement estimator is proposed.

5. Theoretical examples

The derivation of the rank-based central sequence is pivotal for our method. In this section we explain how to derive $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$ in some widely-applied econometric models.

5.1. Discrete-time models

5.1.1. Conditional heteroskedasticity models

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

$$Y_t = \Big(1 + \sum_{j=1}^{q} \theta_j Y_{t-1}^2\Big)^{1/2} \epsilon_t, \tag{5.1}$$

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

$$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\psi_g\left(Z_t(\boldsymbol{\theta})\right)}{1 + \sum_{j=1}^q \theta_j Y_{t-j}^2} \begin{pmatrix} Y_{t-1}^2 \\ \vdots \\ Y_{t-q}^2 \end{pmatrix}, \tag{5.2}$$

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 (5.2) to be expressed in terms of a finite number of past

[12] By using MLE to estimate the skew-t parameters, we are in fact minimizing the Kullback-Leibler divergence and selcting the misspecified skew-t density which is the closest to the actual g; see e.g. White (1982) for a related discussion.

shocks. In Appendix D.1, we show that this is possible via a Volterra series expansion. Here we 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 yields the n-tuple $Z_1^{(n)}(\boldsymbol{\theta}), ..., Z_n^{(n)}(\boldsymbol{\theta})$, with ranks $R_1^{(n)}(\boldsymbol{\theta}), ..., R_n^{(n)}(\boldsymbol{\theta})$. Next consider the problem of a reconstruction of $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)$ (f some chosen reference density) based on those ranks. Since, by definition, $Y_1 = Z_1^{(n)}(\boldsymbol{\theta})$, set $\underset{\sim}{Y}_1 := F^{-1}(R_1^{(n)}(\boldsymbol{\theta})/(n+1))$; start the recurrence

$$\underset{\sim}{Y}_t := \Big(1 + \sum_{j=1}^{q} \theta_j \underset{\sim}{Y}_{t-j}^2\Big)^{1/2} F^{-1}\left(\frac{R_t^{(n)}(\boldsymbol{\theta})}{n+1}\right), \quad t \geq 2, \tag{5.3}$$

then define

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\psi_f\Big(F^{-1}\Big(\frac{R_t^{(n)}(\boldsymbol{\theta})}{n+1}\Big)\Big) - m_{f,(2)}^{(n)}}{1 + \sum_{j=1}^{q} \theta_j \underset{\sim}{Y}_{t-j}^2} \begin{pmatrix} \underset{\sim}{Y}_{t-1}^2 \\ \vdots \\ \underset{\sim}{Y}_{t-q}^2 \end{pmatrix} \tag{5.4}$$

where

$$m_{f,(2)}^{(n)} := \frac{1}{n} \sum_{i=1}^{n} \psi_f\Big(F^{-1}\Big(\frac{i}{n+1}\Big)\Big). \tag{5.5}$$

From the re-centering Lemma 1 in Appendix B.2, it follows that $m_{f,(2)}^{(n)}$ is $o\big(n^{-1/2}\big)$. The asymptotic covariance $\boldsymbol{\Gamma}^*(\boldsymbol{\theta}, f)$ of $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$ 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

$$Y_t = \sum_{j=1}^{p} \vartheta_j Y_{t-j} + \left(1 + \sum_{l=1}^{q} \beta_l Y_{t-l}\right) \epsilon_t, \tag{5.6}$$

where the ϵ_t's are i.i.d. with standardized density g, and $\boldsymbol{\theta} = (\vartheta_1, ..., \vartheta_p, \beta_1, ..., \beta_q)$. If the conditions of Theorem 2.1 in Francq and Zakoïan (2010) are satisfied (see D.2), the ULAN central sequence for $\boldsymbol{\theta}$ reads

$$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Big(1 + \sum_{l=1}^{q} \beta_l Y_{t-l}\Big)^{-1} \begin{pmatrix} -\phi_g\left(Z_t(\boldsymbol{\theta})\right) \begin{pmatrix} Y_{t-1} \\ \vdots \\ Y_{t-p} \end{pmatrix} \\ \psi_g\left(Z_t(\boldsymbol{\theta})\right) \begin{pmatrix} Y_{t-1} \\ \vdots \\ Y_{t-q} \end{pmatrix} \end{pmatrix}. \tag{5.7}$$

A rank-based central sequence for reference density f is obtained by replacing, in (5.7), the residual $Z_t(\boldsymbol{\theta})$ by $F^{-1}(R_t^{(n)}(\boldsymbol{\theta})/(n+1))$, for every t.

We illustrate that construction in the AR(1)-LARCH(1) case, with dynamics

$$Y_t = \vartheta Y_{t-1} + (1+\beta Y_{t-1})\epsilon_t \qquad t \in \mathbb{Z}, \tag{5.8}$$

which is ULAN with central sequence

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

where $Z_t = Z_t(\boldsymbol{\theta}) := (Y_t - \vartheta Y_{t-1})/(1+\beta Y_{t-1})$.

In D.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 residuals $Z_t(\boldsymbol{\theta})$ by $F^{-1}(R_t^{(n)}(\boldsymbol{\theta})/(n+1))$ yields $\underset{\sim}{\boldsymbol{\Delta}}^{(\boldsymbol{n})}(\boldsymbol{\theta}, f)$. We provide here an alternative heuristic argument similar to the one developed for the ARCH(q) case. For $t=1$, arbitrarily put $Z_1^{(n)}(\boldsymbol{\theta}) = Y_1$ and compute 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

$$\underset{\sim}{Y}_t := \vartheta \underset{\sim}{Y}_{t-1} + (1+\beta \underset{\sim}{Y}_{t-1}^2) F^{-1}\left(\frac{R_t^{(n)}(\boldsymbol{\theta})}{n+1}\right), \quad t \geq 2. \tag{5.9}$$

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 $m_{f,(2)}^{(n)}$ is as in (5.5) and

$$m_{f,(1)}^{(n)} = -\frac{1}{n} \sum_{i=1}^{n} \phi_f\left(F^{-1}\left(\frac{i}{n+1}\right)\right). \tag{5.10}$$

From the re-centering Lemma 1 in B.2, it follows that both $m_{f,(1)}^{(n)}$ and $m_{f,(2)}^{(n)}$ are $o\big(n^{-1/2}\big)$, hence can be omitted for n large. 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 block-diagonal form

$$\left(\begin{array}{c|c} I_1(f)\boldsymbol{I}_{p_1\times p_1} & \mathbf{0} \\ \hline \mathbf{0} & I_2(f)\boldsymbol{I}_{p_2\times p_2} \end{array}\right) \boldsymbol{\Upsilon}^{-1}(\boldsymbol{\theta}) \tag{5.11}$$

where $p_1 = p$, $p_2 = q$, and $\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\theta})$ is the Gaussian information matrix as in Chebana and Laïb (2010).

(c) Autoregressive conditional duration (ACD) models for irregularly sampled data. As in the seminal paper of Engle and Russell (1998), let Y_i denote the duration between some $(i-1)$-th and i-th events (e.g., the time elapsed between two successive transactions of some asset); let $\mathcal{F}_{i-1}$ denote the information up to and including event $(i-1)$, and denote by $\Psi_{i-1} := \mathrm{E}(Y_i|\mathcal{F}_{i-1})$ the expected conditional duration. Then, for $\boldsymbol{\theta} = (\beta, \gamma)$, define the *accelerated time process*

$$Y_i = \epsilon_i \Psi_{i-1}, \quad \text{with} \quad \Psi_{i-1} = \Psi(Z_i, \boldsymbol{\theta}) = 1 + \beta Y_{i-1} + \gamma \Psi_{i-2},$$

where the ϵ_i's are i.i.d., positive, and such that $\mathrm{E}(\epsilon_i|\mathcal{F}_{i-1}) = 1$. Engle and Russell (1998) propose a QL estimation procedure (which in this case is based on an exponential reference density), while Drost and Werker (2004) introduce the class of semiparametric ACD models (which does not specify any innovation density) and rely on the standard Bickel et al. semiparametric estimation method. We rather propose here a class of R-estimators for duration models. First note that the ULAN central sequence for ACD models actually is that of a dynamic scale model for Y_i: it follows from Proposition 3.1 that

$$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\psi_g(Z_i(\boldsymbol{\theta}))}{1 + \beta Y_{i-1} + \gamma \Psi_{i-2}} \begin{pmatrix} Y_{i-1} \\ \Psi_{i-2} \end{pmatrix}$$

where $Z_i(\boldsymbol{\theta}) = Y_i / \Psi_{i-1}(\boldsymbol{\theta})$. Similar to the recursion we gave for the ARCH(q) case, arbitrarily putting, for $t = 1$, $Z_1^{(n)}(\boldsymbol{\theta}) = Y_1$ yields $\Psi_0 = 1$ and an 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

$$\underset{\sim}{Y}_i = \underset{\sim}{\Psi}_{i-1} F^{-1}(R_{i-1}^{(n)}(\boldsymbol{\theta})/n + 1), \quad i \geq 2 \quad \text{where} \quad \underset{\sim}{\Psi}_{i-1} = 1 + \beta \underset{\sim}{Y}_{i-1} + \gamma \underset{\sim}{\Psi}_{i-2}. \tag{5.12}$$

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

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\psi_f\left(F^{-1}(R_i^{(n)}(\boldsymbol{\theta})/n + 1)\right)}{1 + \beta \underset{\sim}{Y}_{i-1} + \gamma \underset{\sim}{\Psi}_{i-2}} \begin{pmatrix} \underset{\sim}{Y}_{i-1} \\ \underset{\sim}{\Psi}_{i-2} \end{pmatrix}; \tag{5.13}$$

Lemma 1 in B.2 indeed implies that $m_{f,(2)}^{(n)} = o\big(n^{-1/2}\big)$, hence can be omitted. Since innovations are nonnegative, typical candidate reference densities here are the Gamma, Weibull, or Burr densities.

5.2. Discretely observed continuous-time models

Affine-jump diffusion processes are pivotal in the financial literature; see, e.g., Singleton (2009, Ch.3). Their main characteristic is that the conditional cumulant generating function is exponential-affine. As

a result, the conditional mean and the conditional variance of the discrete-time observed process are also affine and are known in closed-form. These features can be exploited to derive semiparametric dynamic location-scale models for the process observed at discrete-time points. The next two examples illustrate this approach.

(d) Discretely observed mean-reverting jump diffusion. Lévy processes are flexible and widely-applied in finance, as they are able to capture many features of 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 be explored; see, e.g., Bibby et al. (2010) and references therein. In the class of Lévy processes, let us consider the Poisson-Gaussian process $\mathcal{Y}$, which is solution to equation

$$dY_s = -\kappa Y_s ds + d\mathcal{W}_s + d\mathcal{Z}_s, \tag{5.14}$$

where $d\mathcal{W}_s$ is 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, for the discretely observed n-tuple $\{Y_0, Y_h, Y_{2h}, ..., Y_{nh}\}$,

$$\mathrm{E}(Y_{th}|Y_{(t-1)h}) = \frac{\alpha h}{\kappa}\left(1 - \exp\{-\kappa h\}\right) + Y_{(t-1)h}\exp\{-\kappa h\} \tag{5.15}$$

and

$$\mathrm{Var}(Y_{th}|Y_{(t-1)h}) = \frac{1+\eta^2}{2\kappa}\left(1 - \exp\{-2\kappa h\}\right). \tag{5.16}$$

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; Das points out that the bond market often overreacts, i.e., exhibits large moves in the interest rate followed by speedy reversals. The parameter κ measures the speed of mean reversion, and plays the main role: the half-life τ is a function of κ, being the solution to $\exp\{-\kappa\tau\} = 0.5$.

Assume the discrete-time process $\{Y_{th}; t \in \mathbb{Z}\}$ is observed over $n+1$ periods, yielding $(Y_0, Y_h, Y_{2h}, \ldots, Y_{nh})$. Das' estimation of κ is essentially based on an approximate version[13] of the dynamic location-scale model

$$Y_{th} = \frac{\alpha h}{\kappa}(1 - \exp\{-\kappa h\}) + Y_{(t-1)h}\exp\{-\kappa h\} + \left[\frac{1+\eta^2}{2\kappa}\left(1 - \exp\{-2\kappa h\}\right)\right]^{1/2}\epsilon_{th} \tag{5.17}$$

[13] Instead of the exact form (as given in (5.17)) of the conditional first two moments, Das considers a first-order approximation which follows from the Euler scheme applied to the stochastic differential equation (5.14). As a result, Das' approximation entails a bias for the estimated parameter. In contrast, we apply the exact conditional moments and our method does not entail any bias.

In Das' approach, the density g of ϵ_{th} is supposed to be standard normal. If that Gaussian assumption is to be abandoned, several semiparametric extensions of (5.17) are possible. The situation is actually pretty much the same as in the motivating example of Section 1.3 and, for the same reasons, turning to the residual ranks appears as the safest attitude.

Model (5.17) with innovation density g satisfying the usual regularity assumptions is ULAN with respect to $\boldsymbol{\theta} := (\kappa, \alpha, \eta)$, with central sequence (see (3.2))

$$\boldsymbol{\Delta}^{(\boldsymbol{n})}(\boldsymbol{\theta}, g) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{pmatrix} \psi_g(Z_t)\beta_1(\boldsymbol{\theta}) + \phi_g(Z_t)\kappa^2 Y_{(t-1)h}\beta_2(\boldsymbol{\theta}) - \phi_g(Z_t)\beta_2(\boldsymbol{\theta})\beta_3(\boldsymbol{\theta}) \\ \phi_g(Z_t)\beta_4(\boldsymbol{\theta}) \\ \psi_g(Z_t)\beta_5(\boldsymbol{\theta}) \end{pmatrix} \tag{5.18}$$

where

$$Z_t = Z_t(\boldsymbol{\theta}) = \frac{Y_{th} - \frac{\alpha h}{\kappa}(1 - \exp\{-\kappa h\}) - \exp\{-\kappa h\} Y_{(t-1)h}}{\left[\frac{1+\eta^2}{2\kappa}\left(1 - \exp\{-2\kappa h\}\right)\right]^{1/2}}, \tag{5.19}$$

with ($\mathrm{Coth}(x)$ and $\mathrm{Sinh}(x)$ as usual stand for the hyperbolic cotangent and sinus of x, respectively)

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

and $\beta_5(\boldsymbol{\theta}) = \eta/\left(2\kappa + \eta^2\right)$. Canceling $\boldsymbol{\Delta}^{(\boldsymbol{n})}(\boldsymbol{\theta}, g)$ yields M-estimators for $\boldsymbol{\theta}$ (which are not necessarily root-n consistent). Due to the highly nonlinear form of the estimating equations, moreover, numerical implementation is likely to be problematic, and even more so is the derivation of standard semiparametric estimators in the Bickel et al. (1993) style.

The R-estimation methods developed here thus naturally enter into the picture. Projecting $\boldsymbol{\Delta}^{(\boldsymbol{n})}(\boldsymbol{\theta}, f)$ (where f is some chosen reference density) onto the space of residual ranks cancels its second and third components; as for the first one, the terms involving hyperbolic functions disappear, and only $\phi_g(Z_t)\kappa^2 Y_{(t-1)h}\beta_2(\boldsymbol{\theta})$ yields a nondegenerate projection. This means that neither α nor η can be estimated at root-n rate when the density g of ϵ_{th} in (5.17) remains completely unspecified. Going back to equation (5.17), 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 reference density f, the projection onto the σ-algebra of residual ranks of the component of the central sequence associated with κ thus coincides (up to an irrelevant multiplicative constant) 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 (5.19). That rank is the same as the rank $R_t^{(n)}(\kappa)$ of $Z_t^{\dagger}(\kappa) := Y_{th} - \exp\{-\kappa h\} Y_{(t-1)h}$. Then, a rank-based central sequence emerges, of the form (up to a

multiplicative constant and $o_{\mathrm{P}}(1)$ terms)

$$\underset{\sim}{\Delta}^{(n)}(\kappa, f) := n^{1/2} \sum_{i=0}^{s_n} \exp\{-i\kappa h\}(n-i)^{-1} \sum_{t=i+1}^{n} \Big(\phi_f\Big(F^{-1}\Big(\frac{R_t^{(n)}(\kappa)}{n+1}\Big)\Big) F^{-1}\Big(\frac{R_{t-i}^{(n)}}{n+1}\Big) - m_f^{(n)}\Big) \tag{5.20}$$

where $m_f^{(n)} := [n(n-1)]^{-1} \sum_{1 \le i_1 \ne i_2 \le n} \phi_f(F^{-1}(i_1/n+1)) F^{-1}(i_2/n+1)$, with, under $\mathcal{H}_{\boldsymbol{\theta}}^{(n)}$, asymptotic variance $\Gamma^*(\boldsymbol{\theta}, f) = I_1(f)/\big(1 - \exp\{-2\kappa h\}\big)$; semiparametric efficiency here refers to the discrete-time model (5.17) with completely unspecified innovation density g.

Based on (5.20), our method (as described in Section 4) then leads to root-n consistent R-estimators for κ in the rather sophisticated context of a discretized jump diffusion process where the jump parameters are treated as nuisance; in that sense, our R-estimators are robust to a misspecification of the jump process.

(e) Discretely observed Cox-Ingersoll-Ross (CIR) process. The CIR process $\mathcal{Y}$ is often considered for short-term interest rates (e.g., the Fed funds rate as in Das (2002)), stochastic volatility (Heston (1993)), or asset pricing models (Singleton (2009)). It is the solution to the stochastic differential equation

$$dY_s = k(1 - Y_s)dt + \sigma\sqrt{Y_s} d\mathcal{W}_s. \tag{5.21}$$

In Chapter 2 of Singleton (2009), a QL estimator for the model parameters in (5.21) is defined setting a Gaussian dynamic location-scale model for the discrete-time version of $\mathcal{Y}$. In D.4 and D.3, we show how a semiparametric dynamic location-scale extension leads to a semiparametric AR(1)-ARCH(1) model, whose rank-based central sequence can be used in the construction of R-estimators.

6. Numerical examples

6.1. Returns and realized stochastic volatilities

In this section, we study the 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 which are used in modeling and forecasting the realized volatility of assets; see, e.g., Bingham and Kiesel (2002), Corsi (2009), Corsi et al. (2013), and references therein. Before turning to finite-sample performances (see Section 7), let us first analyze the asymptotic ones, as evaluated by asymptotic relative efficiencies.

Asymptotic performance (AREs). We study here the asymptotic relative efficiencies, in model (6.1), with respect to the Gaussian QL, of the R-estimators based on van der Waerden ($\underset{\sim}{\boldsymbol{\Delta}}_{\mathrm{vdW}}^{(n)}$), Wilcoxon ($\underset{\sim}{\boldsymbol{\Delta}}_{W}^{(n)}$),

and Laplace ($\underset{\sim}{\Delta}_L^{(n)}$) rank-based central sequences (associated with normal, logistic and double-exponential reference densities, respectively); their explicit expressions are given in D.5; the centering and scaling constants $m_f^{(n)}$ and $s_f^{(n)}$ are provided explicitly in Hallin and Mélard (1988)—see also Lemma 1 in B.2, which implies $m_f^{(n)} = o\big(n^{-1/2}\big)$.

The performance of all those estimators 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 $\mathrm{J_{SU}}(\gamma, \delta, \mu, \sigma)$, where γ and δ are skewness and kurtosis parameters, respectively, while μ and σ, as usual, stand for location and scale. For $\mu = 0$ and $\sigma = 1$, suitable values of δ and γ may lead to positive skewness values as large as ten, and excess of kurtosis larger than eight.

	$\mathrm{J_{SU}}(\gamma, \delta, \mu, \sigma)$					
	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 Gaussian QL estimator. The underlying process is the variance mixture model (6.1), with Johnson mean zero and variance one innovation density $\mathrm{J_{SU}}(\gamma, \delta, \mu, \sigma)$.

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 Section C.2, and $I_1(g) := I_1(g, g)$. Results are provided in Table 1.

We emphasize that the AREs displayed in this table are uniform in $\boldsymbol{\theta}$; note that, under leptokurtic innovation density g (e.g., $\delta = 0.85$ or 1), all R-estimators considered here quite significantly 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. Time-varying volatility in the presence of skewness and leptokurtosis

Let us consider the semiparametric ARCH(1) model $Y_t = (1 + \theta Y_{t-1}^2)^{1/2}\epsilon_t$, where the i.i.d. ϵ_t's have unspecified density g. For each combination of the two parameter values $\theta = 0.1$ and $\theta = 0.5$, and three

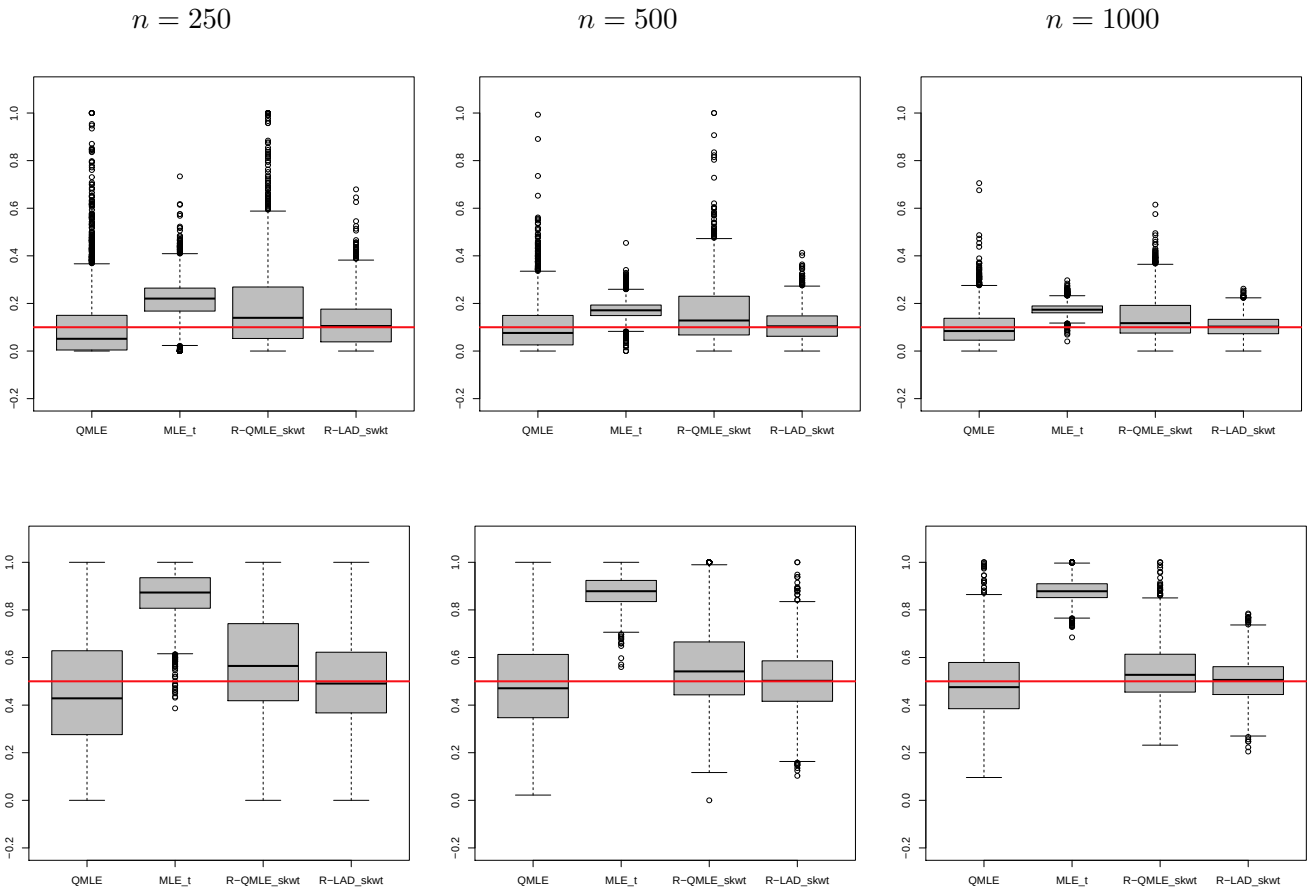

Figure 1: Boxplots for various estimators of θ: (a) Gaussian Quasi-Likelihood estimator (QMLE); (b) skew-t-Maximum Likelihood Estimator (MLE_t); (c) R-estimator with QMLE preliminary and data-driven skew-t reference density (R-QMLE_skwt); (d) R-estimator with LAD preliminary and skew-t data-driven reference density (R-LAD_skwt)). In each plot the horizontal red line indicates the true parameter value: the top plots are for $\theta = 0.1$, while the bottom plots are for $\theta = 0.5$. Monte Carlo size throughout is 2500.

different series lengths $n = 250, 500, 1000$, we simulated 2500 realizations based on a skew generalized error density g with mean zero, standard deviation one, shape parameter one, and skewness parameter five; such densities are increasingly popular in finance (see, e.g., Hansen et al. (2007)). From each realization, we computed the following estimators of θ: (a) the Gaussian QL estimator (QLE), (b) the skew-t-maximum likelihood (MLE_t), (c) an R-estimator (R-QLE_skwt) based on a QLE preliminary, and (d) an R-estimator (R-LAD_skwt) based on a LAD preliminary. Both R-estimators were obtained via data-driven skew-t reference densities. More precisely, following the method described in Section 4.3.2 skew-t distributions with location zero and unit scale were fitted to the (order statistics of) residuals via a maximum likelihood estimation of skewness and degrees of freedom, which accounts for skewness and kurtosis.

In Figure 1 we display the boxplots associated with the four estimators[14] just described. We notice

[14]The computation of rank-based central sequences was based on recursion (5.3), and one-step iterations were performed until

that, for $n = 250$ and $n = 500$, the QLE displays a bias, whereas the R-LAD_skwt is essentially centered about the actual parameter value (both for $\theta = 0.1$ and for $\theta = 0.5$). Since innovations, under our data-generating scheme, have finite fourth-order moment, the QLE is root-n consistent. This can be observed for $n = 1000$, where both the QLE and R-LAD_skwt are centered at the actual parameter value, with R-LAD_skwt displaying higher precision (the R-LAD_skwt yields a sizeable asymptotic efficiency gain over the QLE). Moreover, a comparison of the boxplots of R-QMLE_skwt with those of R-LAD_skwt indicates that the use of LAD as a preliminary estimator tends to improve on the R-estimator performance when the sample size is $n = 250$. However, for the larger sample sizes ($n = 500, 1000$) the difference is less apparent (see e.g. the bottom-right panel for $n = 1000$) and, in line with the asymptotic theory, eventually disappears as $n \to \infty$. Finally, we notice that the MLE_t is severely biased in all settings. Indeed, MLE_t is based on a misspecified reference density which does not meet the Fisher consistency conditions.

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 so-called 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 description

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)) TSRV 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 of September) as training data for model estimation and diagnosis, and the last 50 ones (from October to December) to evaluate forecasting performances.

the desired numerical stabilization was achieved (9 iterations on average).

7.2. Modeling and forecasting

7.2.1. Exploratory analysis and modeling

Log-returns. 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 p-values for $r_t/TSRV_t$ are 0.896 and 0.208 for the 1997 and 1993 data, respectively. Thus, we conclude that a standard normal approximation for the ratios $r_t/TSRV_t$ 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 2 displays some summary statistics for their unconditional distributions. 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 $\sum_{j=1}^{3} \theta_j \log(\mathrm{TSRV}_{t-j})$.

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—the Gaussian 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 C. Results are displayed in Table 3.

In the 1993 training data, all estimation methods considered suggest an AR(1) model for the $\log(\mathrm{TSRV}_t)$ series, while in 1997 the same estimation methods agree on an AR(2) model. Table 3 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 2, we plot the sample partial autocorrelation of the standardized and squared-standardized residuals, as implied

	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 2: 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).

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.

7.2.3. Forecasting

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

Table 4 provides some classical (mean and standard deviation) and robust (median and mean absolute deviation) evaluations of the squared 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, with rank-based methods than with the traditional QL ones.

8. Conclusions

The new R-estimation methodology developed in this paper is particularly well adapted to the semi-parametric analysis of the type of complex and nonlinear time series models considered in econometrics and

	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 3: USD/CHF FX rates: inference on log(TSRV). Gaussian QL and R-estimates of θ_1, θ_2, and θ_3 (along with their estimated standard errors).

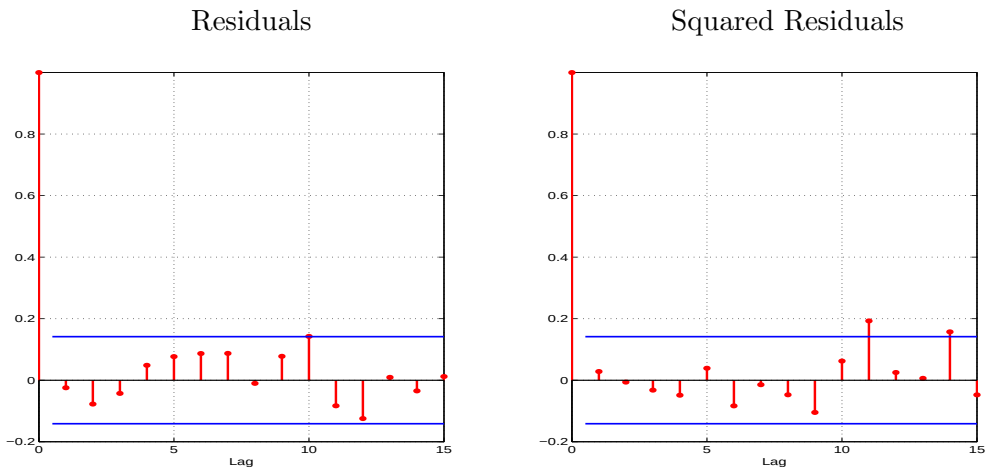

Figure 2: USD/CHF FX rates: diagnostics for the Laplace R-estimator, in the 1993 data. Sample partial autocorrelation of residuals (left) and squared-residuals (right).

finance, for which a more standard semiparametric approach would be hardly feasible, due to the difficulty of obtaining the tangent spaces and implementing the corresponding projections, and the required estimation of innovation densities. Our methodology is flexible (allowing for non-Gaussian and data-driven reference densities), powerful (quite often, outperforming Gaussian QL estimation), robust (preserving root-n consistency under misspecified densities), and computationally less demanding than the standard semiparametric method. Our method also represents a substantial contribution to the literature on rank-based inference for time series, which so far has been focused, mostly, on the simpler problem of rank-based hypothesis testing for conditionally homoskedastic (and generally linear) models. Indeed, our method is the first one extending rank-based inference to autoregressive conditional duration models, AR-ARCH models, and discretely observed affine diffusion processes.

	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 4: USD/CHF FX rates: mean, median, standard deviation, and mean absolute deviation for the squared one-day-ahead prediction errors for both QL and R-estimators. The values for the R-estimators are expressed as proportions of the corresponding QL values.

Acknowledgment

We thank Fulvio Corsi for providing us with the TSRV data. We also thank Adelchi Azzalini, Matteo Barigozzi, Debbie Dupuis, Patrick Gagliardini, Loriano Mancini, Gael Martin, Davy Paindaveine, Elvezio Ronchetti, Olivier Scaillet, Mervyn Silvapulle, Fabio Trojani, Fashid Vahid-Araghi, Maria-Pia Victoria-Feser, Roy Welsch, and Bas Werker, for helpful discussions about an earlier version of this manuscript. We acknowledge the Australian Research Council grant DP150100210 for financial support. Marc Hallin is also member of the Econometrics Group, Tilburg University (Tilburg, The Netherlands), and the Royal Academy of Belgium. His research is supported by an Interuniversity Attraction Pole (2012-2017) of the Belgian Science Policy Office, a Crédit aux Chercheurs of the Fonds de la Recherche Scientifique-FNRS of the Communauté Française de Belgique.

References

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.

Allal, J., Kaaouachi, A., Paindaveine, D., 2001. R-estimation for ARMA models. Journal of Nonparametric Statistics 13, 815–831.

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, 51–73.

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

Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. Journal of the Royal Statistical Society Series B 65, 367–389.

Barndorff-Nielsen, O., Shephard, N., 2001. Non-Gaussian Ornstein-Uhlenbeck-based models and of their uses in financial economics. Journal of the Royal Statistical Society Series B 63, 167–241.

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 Semiparametric Models. Johns Hopkins Univ. Press, Baltimore.

Bingham, N., Kiesel, R., 2002. Semi-parametric modelling in finance: theoretical foundations. Quantitative Finance 2 (4), 241–250.

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

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.

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, 284–304.

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

Dodge, Y., Jurečková, J., 2000. Adaptive Regression. New York: Springer.

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

Drost, F. C., Klaassen, C. A., 1997. Efficient estimation in semiparametric GARCH models. Journal of Econometrics 81, 193–221.

Drost, F. C., Werker, B. J. M., 2004. Semiparametric duration models. Journal of Business & Economic Statistics 22, 40–50.

Engle, R., Gonzalez-Rivera, G., 1991. Semiparametric ARCH. Journal of Business and Economics Statistics 9, 345–359.

Engle, R. F., Bollerslev, T., 1986. Modelling the persistence of conditional variances. Econometric Reviews 5, 1–50.

Engle, R. F., Russell, J. R., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 66, 1127–1162.

Fan, J., Qi, L., Xiu, D., 2014. Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. Journal of Business & Economic Statistics 32, 178–191.

Fernandes, M., Linton, O., Scaillet, O., 2007. Semiparametric methods in econometrics. Journal of Econometrics 141 (1), 1–4.

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

Ghysels, E., Wang, F., 2011. Some useful densities for risk management and their properties. Working paper available at URL: http://www. unc. edu/˜ eghysels/working_papers. html.

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

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

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

Hall, P., Yao, Q., 2003. Inference in ARCH and GARCH models with heavy–tailed errors. Econometrica 71, 285–317.

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., Mehta, C., 2015. R-estimation for asymmetric independent component analysis. Journal of the American Statistical Association 110, 218232.

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

Hallin, M., Paindaveine, D., 2013. Semiparametric efficient one-step R-estimators, unpublished manuscript.

Hallin, M., Puri, M., 1991. Time-series analysis via rank-order theory: signed-rank tests for ARMA models. Journal of Multivariate Analysis 39, 175–237.

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, 181–186.

Hansen, C., McDonald, J., Theodossiou, P., 2007. Some flexible parametric models for partially adaptive estimators of econometric models. Economics (OpenAccess).

Heston, S. L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. Review of Financial Studies 6, 327–343.

Hodges, J., Lehmann, E. L., 1956. The efficiency of some nonparametric competitors of the t-test. The Annals of Mathematical Statistics 34, 324–335.

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, 761–780.

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, 946–961.

LeCam, L., Yang, G., 1990. Asymptotic in Statistics. Springer-Verlag, New York.

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

Lucas, A., 1997. Robustness of the Student t based M-estimator. Communications in Statistics-Theory and Methods 26, 1165–1182.

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, 628–641.

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

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

Newey, W. K., 1990. Semiparametric efficiency bounds. Journal of applied econometrics 5 (2), 99–135.

Pantula, S. G., 1988. Estimation of autoregressive models with ARCH errors. Sankhyā Series B, 119–138.

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

Singleton, K. J., 2009. Empirical Dynamic Asset Pricing: model specification and econometric assessment. Princeton University Press.

Terpstra, J. T., McKean, J. W., Naranjo, J. D., 2001. GR-estimates for an autoregressive time series. Statistics & Probability Letters 51, 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.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica, 1–25.

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

Appendix

A. Proofs

A.1. Proof of Proposition 3.1

Proposition 3.1 follows from Theorem 2.1 in Drost et al. (1997): we just need to check that their Assumptions A-E[15] 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 4.2

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 $\mathrm{P}_{\boldsymbol{\theta},g}^{(n)}$,

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

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

B. Technicalities

B.1. Computation of the rank-based central sequence

The rank-based central sequences $\underset{\sim}{\boldsymbol{\Delta}}{}^{(n)}(\boldsymbol{\theta}, f)$ we eventually are working with are very simply obtained by substituting $F^{-1}(R_t/(n+1))$ for Z_t in $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)$. This, however, requires some justification. Before projecting $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)$, we first need rewriting 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-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: with a slight abuse of notation,

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

[15] In this proof, labels A, B, ... refer to the assumptions in Drost et al. (1997), labels (A), (B), ... to ours.

It may happen that (B.1) only requires a finite number 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 (B.1), $\mathbf{Z}_{t-1}$ 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.

Assumption (D). There exists a sequence of integers $s_n < n$ such that $s_n \uparrow \infty$ as $n \to \infty$ and, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $g \in \mathcal{G}$,

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

where $o_{\mathrm{qm}}(1)$ stands for a sequence that tends to zero in quadratic mean.

Assumption (D) is satisfied by most stationary Markov processes of order p and all q-dependent processes; examples are provided in Section 5. It 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 (B.2) 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 combination of such componentwise monotone functions.

This technical assumption is required in the proof of Proposition 3.2, for the usual asymptotic representation results for rank-based statistics to hold; see Lemma 3.1 in Hallin and Puri (1991) and reference therein. Recall, however, that the class of differences of monotone functions coincides with the class of functions with bounded variation. Restricting to reference densities f for which such functions as ϕ_f or ψ_f have bounded variation has little practical consequences and only discards pathological cases. As for the (marginal) dependence on lagged values of the Z_t's, it is linear in AR-type models, or polynomial when based on a Volterra expansion (as in D.1 and D.2), hence also has bounded variation.

The rank-based score in (3.5) is defined as the projection of $\dot{\boldsymbol{l}}^{s_n}(Z_t, \mathbf{Z}_{t-1}^{s_n}, \boldsymbol{\theta}, f)$ onto $\mathcal{B}^{(n)}(\boldsymbol{\theta})$. Letting $s_n = s$ and $\boldsymbol{R}_t^{s_n} := (R_t^{(n)}, \ldots, R_{t-s}^{(n)})$, we set (recall that the score function is square-integrable), under Assumption (E),

$$\boldsymbol{a}_f(\boldsymbol{R}_t^{s_n}; \boldsymbol{\theta}) := \mathrm{E}_{\boldsymbol{\theta},f} \left[\dot{\boldsymbol{l}}^{s_n}(Z_t, \mathbf{Z}_{t-1}^{s_n}, \boldsymbol{\theta}, f) | R_t^{(n)}, \ldots, R_{t-s}^{(n)} \right] \tag{B.3}$$

(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*

$$\boldsymbol{a}_f^{(n)}(\boldsymbol{R}_t^{s_n}; \boldsymbol{\theta}) := \dot{\boldsymbol{l}}^{s_n} \left(F^{-1}\left(\frac{R_t^{(n)}}{n+1} \right), \ldots, F^{-1}\left(\frac{R_{t-s}^{(n)}}{n+1} \right), \boldsymbol{\theta}, f \right); \tag{B.4}$$

contrary to the exact ones, those approximate scores are straightforwardly evaluated.

With approximate scores, the rank-based central sequence takes the form

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f) = \frac{1}{\sqrt{n-s}} \sum_{t=s+1}^{n} \boldsymbol{a}_f^{(n)}(\boldsymbol{R}_t^s; \boldsymbol{\theta}) - \boldsymbol{m}_f^{(n)} \tag{B.5}$$

with the re-centering $\boldsymbol{m}_f^{(n)} := \sqrt{n-s}\mathrm{E}\left[\sum_{t=s+1}^{n} \boldsymbol{a}_f^{(n)}(\boldsymbol{R}_t^s; \boldsymbol{\theta}) \right]$. By standard U-statistics results:

$$\boldsymbol{m}_f^{(n)} = \frac{\sqrt{n-s}}{n(n-1)\cdots(n-s)} \sum_{1 \le i_1 \ne \cdots \ne i_s \le n} \cdots \sum \boldsymbol{a}_f^{(n)}\left(\frac{i_1}{n+1}, \ldots, \frac{i_s}{n+1}; \boldsymbol{\theta} \right) + o_P(1), \tag{B.6}$$

so that $\boldsymbol{m}_f^{(n)}$ indeed qualifies as a centering; see Lemma 1 in B.2.

B.2. A re-centering Lemma

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

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

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

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

$$\max_{1\le i\le n} \frac{1}{n}\Big|h\Big(\frac{i}{n+1}\Big)\Big| = o\big(n^{-1/2}\big). \tag{B.7}$$

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 $\int_{u^-}^{u^+} h(u)\ du = 0$. Defininning $i^- := \lceil (n+1)u^- \rceil$, $i_0 = \lfloor (n+1)u_0 \rfloor$, 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,

$$0 \le -\frac{n}{n+1} S_n^{--} \le -\int_0^{u^-} h(u)\ du \quad \text{and} \quad 0 \le \frac{n}{n+1} S_n^{++} \le \int_{u^+}^1 h(u)\ du, \tag{B.8}$$

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,

$$-\frac{n}{n+1} S_n^{-} = \begin{cases} \underline{D}_n^- & + & \Big(u^- - \dfrac{i^- - 1}{n+1}\Big)\Big|h\Big(\dfrac{i^-}{n+1}\Big)\Big| \\ \bar{D}_n^- & + & \Big(\dfrac{i_0+1}{n+1} - u_0\Big)\Big|h\Big(\dfrac{i_0}{n+1}\Big)\Big| - \Big(\dfrac{i^-}{n+1} - u^-\Big)\Big|h(u^-)\Big| \end{cases} \tag{B.9}$$

and

$$\frac{n}{n+1} S_n^{+} = \begin{cases} \underline{D}_n^+ & + & \Big(\dfrac{i^+ + 1}{n+1} - u^+\Big) h\Big(\dfrac{i^+}{n+1}\Big) \\ \bar{D}_n^+ & + & \Big(u_0 - \dfrac{i_0}{n+1}\Big) h\Big(\dfrac{i_0+1}{n+1}\Big) - \Big(u^+ - \dfrac{i^+}{n+1}\Big) h(u^+) \end{cases} \tag{B.10}$$

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 (B.9), (B.10) and (B.8) that

$$\begin{aligned} &\underline{D}_n^+ - \bar{D}_n^- + \int_0^{u^-} h(u)\ du + o\big(n^{-1/2}\big) \qquad\qquad (\text{B.11}) \\ &= \underline{D}_n^+ - \bar{D}_n^- + \Big(\frac{i^+ + 1}{n+1} - u^+\Big) h\Big(\frac{i^+}{n+1}\Big) + \Big(\frac{i_0+1}{n+1} - u_0\Big) h\Big(\frac{i_0}{n+1}\Big) - \Big(\frac{i^-}{n+1} - u^-\Big) h\big(u^-\big) + \int_0^{u^-} h(u)\ du \\ &\le \frac{n}{n+1} S_n \\ &\le \bar{D}_n^+ - \underline{D}_n^- + \Big(u_0 - \frac{i_0}{n+1}\Big) h\Big(\frac{i_0+1}{n+1}\Big) - \Big(u^+ - \frac{i^+}{n+1}\Big) h\big(u^+\big) + \Big(u^- - \frac{i^- - 1}{n+1}\Big) h\Big(\frac{i^-}{n+1}\Big) + \int_{u^+}^1 h(u)\ du \\ &= \bar{D}_n^+ - \underline{D}_n^- + \int_{u^+}^1 h(u)\ du + o\big(n^{-1/2}\big). \qquad\qquad (\text{B.12}) \end{aligned}$$

Now,

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

and

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

It follows that the lower and upper bounds in (B.11) and (B.12) reduce to

$$\int_0^{u^-} h(u)\ du + o\big(n^{-1/2}\big) \quad \text{and} \quad \int_{u^+}^1 h(u)\ du + o\big(n^{-1/2}\big),$$

respectively, and their difference 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 non decreasing square-integrable functions to which Lemma 1 applies.

C. Cross-information quantities

C.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 $\mathbf{\Gamma}(\boldsymbol{\theta}, f, g)$ in (4.2). Constructing such an estimator is a delicate task, since $\mathbf{\Gamma}(\boldsymbol{\theta}, f, g)$ 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 (2013).

Very often, though, the matrix $\mathbf{\Gamma}(\boldsymbol{\theta}, f, g)$ 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 $\mathbf{\Gamma}(\boldsymbol{\theta}, f, g) = \mathcal{J}(f, g)\mathbf{\Upsilon}^{-1}(\boldsymbol{\theta})$, where $\mathcal{J}(f, g)$ is a scalar quantity depending on f and g only, while $\mathbf{\Upsilon}^{-1}(\boldsymbol{\theta})$ only depends on $\boldsymbol{\theta}$. In some others, $\mathbf{\Gamma}(\boldsymbol{\theta}, f, g)$ 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 $\boldsymbol{\theta} \in \mathbf{\Theta}$ and $f, g \in \mathcal{G}$, the cross-information matrix $\mathbf{\Gamma}(\boldsymbol{\theta}, f, g)$

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

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

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

$$\mathbf{\Gamma}(\boldsymbol{\theta}, f, g) = \left(\begin{array}{c|c} I_1(f, g)\boldsymbol{I}_{p_1 \times p_1} & \mathbf{0} \\ \hline \mathbf{0} & I_2(f, g)\boldsymbol{I}_{p_2 \times p_2} \end{array}\right) \mathbf{\Upsilon}^{-1}(\boldsymbol{\theta}) \tag{C.13}$$

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 $\mathbf{\Upsilon}^{-1}(\boldsymbol{\theta})$ 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 (C.13) as soon as $\mathcal{G}$ is restricted to symmetric (with respect to 0) densities—see Section C.2 for details – an assumption which is quite

common in the literature, see, e.g., Gouriéroux 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

$$\underset{\sim}{\boldsymbol{\theta}}_f^{(n)} := \hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\boldsymbol{\Upsilon}(\hat{\boldsymbol{\theta}}^{(n)}) \left(\begin{array}{c|c} \hat{I}_1^{-1}(f,g)\boldsymbol{I}_{p_1\times p_1} & \mathbf{0} \\ \hline \mathbf{0} & \hat{I}_2^{-1}(f,g)\boldsymbol{I}_{p_2\times p_2} \end{array}\right) \underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\hat{\boldsymbol{\theta}}^{(n)}, f). \tag{C.14}$$

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

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

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

$$(\lambda_{*1}^{(n)},\lambda_{*2}^{(n)}) := \inf_{(\lambda_1,\lambda_2)\in\mathbb{R}^+\times\mathbb{R}^+} \left\{\lambda_1,\lambda_2 | \underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\hat{\boldsymbol{\theta}}^{(n)})'\boldsymbol{\Upsilon}(\hat{\boldsymbol{\theta}}^{(n)})\boldsymbol{\Upsilon}(\tilde{\boldsymbol{\theta}}^{(n)}(\lambda_1,\lambda_2))\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\tilde{\boldsymbol{\theta}}^{(n)}(\lambda_1,\lambda_2)) < 0\right\}. \tag{C.15}$$

C.2. Examples

Let us show that, under suitable assumptions on innovation densities, Assumption (G) holds for all models considered in Section 5. To illustrate this point, let us define

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

Those four (cross-)information quantities enter the definition of $\boldsymbol{\Gamma}(\boldsymbol{\theta},f,g)$. Assumption (G1) clearly holds when $I_{12}(f,g)$ and $I_{21}(f,g)$ (which appear in the off-diagonal blocks, if any, of $\boldsymbol{\Gamma}(\boldsymbol{\theta},f,g)$) both vanish --a condition which is clearly satisfied when f and g both are symmetric. Here are a few examples from Section 5.1.

- *ARCH(q).* The central sequence in Eq. (5.2) 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 (2013) 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 (2013) applies.

D. Analytical derivation of the rank-based central sequences in Section 5

D.1. ARCH(q)

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

$$Y_t = \left(1+\sum_{j=1}^q \theta_j Y_{t-j}^2\right)^{1/2}\epsilon_t, \tag{D.16}$$

where the ϵ_t's are i.i.d. with density g, $\boldsymbol{\theta} = (\theta_1,...,\theta_q)$ where $\theta_j > 0$ for $j = 1,..,q$ and such that $\sum_{j=1}^q \theta_j \leq \rho < 1$.

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

$$\boldsymbol{\Delta}^{(\boldsymbol{n})}(\boldsymbol{\theta}, g) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\psi_g(Z_t(\theta))}{1+\sum_{j=1}^q \theta_j Y_{t-j}^2} \begin{pmatrix} Y_{t-1}^2 \\ \vdots \\ Y_{t-q}^2 \end{pmatrix}. \tag{D.17}$$

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

$$w_t(k) := \sum_{j_1,\ldots,j_k \geq 1} \prod_{r=1}^{k} \theta_{j_r} \prod_{r'=0}^{k} \epsilon^2_{t-\sum_{h=0}^{r'} j_h} \tag{D.18}$$

(with $j_0 = 0$) converges in probability (this follows from stationarity of Y_t, see, e.g., Giraitis et al. (2000) 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,

$$Y_t^2 = \tilde{Y}_t^2 + \sum_{k\geq s_n} w_t(k), \quad \text{with} \quad \tilde{Y}_t^2 := \epsilon_t^2 + \sum_{k=1}^{s_n} w_t(k). \tag{D.19}$$

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

Combining (D.19) and (D.17), we get

$$\boldsymbol{\Delta}^{(\boldsymbol{n})}(\boldsymbol{\theta}, g) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\psi_g\left(Z_t\right)}{1+\sum_{j=1}^q \theta_j \left(Z_{t-j}^2 + \sum_{k\geq 1} w_{t-j}(k)\right)} \begin{pmatrix} Z_{t-1}^2 + \sum_{k\geq 1} w_{t-1}(k) \\ \vdots \\ Z_{t-q}^2 + \sum_{k\geq 1} w_{t-q}(k) \end{pmatrix} \tag{D.20}$$

where, with a slight abuse of notation, $w_{t-j}(k)$ is as in (D.18), 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 (D.17) by another central sequence depending on a finite number s of lags only, which we also denote as $\boldsymbol{\Delta}^{(\boldsymbol{n})}(\boldsymbol{\theta}, g)$. In view of Section B.1 and (D.20), 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 (5.5), is such that the expected value of $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, g)$ is exactly zero for every n.

Lemma 1 in B.2 implies that $m^{(n)}_{g,(2)} = o\big(n^{-1/2}\big)$, hence can be omitted, yielding

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

Of course, rather than $\underset{\sim}{\Delta}^{(n)}(\boldsymbol{\theta}, g)$, which is associated with the unknown actual density g, the rank-based central sequence $\underset{\sim}{\Delta}^{(n)}(\boldsymbol{\theta}, f)$ associated with the reference density f is to computed for R-estimation purposes.

D.2. AR(p)-LARCH(q)

Consider the discrete-time bilinear process with dynamics

$$Y_t = \sum_{j=1}^{p} \vartheta_j Y_{t-j} + \left(1 + \sum_{l=1}^{q} \beta_l Y_{t-l} \right) \epsilon_t, \tag{D.21}$$

where $\epsilon_t \sim g$, with mean zero and unit variance, and $\boldsymbol{\theta} = (\vartheta_1, ..., \vartheta_p, \beta_1, ..., \beta_q)$ satisfies Assumptions A_1-A_3 in Giraitis and Surgailis (2002). Model (D.21) 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 $\inf_{\theta \in \Theta} \left(1 + \sum_{l=1}^{q} \beta_l Y_{t-l} \right)$ 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:

$$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(1 + \sum_{l=1}^{q} \beta_l Y_{t-l} \right)^{-1} \begin{pmatrix} -\phi_g\left(Z_t(\boldsymbol{\theta}) \right) \begin{pmatrix} Y_{t-1} \\ \vdots \\ Y_{t-p} \end{pmatrix} \\ \psi_g\left(Z_t(\boldsymbol{\theta}) \right) \begin{pmatrix} Y_{t-1} \\ \vdots \\ Y_{t-q} \end{pmatrix} \end{pmatrix}. \tag{D.22}$$

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

$$Y_t = \sum_{k=1}^{\infty} \sum_{j_k < .. < j_1 < t} u_{t-j_1} w_{j_1 - j_2} ... w_{j_{k-1} - j_k} \epsilon_{j_1} ... \epsilon_{j_k}, \tag{D.23}$$

which depends on a infinite number of lagged shocks.

From Theorem 2.2 in Giraitis and Surgailis (2002) it follows that $Y_t = Y^s_t + o_{\mathrm{P}}(1)$, where Y^s_t is obtained by truncating Eq. (D.23) 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,

$$e_{t-j} = \frac{Y_{t-j}}{(1+\sum_{l=1}^{q}\beta_l Y_{t-l})} - \frac{Y_{t-j}^s}{(1+\sum_{l=1}^{q}\beta_l Y_{t-l}^s)} = o_{\mathrm{P}}(1), \tag{D.24}$$

so that, letting

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

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

and

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

we have that

$$\boldsymbol{\Delta}^{(\boldsymbol{n})}(\boldsymbol{\theta},g) \quad = \quad \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}\boldsymbol{\Delta}_{(\mathbf{1})}^{(\boldsymbol{t},\boldsymbol{s})}(\boldsymbol{\theta},g)\\ \boldsymbol{\Delta}_{(\mathbf{2})}^{(\boldsymbol{t},\boldsymbol{s})}(\boldsymbol{\theta},g)\end{pmatrix} \tag{D.25}$$

is another version of the central sequence in (D.22), since it approximates $\boldsymbol{\Delta}^{(\boldsymbol{n})}(\boldsymbol{\theta},g)$ up to $o_{\mathrm{P}}(1)$. The rank-based central sequence $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta},g)$ again is obtained as the approximate-score projection of this central sequence $\boldsymbol{\Delta}^{(\boldsymbol{n})}(\boldsymbol{\theta},g)$ onto the invariant σ-field generated by the ranks, namely replacing, in (D.25), 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 (5.8). The truncated Volterra series here takes the form

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

which is such that $Y_t = Y_t^s + o_{\mathrm{P}}(1)$; an alternative representation for Y_t^s follows from iterating Eq. (5.8):

$$Y_t^s = \epsilon_t + \sum_{k=1}^{s}\epsilon_{t-k}\prod_{m=t-k+1}^{t}(\beta_1\epsilon_m+\vartheta_1), \tag{D.26}$$

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

$$\begin{aligned}\boldsymbol{\Delta}^{(\boldsymbol{n})}(\boldsymbol{\theta},g) \quad &= \quad \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}-\phi_g\left(Z_t\right)\\ \psi_g\left(Z_t\right)\end{pmatrix}\frac{\sum_{k=1}^{s}(\beta_1/\vartheta_1)^k\sum_{j_k,...,j_1<t-1}\vartheta_1^{t-j_k}Z_{j_1},...,Z_{j_k}}{1+\beta_1\sum_{k=1}^{s}(\beta_1/\vartheta_1)^k\sum_{j_k,...,j_1<t-1}\vartheta_1^{t-j_k}Z_{j_1},...,Z_{j_k}}\\ &= \quad \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}-\phi_g\left(Z_t\right)\\ \psi_g\left(Z_t\right)\end{pmatrix}\frac{Z_t+\sum_{k=1}^{s}Z_{t-k}\prod_{m=t-k+1}^{t}(\beta_1 Z_m+\vartheta_1)}{1+\beta_1(Z_t+\sum_{k=1}^{s}Z_{t-k}\prod_{m=t-k+1}^{t}(\beta_1 Z_m+\vartheta_1))},\end{aligned} \tag{D.27}$$

where the last expression, which is easier to work with, follows from Eq. (D.26). 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

$$\underset{\sim}{\Delta}^{(t,s)}_{(1)}(\boldsymbol{\theta},g) := \left(-\underset{\sim}{\zeta}_t\phi_g\left(G^{-1}\left(\frac{R^{(n)}_t}{n+1}\right)\right)-m^{(n)}_{g,(1)}\right) \times\left(G^{-1}\left(\frac{R^{(n)}_t}{(n+1)}\right)+\sum_{k=1}^{s}G^{-1}\left(\frac{R^{(n)}_{t-k}}{n+1}\right)\prod_{m=t-k+1}^{t}\left(\beta_1 G^{-1}\left(\frac{R^{(n)}_m}{n+1}\right)+\vartheta_1\right)\right),$$

and

$$\underset{\sim}{\Delta}^{(t,s)}_{(2)}(\boldsymbol{\theta},g) := \underset{\sim}{\zeta}_t\left(\psi_g\left(G^{-1}\left(\frac{R^{(n)}_t}{n+1}\right)\right)-m^{(n)}_{g,(2)}\right) \times\left(G^{-1}\left(\frac{R^{(n)}_t}{(n+1)}\right)+\sum_{k=1}^{s}G^{-1}\left(\frac{R^{(n)}_{t-k}}{n+1}\right)\prod_{m=t-k+1}^{t}\left(\beta_1 G^{-1}\left(\frac{R^{(n)}_m}{n+1}\right)+\vartheta_1\right)\right),$$

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

$$\underset{\sim}{\Delta}^{(n)}(\boldsymbol{\theta},g) := \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}\Delta^{(n)}_{(1)}(\boldsymbol{\theta},g)\\ \Delta^{(n)}_{(1)}(\boldsymbol{\theta},g)\end{pmatrix}. \tag{D.28}$$

As in the AR case of Hallin and Werker (2003) and the case of ARCH processes discussed in D.1, Lemma 1 in B.2 implies that $m^{(n)}_{g,(1)}$ and $m^{(n)}_{g,(2)}$ are $o\big(n^{-1/2}\big)$, hence can be omitted.

Here again, rather than $\underset{\sim}{\Delta}^{(n)}(\boldsymbol{\theta},g)$, which is associated with the unknown actual density g, the rank-based central sequence $\underset{\sim}{\Delta}^{(n)}(\boldsymbol{\theta},f)$ associated with the reference density f is to computed for R-estimation purposes.

D.3. AR(p)-ARCH(q)

Consider the process with dynamics

$$Y_t=\sum_{j=1}^{p}\vartheta_j Y_{t-j}+\left(1+\sum_{l=1}^{q}\beta_l Y^2_{t-j}\right)^{1/2}\epsilon_t, \tag{D.29}$$

where the ϵ_t's are i.i.d. with standardized density g, $\boldsymbol{\theta}=(\vartheta_1,...,\vartheta_p,\beta_1,...,\beta_q)$, 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:

$$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta},g) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{-\phi_g\left(Z_t(\boldsymbol{\theta})\right)}{\left(1+\sum_{l=1}^{q}\beta_l Y^2_{t-l}\right)^{1/2}}\begin{pmatrix}Y_{t-1}\\ \vdots\\ Y_{t-p}\end{pmatrix}'\quad\frac{\psi_g\left(Z_t(\boldsymbol{\theta})\right)}{1+\sum_{l=1}^{q}\beta_l Y^2_{t-l}}\begin{pmatrix}Y^2_{t-1}\\ \vdots\\ Y^2_{t-q}\end{pmatrix}'\right)'. \tag{D.30}$$

Using the results in Hansen (1991), one can show that there exists an asymptotically equivalent version of (D.30) 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 such a rank-based central sequence heuristically by (i) using a recurrence similar to (5.3) and (5.9), starting from $Y_0=0$; (ii) replacing Z_t in (D.30) 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 (5.5). 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 $\boldsymbol{\Gamma}^*(\boldsymbol{\theta},f)$ of $\underset{\sim}{\Delta}^{(n)}(\boldsymbol{\theta},f)$ under $\mathcal{H}^{(n)}_{\boldsymbol{\theta}}$ (which coincides with the semiparametric information matrix under $\mathcal{H}^{(n)}_{\boldsymbol{\theta},f}$) is of the form (5.11), with $\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\theta})$ derived as in Theorem 3.1 of Pantula (1988).

D.4. Discretely observed Cox-Ingersoll-Ross (CIR) processes

The CIR process is the solution to

$$dY_s = k(1 - Y_s)dt + \sigma\sqrt{Y_s}d\mathcal{W}_s, \tag{D.31}$$

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

$$\begin{aligned} Y_{th} &= (1 - \exp\{-kh\}) + Y_{(t-1)h}\exp\{-kh\} \\ &\quad + \left[Y_{(t-1)h}\frac{\sigma^2}{k}\left(\exp\{-kh\} - \exp\{-2kh\}\right) + \frac{\sigma^2}{2k}\left(1 - \exp\{-kh\}\right)^2\right]^{1/2}\epsilon_{th}, \end{aligned} \tag{D.32}$$

with i.i.d. standard normal ϵ_{th}'s. The resulting QL estimator typically exhibits quite large standard errors (see,e.g., Fan et al. (2014)) 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 (D.32) is an AR(1)-ARCH(1) model; this brings us back to (D.30).

D.5. Rank-based central sequence for the realized volatility numerical analysis

The rank-central sequence for the log-TSRV model of Section 6.1 are obtained using

(a) a Gaussian reference density f, yielding a rank-based central sequence $\underset{\sim}{\Delta}^{(n)}_{\mathrm{vdW}}$ which is a linear combination of the van der Waerden correlation coefficients

$$\underset{\sim}{r}^{(n)}_{\mathrm{vdW};i} = \left(s^{(n)}_{vdW}\right)^{-1}\left\{(n-i)^{-1}\sum_{t=i+1}^{(n)}\Phi^{-1}\left(\frac{R^{(n)}_t}{n+1}\right)\Phi^{-1}\left(\frac{R^{(n)}_{t-i}}{n+1}\right) - m^{(n)}_{vdW}\right\}, \tag{D.33}$$

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

$$\underset{\sim}{r}^{(n)}_{W;i} = \left(s^{(n)}_{W}\right)^{-1}\left\{(n-i)^{-1}\sum_{t=i+1}^{(n)}\left(\frac{R^{(n)}_t}{n+1} - \frac{1}{2}\right)\log\left(\frac{R^{(n)}_{t-u}}{n+1-R^{(n)}_{t-i}}\right) - m^{(n)}_{W}\right\}; \tag{D.34}$$

(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)

$$\begin{aligned} \underset{\sim}{r}^{(n)}_{L;i} = \left(S^{(n)}_{L}\right)^{-1}\Bigg\{(n-i)^{-1}\sum_{t=i+1}^{(n)}\mathrm{sign}\left(\frac{R^{(n)}_t}{n+1} - \frac{1}{2}\right)\left[\log\left(2\frac{R^{(n)}_{t-u}}{n+1}\right)I\left[\frac{R^{(n)}_{t-u}}{n+1} \le \frac{1}{2}\right]\right] \\ - \left[\log\left(2 - 2\frac{R^{(n)}_{t-u}}{n+1}\right)I\left[\frac{R^{(n)}_{t-u}}{n+1} > \frac{1}{2}\right]\right] - m^{(n)}_{L}\Bigg\}. \end{aligned} \tag{D.35}$$

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