Validity-Robust Estimation in Semiparametric Nonlinear Time Series Models

Marc Hallin
Université Libre de Bruxelles

London, December 11 2015

1.1 Gaussian dynamic location-scale models

Dynamic location-scale processes are essential tools in time series econometrics, with 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.

Probabilistic properties have been studied extensively and in great details; statistical analysis is less exhaustive, and still presents several challenges.

Among them 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.

In particular, QL estimators erroneously are surmised to be root-n consistent and asymptotically normal under very general conditions (actually, this requires, essentially, finite fourth-order moments).

The trouble is that those 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 rapidly deteriorates in the presence of skewness or excess kurtosis, two characteristics which are quite common in financial data.

Finally, Gaussian QL estimators are highly nonrobust, and 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, 2014) for LARCH and GARCH, ...

Remedies?

• Replacing the Gaussian reference density with more appropriate pseudo densities (e.g. Student ones), defining non-Gaussian QL estimators does not work: Fisher-consistecy under misspecified densities is lost, leading to root-n inconsistent estimates.

• The bootstrap approach by Hall and Yao (2003) is recovering, under certain conditions, the non-Gaussian asymptotic distribution of the Gaussian QL estimator but 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, Klaassen, Wellner and Ritov (actually restricted to independent observations; the time-series case is treated by Drost, Klaassen and Werker (1997)) under which the innovation density—call it g—remains unspecified is more realistic highly advisable.

Typical examples of that approach are Linton (1993), Wefelmeyer (1996), Drost and Klaassen (1997), Drost, Klaassen and Werker (1997), and Drost, and Werker (2004).

Standard as it is, not without serious difficulties: methodologically and computationally heavy; distinct possible semiparametric extensions inducing distinct efficient estimators, the validity of which depends on the semiparametric model adopted.

1.3 Discretely observed non-Gaussian Ornstein-Uhlenbeck processes (1)

A motivating example

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

$$dY_s = -\theta Y_s ds + dL_s.$$

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 Shepard (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. It can be shown that

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

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

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

In the classical case under which L_s is Brownian motion, ϵ_{th} is standard normal,

$$m(\theta)Y_{(t-1)h} = \mathrm{E}[Y_{th}|Y_{(t-1)h}] \quad \text{and} \quad v^2(\theta) = \mathrm{Var}[Y_{th}|Y_{(t-1)h}]$$

are the conditional mean and variance, respectively, of Y_{th}.

Call this the *Gaussian case*.

In the general Lévy-driven case, both the distribution of the Lévy process L_s and the value of θ enter the characterization of the discrete-time innovation density g, generating a complex class of possible distributions.

Several semiparametric extensions of the Gaussian case therefore have been considered in the literature: they all consider the model equation

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

with

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

and independently and identically distributed ϵ_{th}'s having density g, where either

independently and identically distributed ϵ_{th}'s having density g, where either

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

(ii) g 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 in the family $\mathcal{G}_{\text{HKW1}}$ of all densities (Hallin, Koell and Werker (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 in the family $\mathcal{G}_{\text{HKW2}}$ of all densities (Hallin, Koell and Werker (2000)) with (median zero and)

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

Call $\mathcal{E}_{(i)}, \ldots, \mathcal{E}_{(iv)}$, respectively, the resulting semiparametric models.

Note that the family $\mathcal{G}_0$ contains the other three, and contains the innovation densities of all discretized versions of the original process; 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 of $\mathcal{Y}$ has innovation density g.

The standard Bickel et al. 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

$$\int_{-\infty}^{\infty} (\dot{g}(z)/g(z))^2 g(z) dz < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} (1 + z\dot{g}(z)/g(z))^2 g(z) dz < \infty,$$

namely, g has finite Fisher information for location and for scale—which is less demanding, though, than finite fourth-order moments.

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

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

where $\epsilon_{th}(\theta) := (Y_{th} - m(\theta) Y_{(t-1)h})/v(\theta)$.

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.

• Step (b) requires a theoretical derivation of the so-called tangent space projection $\Delta^{*(n)}(\theta, g)$ (the *semiparametrically efficient*, at g and θ, central sequence) of $\Delta^{(n)}(\theta, g)$.

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.

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

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

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

$$\begin{aligned}\Delta^{*(n)}(\theta, g) \;=\; -\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Bigg\{ & \frac{\partial_\theta m(\theta)}{v(\theta)} \frac{\dot{g}(\epsilon_{th}(\theta))}{g(\epsilon_{th}(\theta))} Y_{(t-1)h} \\ & + c_g^{-1}(\theta) \partial_\theta v(\theta) \Big(v^2(\theta) \epsilon_{th}^2(\theta) - v(\theta) \Big) \\ & - \hat{\mu}_3 \epsilon_{th}(\theta) \Bigg\}\end{aligned}$$

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

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

(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:

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

This calls for several immediate remarks.

(1) First, the semiparametrically efficient central sequences $\Delta^{*(n)}(\theta, g)$ are considerably more complicated than $\Delta^{(n)}(\theta, g)$; their derivation is nontrivial, model-specific, to be performed on a case-by-case basis.

(2) Second, semiparametrically efficient central sequences depend on g and its derivative $\dot{g}$, both unknown. For $f \neq g$, typically,

$$\mathrm{E}_g[\Delta^{*(n)}(\theta, f)] \neq 0$$

(violating the Fisher consistency condition), so 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.

(3) 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.

(4) 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 $\mathcal{E}_{(i)}$.

The objective of this talk is to propose another semiparametric approach, based on residual ranks (the ranks of the ϵ_{th}'s), 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.

1.4 R-estimation: an alternative semiparametric approach

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).

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 $\mathbf{\Delta}^{*(n)}(\theta, \hat{g}^{(n)})$, constructing a root-$n$ consistent and asymptotically normal one-step R-estimator.

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 prior preferences or past experience.

• It also can be data-driven as soon as it only depends on the order statistic of 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.

Other attempts have been made to introduce R-estimation in the context of time series models: see, among others. 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-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), and have a clear link with invariance, hence with semiparametric efficiency at the reference density: see Hallin and Werker (2003).

We conclude by showing how our rank-based procedures apply and yield a root-n consistent and asymptotically normal estimator of the parameter θ, even in the presence of misspecified innovation density ($f \neq g$).

In the very general semiparametric experiment $\mathcal{E}_{(i)}$ (completely unspecified density g), the central sequence, at reference density f with cumulative distribution function F, takes the form

$$\Delta^{(n)}(\theta, f) = \sqrt{n} \sum_{i=1}^{n-1} \theta^{i-1} r_{f,i}^{(n)} - \frac{1}{\sqrt{n}} \frac{1}{8\theta^2 v^2(\theta)} \sum_{t=1}^{n} \left(1 + \epsilon_{th}(\theta) \frac{\dot{f}(\epsilon_{th}(\theta))}{f(\epsilon_{th}(\theta))} \right)$$

with (writing $\epsilon_t(\theta)$ for $\epsilon_{th}(\theta) := (Y_{th} - m(\theta) Y_{(t-1)h})/v(\theta)$)

$$r_{f,i}^{(n)} = \frac{1}{n-i} \sum_{t=i+1}^{n} \frac{\dot{f}\left(\epsilon_t(\theta)\right)}{f\left(\epsilon_t(\theta)\right)} \epsilon_{t-i}(\theta).$$

Projecting that central sequence onto the ranks yields

$$\underset{\sim}{\Delta}^{(n)}(\theta, f) = \sqrt{n} \sum_{i=1}^{n-1} \theta^{i-1} \underset{\sim}{r}_{f,i}^{(n)}$$

with

$$\underset{\sim}{r}_{f,i}^{(n)} = \frac{1}{s^{(n)}} \left[\frac{1}{n-i} \sum_{t=i+1}^{n} \frac{\dot{f}\left(F^{-1}\left(\frac{R_t^{(n)}}{n+1} \right) \right)}{f\left(F^{-1}\left(\frac{R_t^{(n)}}{n+1} \right) \right)} F^{-1}\left(\frac{R_{t-i}^{(n)}}{n+1} \right) - m^{(n)} \right],$$

where $R_t^{(n)}$ denotes the rank of $\epsilon_{th}^{(n)}(\theta)$ among $\epsilon_h^{(n)}(\theta), \ldots, \epsilon_{nh}^{(n)}(\theta)$, and $m^{(n)}$ and $s^{(n)}$ are exact standardizing constants.

In the particular case of a Gaussian reference density, $\underset{\sim}{r}^{(n)}_{f,i}$ takes the form of a *van der Waerden autocorrelation coefficient*

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

where $m^{(n)}_{\mathsf{vdW}} = O(n^{-1})$ can be omitted and

$$s^{(n)}_{\mathsf{vdW}} = \frac{1}{n} \sum_{j=1}^{n} \left(\Phi^{-1}\left(\frac{j}{n+1} \right) \right)^2 + O(n^{-1})$$

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 huge 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 $\underset{\sim}{\Delta}^{(n)}(\theta, f)$, which does not depend on g, 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 $f \neq g$.

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

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

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

The interpretation of $m(\boldsymbol{y}, \boldsymbol{\theta})$ and $v(\boldsymbol{y}, \boldsymbol{\theta})$ depends on 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
- if g is assumed to have median zero and interquartile range one, $m(\boldsymbol{y}, \boldsymbol{\theta})$ is the median, and $v(\boldsymbol{y}, \boldsymbol{\theta})$ the interquartile range, of Y_t conditional on $\mathbf{Y}_{t-1} = \boldsymbol{y}$.

Denote by $\mathrm{P}_{\boldsymbol{\theta},g}^{(n)}$ the joint distribution, under parameter value $\boldsymbol{\theta}$ and density g, of $\mathbf{Y}^{(n)}$

Since different (and mostly arbitrary) choices of $\mathcal{G}$ may lead to different tangent space projections and definitions of semiparametric efficiency we hereafter consider for $\mathcal{G}$ the most general choice $\mathcal{G}_0$ of the family of all nonvanishing densities over the real line.

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 $\mathrm{E}_{\boldsymbol{\theta},g}$ expectations under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},g}$, both $\mathrm{E}_{\boldsymbol{\theta},g}[\dot{\boldsymbol{m}}(\mathbf{Y}_{t-1}, \boldsymbol{\theta})]$ and $\mathrm{E}_{\boldsymbol{\theta},g}[\dot{\boldsymbol{v}}(\mathbf{Y}_{t-1}, \boldsymbol{\theta})]$ exist and are finite.

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

(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.

Denote by

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

the residuals associated with the parameter value $\boldsymbol{\theta}$. Clearly,

- the hypothesis $\mathcal{H}_{\boldsymbol{\theta}}^{(n)}$: parameter value is $\boldsymbol{\theta}$ holds iff the residuals $Z_t(\boldsymbol{\theta})$ are i.i.d.

- the hypothesis $\mathcal{H}_{\boldsymbol{\theta}}^{(n)}$: parameter value is $\boldsymbol{\theta}$ and innovation density is g holds iff the residuals $Z_t(\boldsymbol{\theta})$ are i.i.d. with density g

Defining

$$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g) := n^{-1/2} \sum_{t=1}^{n} \dot{\boldsymbol{l}}(Z_t, \mathbf{Z}_{t-1}, \boldsymbol{\theta}, g)$$

and

$$\boldsymbol{\Gamma}(\boldsymbol{\theta}, g) := \mathrm{E}_{\boldsymbol{\theta}, g}\left[\dot{\boldsymbol{l}}(Z_t, \mathbf{Z}_{t-1}, \boldsymbol{\theta}, g)\dot{\boldsymbol{l}}'(Z_t, \mathbf{Z}_{t-1}, \boldsymbol{\theta}, g)\right],$$

where

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

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 holds

Proposition. 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 $\boldsymbol{\tau}_n \in \mathbb{R}^p$,

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

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 $n \to \infty$.

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 $\boldsymbol{\Delta}^{(n)}(\hat{\boldsymbol{\theta}}^{(n)}, g)$-based one-step update of some preliminary root-n consistent estimator $\hat{\boldsymbol{\theta}}^{(n)}$:

$$\hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\boldsymbol{\Gamma}^{-1}(\hat{\boldsymbol{\theta}}^{(n)}, g)\boldsymbol{\Delta}^{(n)}(\hat{\boldsymbol{\theta}}^{(n)}, g)$$

Parametric efficiency, in general, cannot be reached in the semiparametric context where g is unspecified, and the best one can go for is the semiparametric efficiency bound $\boldsymbol{\Gamma}^{*-1}(\boldsymbol{\theta}, g)$.

The *semiparametrically efficient central sequence* $\boldsymbol{\Delta}^{*(n)}(\boldsymbol{\theta}, g)$, obtained by projecting the central sequence $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g)$ along the so-called *tangent space*, is the tool one needs to construct estimators that reach that semiparametric efficiency bound

- as the one-step update of some preliminary root-n consistent estimator $\hat{\boldsymbol{\theta}}^{(n)}$:

$$\hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\boldsymbol{\Gamma}^{*-1}(\hat{\boldsymbol{\theta}}^{(n)}, g)\boldsymbol{\Delta}^{*(n)}(\hat{\boldsymbol{\theta}}^{(n)}, g)$$

... this, however, is not implementable, as g is still unknown; accordingly, it is replaced with

$$\hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\boldsymbol{\Gamma}^{*-1}(\hat{\boldsymbol{\theta}}^{(n)}, \hat{g}^{(n)})\boldsymbol{\Delta}^{*(n)}(\hat{\boldsymbol{\theta}}^{(n)}, \hat{g}^{(n)})$$

where $\hat{g}^{(n)}$ is some kernel estimator of g.

As announced, in order to avoid kernel estimation of g, we rather are using the projection $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$ of $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)$ (f some chosen reference density) 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 interpreted as conditional expectation).

• Computing that projection is easy: it has been shown (Hallin and Werker 2003) that it can be obtained very simply by substituting, in $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)$,

$$F^{-1}\left(\frac{R_t(\boldsymbol{\theta})}{n+1}\right) \qquad \text{for} \qquad Z_t(\boldsymbol{\theta})$$

where $R_t(\boldsymbol{\theta})$ is the rank of $Z_t(\boldsymbol{\theta})$ among $Z_1(\boldsymbol{\theta})$, ..., $Z_n(\boldsymbol{\theta})$ (the so-called *approximate score form*).

No painful tangent space computation here! In case $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)$ involves infinitely many lagged $Z_t(\boldsymbol{\theta})$'s, adequate truncations can be performed.

- Obviously,

$$\mathrm{E}_f[\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)] = \mathrm{E}_f[\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)] = 0$$

(expectation of a conditional expectation). The distribution-freeness of ranks then entails (Fisher consistency) that also

$$\mathrm{E}_g[\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)] = 0,$$

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.

More precisely, it can be shown that

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

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

and

$$\lim_{n\to\infty} \underset{\sim}{\boldsymbol{\Gamma}}^{(n)}(\boldsymbol{\theta}, f) = \lim_{n\to\infty} \boldsymbol{\Gamma}^{*(n)}(\boldsymbol{\theta}, f) =: \boldsymbol{\Gamma}^{*}(\boldsymbol{\theta}, f),$$

where $\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}(\boldsymbol{\theta}, f)$ *is the variance (does not depend on* g*) of* $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}(\boldsymbol{\theta}, f)$ *and* $\boldsymbol{\Gamma}^{*}(\boldsymbol{\theta}, f)$ *is the* semiparametric information matrix *(at density* f*).*

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

4.1 Theoretical construction

As a test statistic, the quadratic form

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

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

Despite its simplicity and intuitive appeal, this definition, which in a much simpler context goes back to Hodges and Lehmann (1956), often runs into serious numerical difficulties related with the non-convex form of $\boldsymbol{\theta} \mapsto \underset{\sim}{Q}_{\mathrm{HL}}^{(n)}(\boldsymbol{\theta}, f)$, especially when the dimension of the parameter θ gets large.

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*.)

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*

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

(methods exist for the construction of such estimators), and

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

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

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

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

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

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

In particular, under under $\mathrm{P}_{\boldsymbol{\theta},f}^{(n)}$,

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

4.2 Implementation details.

- 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, Qi, and Xiu (2014) or, in the presence of outliers and data contamination, the bounded-influence estimators by Mancini, Ronchetti, and Trojani (2005).

The impact of that choice 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.

- 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 $\underset{\sim}{\boldsymbol{\theta}}_f^{(n)}$.

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.

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, 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

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

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.

Other parametric families of course can be considered, very much in the same way, such as the stable family, or the so-called *skew generalized error* family.

- Estimation of cross-information quantities

An important issue is the need, in part (ii) of Assumption (F), for a consistent estimator of the cross-information matrix

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

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.

In the present context, 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

$$\mathbf{\Gamma}(\boldsymbol{\theta}, f, g) = \mathcal{J}(f, g)\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\theta}),$$

where $\mathcal{J}(f, g)$ is a scalar quantity depending on f and g only, while $\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\theta})$ only depends on $\boldsymbol{\theta}$.

More generally, $\mathbf{\Gamma}(\boldsymbol{\theta}, f, g)$ is block-diagonal, with J blocks, each of which is enjoying a similar factorization.

A precise formulation of that simplifying assumption is as follows.

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

(G1) is block-diagonal, with J full-rank blocks of the form

$$\mathcal{J}_1(f,g)\boldsymbol{\Upsilon}_1^{-1}(\boldsymbol{\theta}), \ldots, \mathcal{J}_{\mathrm{J}}(f,g)\boldsymbol{\Upsilon}_{\mathrm{J}}^{-1}(\boldsymbol{\theta})$$

where the scalar *cross-information quantities* $\mathcal{J}_j(f,g)$ only depend on f and g, while the $\boldsymbol{\Upsilon}_j(\boldsymbol{\theta})$ matrices only depend on $\boldsymbol{\theta}$, $j = 1, \ldots, \mathrm{J}$;

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

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

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

where $\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\theta})$ is the asymptotic covariance matrix of the Gaussian quasi-likelihood estimator.

In particular, Assumption (G) holds with the above $\boldsymbol{\Gamma}(\boldsymbol{\theta}, f, g)$ as soon as $\mathcal{G}$ is restricted to symmetric (with respect to 0) densities—an assumption which is quite common in the literature.

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

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

Cassart, Hallin and Paindaveine (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

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

5.1 Discrete-time models

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

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

where the ϵ_t's are i.i.d. with 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

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

where $\boldsymbol{\theta} := (\theta_1, \ldots, \theta_q)$ and $Z_t(\boldsymbol{\theta}) := Y_t / \left(1 + \sum_{j=1}^{q} \theta_j Y_{t-1}^2\right)^{1/2}$.

The definition of a rank-based central sequence requires $(Y_{t-1}, \ldots, Y_{t-q})$ to be expressed in terms of a finite number of past shocks.

This is possible

- via a Volterra series expansion, or

- via a simple recurrence. 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 recursively yields the n-tuple $Z_1^{(n)}(\boldsymbol{\theta}), ..., Z_n^{(n)}(\boldsymbol{\theta})$, with ranks $R_1^{(n)}(\boldsymbol{\theta}), ..., R_n^{(n)}(\boldsymbol{\theta})$. 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

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

then define

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

where

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

It can be shown that $m_{f,(2)}^{(n)}$ is $o\left(n^{-1/2}\right)$. 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 by Kristensen and Rahbek (2005) (Theorem 2.1).

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

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

where the ϵ_t's are i.i.d. with density g, and $\boldsymbol{\theta} = (\vartheta_1, ..., \vartheta_p, \beta_1, ..., \beta_q)$. (If standard conditions are satisfied), the ULAN central sequence for $\boldsymbol{\theta}$ reads

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

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

In the AR(1)-LARCH(1) case, with dynamics

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

which is ULAN with central sequence

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

where $Z_t = Z_t(\boldsymbol{\theta}) := (Y_t - \vartheta Y_{t-1})/(1 + \beta Y_{t-1})$, this is achieved, again, either via a truncated Volterra expansion of Y_t in terms of its innovations, or a recurrence producing a rank-based "reconstruction" $\underset{\sim}{Y}_t$ of Y_t. This yields

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

where

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

and

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

It can be shown that both $m_{f,(1)}^{(n)}$ and $m_{f,(2)}^{(n)}$ are $o\left(n^{-1/2}\right)$, hence can be omitted for n large. 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

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

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

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 specifiy any innovation density) and rely on the standard Bickel et al. semiparametric estimation method.

We rather propose here a class of R-estimators for those duration models. First note that the ULAN central sequence for ACD models actually is that of a dynamic scale model for Y_i:

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

where $Z_i(\boldsymbol{\theta}) = Y_i / \Psi_{i-1}(\boldsymbol{\theta})$.

Similar to the recursion 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

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

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

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

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 central to the financial literature. 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.

(d) Discretely observed mean-reverting jump diffusion

In the class of Lévy processes, let us consider the Poisson-Gaussian process $\mathcal{Y}$, which is solution to equation

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

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 $\{Y_0, Y_h, Y_{2h}, ..., Y_{nh}\}$,

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

and

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

That class of models has been considered by Das (2002) in the dynamic analysis of bond markets. 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 of the dynamic location-scale model

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

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 are possible. The situation is actually pretty much the same as in the motivating example and, for the same reasons, turning to the residual ranks appears as the safest attitude.

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

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

where

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

with

$$\beta_1(\boldsymbol{\theta}) = \frac{1}{2}\left[h(\mathrm{Coth}(h\kappa)-1) - \frac{\eta^2}{\kappa\left(2\kappa+\eta^2\right)}\right], \qquad \beta_2(\boldsymbol{\theta}) = \alpha(1-\exp(h\kappa)+h\kappa)$$

$$\beta_3(\boldsymbol{\theta}) = \kappa^{-3/2}h\exp(-h\kappa/2)\left[\left(2\kappa+\eta^2\right)\mathrm{Sinh}(h\kappa)\right]^{-1/2}, \qquad \beta_4(\boldsymbol{\theta}) = h\left(1-\exp(-h\kappa)\right)^{1/2}$$

and $\beta_5(\boldsymbol{\theta}) = \eta/\left(2\kappa+\eta^2\right)$ (Coth(x) and Sinh(x) as usual stand for the hyperbolic cotangent and sinus of x, respectively).

Canceling $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, g)$ 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. style.

The R-estimation methods developed here thus naturally enter into the picture. Projecting $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)$ (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} remains completely unspecified.

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 $Z_t(\boldsymbol{\theta})$. 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)

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

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 (**??**) with completely unspecified innovation density g.

Our method 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, stochastic volatility, or asset pricing models. It is the solution to the stochastic differential equation

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

In Chapter 2 of Singleton (2009), a QL estimator is defined setting a Gaussian dynamic location-scale model for the discrete-time version of $\mathcal{Y}$. It can be shown that 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.

6.1 Asymptotic Relative Efficiencies (AREs)

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

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

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.

We compute AREs, with respect the QL of the R-estimators based on the van der Waerden ($\underset{\sim}{\Delta}^{(n)}_{\text{vdW}}$; normal f), Wilcoxon ($\underset{\sim}{\Delta}^{(n)}_{W}$; logistic f), and Laplace ($\underset{\sim}{\Delta}^{(n)}_{L}$; double-exponential f) rank-based central sequences, under densities g in the Johnson family (Jones and Pewsey (2009) and Ghysels and Wang (2011)).

Denote by $J_{SU}(\gamma, \delta, \mu, \sigma)$—where γ and δ are skewness and kurtosis parameters, respectively, while μ and σ, as usual, stand for location and scale— the general density in that family.

For $\mu = 0$ and $\sigma = 1$, suitable values of δ and γ are may lead to positive skewness values as large as ten, and excess of kurtosis larger than eight.

$J_{SU}(\gamma, \delta, \mu, \sigma)$

	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

6.2 Finite-sample performances (in the presence of skewness and leptokurtosis)

Let us consider the semiparametric ARCH(1) model

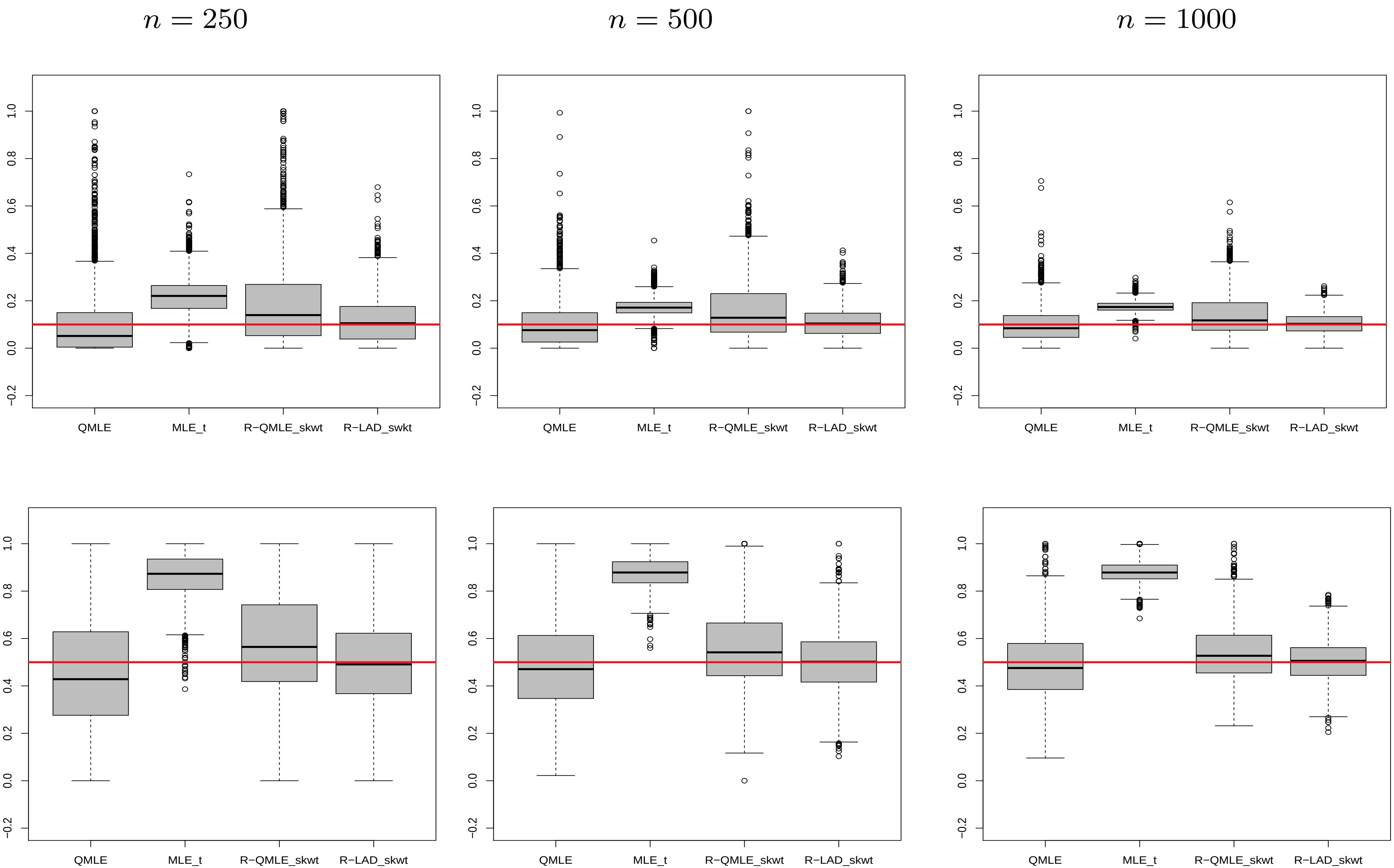
$$Y_t = (1 + \theta Y_{t-1}^2)^{1/2} \epsilon_t,$$

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 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.

- From each realization, we computed (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.



$n = 250$
$n = 500$
$n = 1000$
QMLE
MLE_t
R−QMLE_skwt
R−LAD_skwt
−0.2
0.0
0.2
0.4
0.6
0.8
1.0

6.3 An empirical analysis of the USD/CHF exchange rate

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 (Aït-Sahalia et al. 2005).

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.

From the high-frequency quotes, we compute 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.

• *Log-returns.* 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.

(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".

	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

USD/CHF FX rates: descriptive statistics

	1993		1997	
	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

USD/CHF FX rates: descriptive statistics

• The autocorrelation analysis (unreported) of the training data suggests that AR(p) processes, $p \leq 3$, are suitable models fot log(TSRV). Thus, we set that the conditional mean of the log(TSRV) is of the form $\sum_{j=1}^{3} \theta_j \log(\text{TSRV}_{t-j})$.

• We thus set up a normal mean-variance mixture model, of the form

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

with TSRV_t playing the role of ς_t.

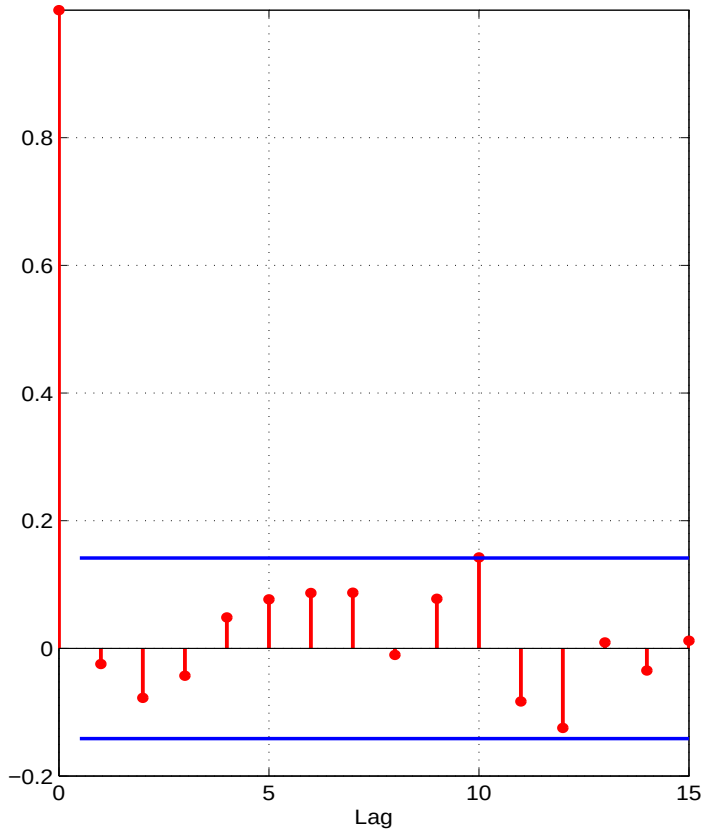
• We estimate the model parameters from the data in each training period, and assess the quality of the various estimates—the Gaussian QL, and the van der Warden (vdW), Wilcoxon (W), and Laplace (L) R-estimators—via their asymptotic standard errors.

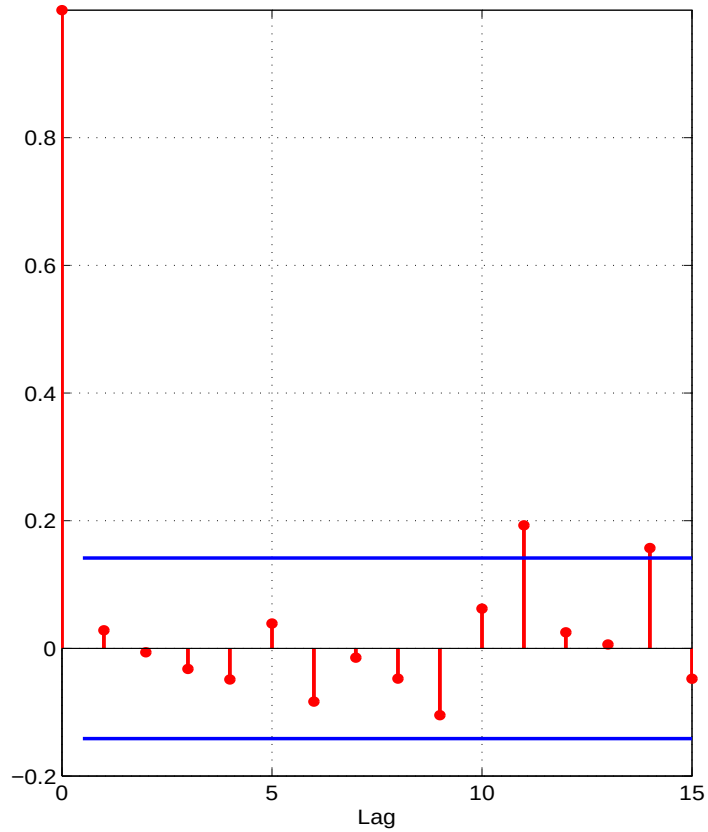
	1993				1997			
	QL	vdW	W	L	QL	vdW	W	L
θ_1	0.2762	0.3204	0.3525	0.4014	0.3719	0.3517	0.3677	0.3921
	(0.072)	(0.051)	(0.070)	(0.045)	(0.071)	(0.063)	(0.080)	(0.077)
θ_2	0.0969	0.0988	0.0768	0.0190	0.1323	0.1586	0.1408	0.1761
	(0.074)	(0.061)	(0.075)	(0.048)	(0.076)	(0.066)	(0.085)	(0.081)
θ_3	-0.0371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

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).

To assess the validity of the fitted models, we considered a standard correlogram analysis of residuals. Below, we plot the sample residual autocorrelations of residuals and squared residuals, implied by the Laplace R-estimator for the training period January-September 1993. None of the plots provide any evidence of autocorrelation outside Bartlett's two-standard-error bands for white noise.

Similar plots (unreported) for the QL estimator and the other R-estimators, yield similar but less good results.





Forecasting. We computed, for each day in the October-December period (still 1993 and 1997), the squared one-day-ahead prediction errors of forecasts based on each estimator computed from the training period.

The table below 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.

	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%

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.

8. Conclusions

The methods we are proposing are

- valid (root-n consistent and asymptotically normal) under a much broader family of densities than Gaussian QL estimators;

- simple (no derivation of tangent spaces, no need of kernel estimation of innovation densities);

- extremely flexible (possibility of data-driven scores);

- efficient (depending on the choice of scores, between semiparametrically efficient at selected f and "close to semiparametrically efficient" (uniformly in g);

- have a better resistance to outliers and extreme observations than Gaussian QL estimators.

They definitely should enter econometrics daily practice.