A Simple R-estimation method for semiparametric duration models

Marc Hallin\textsuperscript{a}, Davide La Vecchia\textsuperscript{b,*}

\textsuperscript{a} ECARES, Université libre de Bruxelles, Belgium
\textsuperscript{b} Research Center for Statistics, Geneva School of Economics and Management, University of Geneva, Switzerland

\textbf{A B S T R A C T}

Modeling nonnegative financial variables (e.g., durations between trades or volatilities) is central to a number of studies across econometrics, and still poses several statistical challenges. Among them, the efficiency aspects of semiparametric estimation remain pivotal. In this paper, we concentrate on estimation problems in autoregressive conditional duration models with unspecified innovation densities. Exponential quasi-likelihood estimators (QMLE) are the usual practice in that context, since they are easy-to-implement and preserve Fisher-consistency. However, the efficiency of the QMLE rapidly deteriorates away from the reference exponential density. To cope with the QMLE's lack of accuracy, semiparametrically efficient procedures have been introduced. These procedures are obtained using the classical tangent space approach; they require kernel estimation and quite large sample sizes. We propose rank-based estimators (R-estimators) as a substitute. Just as the QMLE, R-estimators remain root-$n$ consistent, irrespective of the underlying density, and rely on the choice of a reference density (which, however, needs not be the exponential one), under which they achieve semiparametric efficiency. Moreover, R-estimators neither require tangent space calculations nor kernel estimation. Numerical results illustrate that R-estimators based on the exponential reference density outperform the QMLE under a large class of actual innovation densities, such as the Weibull or Burr densities. A real-data example about modeling the price range of the Swiss stock market index concludes the paper.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

1.1. Rank-based inference in time series

Exact inference is a red thread running through Jean-Marie Dufour’s entire scientific life, providing a strong coherence to his otherwise quite broad and diverse list of contributions. That interest in exact inference can be traced back to his early career publications in rank-based inference. Rank-based methods, in the early eighties, were essentially limited to linear models with independent observations (more precisely, to models driven by exchangeable noise), and Jean-Marie Dufour pioneered the idea of using them in a time series and econometric context: see Dufour (1981), Dufour et al. (1982) and Dufour and Roy (1985) on rank-based testing against serial dependence, or Dufour and Hallin (1987) on rank-based tests for AR(1) coefficients.

* Corresponding author.

E-mail addresses: mhallin@ulb.ac.be (M. Hallin), davide.lavecchia@unige.ch (D. La Vecchia).

https://doi.org/10.1016/j.jeconom.2020.04.036

0304-4076/© 2020 Elsevier B.V. All rights reserved.
Since then, rank-based tests have been extended, quite considerably, into several directions and increasingly sophisticated time-series models, starting with the linear ARMA ones (Hallin et al., 1985; Hallin and Puri, 1994), bilinear models (Benghabrit and Hallin, 1992, 1996), VARMA (Hallin et al., 1989) and elliptical VARMA models (Hallin and Paindaveine, 2002, 2004b,a, 2005), and, more recently, ARCH-GARCH models (Andrew and Werker, 2015), unit-root, and cointegration problems (Hallin et al., 2006, 2011). The development of R-estimation was somewhat slower (Koul and Saleh, 1993; Andrews, 2012; Mukherjee and Bai, 2002), but now includes R-estimators for a variety of linear and nonlinear time-series models (see Hallin and La Vecchia, 2017), including e.g. AR-ARCH or discretely observed diffusions with jumps.

In this paper, we develop R-estimation for autoregressive conditional duration (ACD) models, or, more generally, in the setting of multiplicative error models. While the main theoretical tools needed for the development of our R-estimators are those described in Hallin and La Vecchia (2017), the construction of R-estimators in the ACD context is not straightforward: a key assumption in Hallin and La Vecchia (2017) on the structure of cross-information matrices indeed does not hold here, and covariance and information matrices do not admit explicit forms. The methods proposed in Hallin and La Vecchia (2017) thus need to be adapted to the ACD context.

1.2. Duration models in econometrics and finance

Duration is commonly defined as the time interval between consecutive events. The autoregressive conditional duration (ACD) model was proposed by Engle and Russell (1998) to analyze irregularly spaced (ultra) high-frequency financial transaction data. Since then, duration models have been widely applied in economics and finance; see the seminal papers by Dufour and Engle (2000), Engle (2000), and the related literature. A short survey about possible applications in economics and finance is available in Pacurar (2008), while Tsay (2005) provides a book-length introduction to ACD models.

The growing interest in duration models is related to the availability of financial data recorded at high frequency, like e.g. tick-by-tick data. The high-frequency data has inspired research across finance, econometrics and statistics. We refer to Hasbrouck (2007), Aït-Sahalia and Jacod (2014) and Hautsch (2011).

Besides intra-day time intervals between consecutive financial transactions, duration models, which are multiplicative error models (see Engle, 2002b), more generally apply to time series with non-negative observations. As an example, we mention the daily range $g$ of an asset, a measure which is applied to proxy the asset volatility (see, e.g., Parkinson, 1980; Kunitomo, 1992 and related literature). Chou (2005) considers the class of conditional autoregressive range (CARR) model is equivalent to the ACD(1,1) model, hence duration models can be effectively applied to conduct inference on asset volatilities. Similar considerations apply to the autoregressive conditional volume process introduced by Manganelli (2005). Moreover, ACD models not only apply to different types of financial variables, but also to different types of sampling schemes. Indeed, in case of financial durations, ACD models are suitable when the process under study is observed in event time, so that observations are irregularly spaced in time. However, the same ACD models also are effective in the analysis of processes in calendar time, such as aggregated trading activities (e.g., daily volumes) based on equidistant time intervals.

1.3. R-estimation in duration models

The statistical analysis of duration models still raises problems related, mainly, with the accuracy of model parameter estimators. The standard duration model (see Section 3.1) depends on an unspecified Euclidean parameter of interest, and is driven by unobserved shocks $\epsilon_t$—call them innovations—with innovation density $g$, say. That innovation density typically should remain unspecified, in which case the model is of a semiparametric nature.

Innovation densities, in duration models, are supported on $(0, \infty)$ rather than the real line; obvious candidates are the exponential, Weibull, Burr, or Gamma densities, generally with increasing failure rates. Although it can be argued that exponential distributions, being typical for waiting times, are somehow “natural” in the context, empirical studies (including Engle and Russell, 1998 and Engle, 2000) indicate that the assumption of exponential innovations is quite unlikely to hold in practice. Drost and Werker (2004) nevertheless show that, in case the $\epsilon_t$’s are i.i.d. with density $g$ in a broad class $G$ of densities, exponential quasi-likelihood estimators (henceforth, QMLEs) remain root-$n$ consistent irrespective of $g \in G$. In contrast, quasi-maximum likelihood estimators based on log-normal, Burr, or Weibull densities—namely, M-estimators based on the maximization of (misspecified) log-normal, Burr, or Weibull likelihoods—fail to be root-$n$ consistent under misspecification.

Even if empirical evidence strongly suggests that actual innovation densities are not exponential, QMLEs have become standard practice, thanks also to their simplicity. Unfortunately, QMLEs also can suffer from a very significant efficiency loss away from the reference exponential density. The usual remedy to that major drawback is the classical semiparametric approach associated with the names of Bickel, Klaassen, Wellner and Ritov; see Bickel et al. (1993) and Chapter 25 of Vaart (1998) for the case of independent observations, or Drost et al. (1997) for the time-series case. In that approach, estimation

---

1 The range is, roughly speaking, the difference, in log scale, between the highest and the lowest price at which the stock is traded during a given day.
equations are based on semiparametric score functions resulting from the so-called tangent space projections, after plugging in a nonparametric estimator of the actual density $g$. Building on Drost et al. (1997), Drost and Werker (2004) (see also Ranasinghe and Silvapulle, 2011) apply this method in the context of ACD model, not only relaxing the specification of the innovation density $g$, but even giving up the assumption of i.i.d. innovations (replacing it with a full range of dependence structures yielding for $\epsilon_i$ a conditional density $g_i$ itself depending on the past). Droste and Werker derive semiparametrically efficient estimators using (Bickel et al., 1993) methodology. However, that success comes at a cost: besides its analytical sophistication, the semiparametric method requires the estimation, by means of kernel methods, of the density $g$ or the conditional densities $g_i$, their derivatives, and some (conditional) expectations. Besides requiring large samples, this implies the delicate selection of tuning parameters, which (may) have a strong impact on the actual value of the estimators. Sample splitting is generally required as well.

In this paper, we adopt a simpler semiparametric approach, based on R-estimation (via residual ranks): no kernel density estimation and no sample-splitting here. On the other hand, unlike Droste and Werker (2004), we are restricted to the case of i.i.d. (actually exchangeable) $\epsilon_i$’s. Just as QMLEs, our R-estimators rely on the choice of some reference density $f$, under which R-estimators achieve semiparametric efficiency. The density $f$ needs not to coincide with the actual unspecified $g$ and it is not restricted to be exponential for the resulting R-estimators to remain root-$n$ consistent under $g \neq f$. In our numerical exercises, we illustrate that using as $f$ the traditional exponential reference density $f_{exp}$ yields a simple-to-implement R-estimator, that outperforms the QMLE, under commonly used innovation densities (like e.g. the Weibull).

2. Motivating example

Let us consider the simple ACD(1,0) model, with equation

$$Y_i = \epsilon_i \psi_{i-1}, \quad \psi_{i-1} = 1 + \beta Y_{i-1}, \quad i \in \mathbb{Z},$$

(1)

where $\epsilon_i$ has a Weibull density $g = g^W_\zeta$ with scale one and shape parameter $\zeta \geq 1$.

Trajectories for this model are easily simulated for given values of $\beta$ and $\zeta$, and can be used in a numerical comparison of the efficiencies, as estimators of $\beta$ under such Weibull innovations, of

(a) the exponential QMLE,
(b) the correctly specified Weibull maximum likelihood estimator (MLE), and
(c) (anticipating on the results of the next sections) the R-estimator with exponential reference density

as functions of $\zeta$. Those three estimators are root-$n$ consistent and asymptotically normal, so that their efficiencies are captured by their asymptotic variances. We perform our comparison under different tailweights, as obtained by varying the shape parameter $\zeta$, which produces different degrees of skewness and kurtosis in the innovation density $g^W_\zeta$, note that for $\zeta = 1$ the Weibull density coincides with the exponential while, for $\zeta > 1$, it belongs to the class of increasing failure rate densities. To visualize the variety of shapes thus considered, we first plot in Fig. 1 (left and middle panels) the Weibull densities $g^W_\zeta$ associated with some representative values of $\zeta$, and the corresponding failure rates.

Fig. 1. Left panel: a plot of Weibull densities $g^W_\zeta$ for various values of the shape parameter $\zeta$ and adjusted scale to preserve mean one. Middle panel: failure rates of $g^W_\zeta$, for different values of $\zeta$ and adjusted scale to preserve mean one. Right panel, dot-dashed line: ARE, under $g^W_\zeta$, of the MLE with respect to the QMLE of $\beta$, for different values of $\zeta$ in abscissa. Solid line: ARE, under $g^W_\zeta$, of the MLE with respect to the R-estimator of $\beta$, for different values of $\zeta$ in abscissa.
Next, we simulated, over a grid of \( \zeta \) value in [1 2.2], 1500 trajectories for \( \{Y_i^\zeta, \ i = 1, \ldots, 3500\} \) from the ACD(1,0) model (1), with \( \beta = 0.5 \). For each trajectory, we computed the MLE and QMLE (using the R routine acdFit\(^2\)), and the R-estimator (see Section 5) with exponential reference density. For each value of \( \zeta \), we evaluated the variances, over the 1500 replications, of those three estimators, and the ratios of the QMLE and R-estimator variances, respectively, to the variance of the MLE (the parametric efficiency bound). The resulting ratios are finite-sample evaluations of the asymptotic relative efficiencies \( \text{ARE}_\zeta (\text{MLE/QMLE}) \) and \( \text{ARE}_\zeta (\text{MLE/R-Estim}) \) of the MLE and the R-estimator with respect to the QMLE, under Weibull innovation density \( g_\zeta \).

The results are shown in Fig. 1 (right panel), where those AREs are plotted against the values of \( \zeta \). The plots indicate that, whereas the QMLE performs as well as the MLE for \( \zeta = 1 \), its performances rapidly deteriorate as \( \zeta \) moves away from one, with an \( \text{ARE}_\zeta (\text{MLE/QMLE}) \) value larger than 4 for \( \zeta > 2 \). Thus, we conclude that blindly relying on exponential quasi-maximum likelihood can be quite costly in terms of accuracy.

3. The autoregressive conditional duration model

3.1. General setting

As in the seminal paper of Engle and Russell (1998), let \( Y_t \) denote the duration between some \((i-1)\)th and \(i\)th events (typically, the time elapsed between two successive transactions of some asset). We say that the process \( \{Y_t\}_{t \in \mathbb{Z}} \) satisfies the assumption of the accelerated time or autoregressive conditional duration (ACD\((p, q)\)) model with parameters \( \theta = (\alpha, \beta, \gamma)' = (\alpha, \beta_1, \ldots, \beta_p, \gamma_1, \ldots, \gamma_q)' \) if

\[
Y_t = \epsilon_t \psi_{t-1}, \quad \text{with} \quad \psi_{t-1} = \alpha + \sum_{j=1}^p \beta_j Y_{t-j} + \sum_{l=1}^q \gamma_l \psi_{t-l-1}, \quad i \in \mathbb{Z}
\]

(2)

where the \( \epsilon_t \)s are i.i.d., with nonvanishing (over \((0, \infty)\)) density \( g \), and \( \epsilon_t \) independent of \( \psi_{t-1}, \psi_{t-2}, \ldots \). Clearly, under that model, and provided that \( E(\epsilon_t) = 1 \), the conditional mean duration time \( E(Y_t|F_{t-1}) \), where \( F_t \) stands for the \( \sigma \)-field generated by \( Y_1, Y_2, \ldots, Y_t \), is \( \psi_{t-1} \). Exogenous variables also can be included in the expression for \( \psi_{t-1} \), which allows to capture the influence, on the distribution of future durations, of observable covariates. The theory developed in this paper can easily accommodate for the presence of such covariates but, for the sake of notational simplicity, we do not pursue with this.

Let \( Y^{(n)} := (Y_1, \ldots, Y_n) \) be the observed finite realization of some solution of (2). Let us assume we also observe initial values \( Y_{-p+1}, \ldots, Y_{-1}, Y_0 \) and \( (\psi_q, \ldots, \psi_1) \), and base our (asymptotic) analysis on the likelihoods of \( Y^{(n)} \) conditional on those initial values. If \( \theta \) is such that a stationary solution exists (see page 228 of Tsay, 2005 for positivity and stationarity conditions), the asymptotic influence of those starting values is nil, and we safely can put \( Y_{-p+1} = \ldots = Y_{-1} = Y_0 = 0 \), along with \( \psi_q = \cdots = \psi_1 = 0 \) (so that \( \psi_0 = 1 \)). In practice, however, a more reasonable choice would be, for instance, \( Y_{-p+1} = \ldots = Y_{-1} = Y_0 = Y \), where \( Y := n^{-1} \sum_{i=1}^n Y_i \) is an unbiased estimator of the unconditional stationary mean \( EY_t \) of \( Y_t \). Moreover, the law of iterated expectations implies that \( E\psi_t = EY_t \), so that \( Y \) is an estimator of the unconditional stationary mean of \( \psi_t \). The global influence of those starting values is asymptotically negligible, and so is the mild dependence induced by \( Y \).

Based on those initial values, \( \psi_t \), in view of (2), can be expressed linearly (the coefficients are products and powers of \( \alpha \)'s, \( \beta \)'s and \( \gamma \)'s) as a function of the past observations and the parameter value \( \theta \):

\[
\psi_t = \psi(Y_t, Y_{t-1}, \ldots, Y_1; \theta) \quad \text{for} \quad i = 1, \ldots, n.
\]

For any given values of the \( Y_t \)'s, that function \( \psi_t \) clearly is differentiable with respect to \( \theta \), with gradient \( \hat{\psi}_t := \text{grad}_\theta \psi_t \), say, taking values \( \hat{\psi}_t \). Obtaining exact analytical expressions for \( \psi_t \) and \( \hat{\psi}_t \) seems difficult. However, both \( \psi_t \) and \( \hat{\psi}_t \) can be computed recursively from the initial values \( (Y_{-p+1}, \ldots, Y_{-1}, Y_0) \) and \( (\psi_q, \ldots, \psi_1) \), along with \( (\psi_{t-1}, \ldots, \psi_{t-q}) \), for \( \hat{\psi}_t \); we refer to Section 3.2 for examples of the linear recursions involved.

3.2. Relevant special cases

For \( p = q = 1 \) (the ACD(1,1) model), which is, by far, the ACD model most frequently considered in practice, the estimation problem for \( \theta = (\alpha, \beta, \gamma)' \) is pretty similar to that of the parameters of the GARCH(1,1) model considered in Bollerslev (1986). Positivity and the stationarity conditions require nonnegative values of \( \beta \) and \( \gamma \) such that \( \beta + \gamma < 1 \), so that \( O(\gamma^j) \) quantities below go to zero exponentially fast as \( i \to \infty \), and \( \psi_t = \psi(Y_t, Y_{t-1}, \ldots, Y_1; \theta) = \alpha(1 - \gamma)^{-1} + \beta \sum_{j=0}^{\infty} \gamma^j Y_{t-j} + O(\gamma^j) \). Moreover, differentiating the second equation in (2) with respect to \( \theta \) yields

\[
\hat{\psi}_t = \gamma \hat{\psi}_{t-1} + (1, Y_t, \psi_{t-1})'.
\]

\(^2\) Calculation performed in R studio, with R version 3.2.5 (2016-04-14), year 2016, platform x86_64-w64-mingw32. Processor: Intel(R) Core(TM) i7-3770 CPU @ 3.40 GHz.
which provides a linear recursion for $\Psi_i$. Again, the asymptotic influence of initial values is nil, hence they safely can be put to zero. For $p = 1$ and $q = 0$ (ACD(1,0) case), we obtain (with $\theta := (\alpha, \beta)$)

$$\Psi_i = \psi_i(Y_i, Y_{i-1}, \ldots, Y_1; \theta) = \alpha + \beta Y_i$$

and, from (3), $\dot{\Psi}_i = \dot{\psi}_i(Y_i, Y_{i-1}, \ldots, Y_1; \theta) = (1, Y_i)'$.

3.3. Assumptions

Our objective is to conduct inference about $\theta$ while avoiding both a complete parametric specification of the data-generating mechanism and a kernel estimate of the underlying innovation density. Throughout, we assume that (2) holds, with unspecified parameter value $\theta$ and innovation density $g$ in the class $\mathcal{G}$ of densities that do not vanish on $(0, \infty)$ and satisfy a few regularity assumptions described below. The resulting statistical models thus are semiparametric ones, with Euclidean parameter of interest $\theta$, and infinite-dimensional nuisance $g$.

Denoting by $P_{\theta, g}^{(n)}$ the joint distribution, under (2), of $Y^{(n)}$ (conditionally on the initial values, as explained in Section 3.1), consider the semiparametric families $\mathcal{P}^{(n)} = \{P_{\theta, g}^{(n)} : \theta \in \Theta, g \in \mathcal{G}\}$, $n \in \mathbb{N}$, where $\Theta$ and $\mathcal{G}$ are such that the following assumptions (Assumptions (A) and (B), but also Assumption (C) stated in Section 4.1) hold.

Assumption (A). The ACD model (2) admits a positive and stationary solution.\(^3\)

Assumption (B). For all $z \in \mathbb{R}^+$, the density $g(z)$ is strictly positive, and $\int_{\mathbb{R}^+} zg(z) \, dz = 1$.

(B2) The mapping $z \rightarrow g(z)$ is absolutely continuous on finite intervals, i.e. there exists an a.e. derivative $\dot{g}$ such that, for all $0 \leq a < b < \infty$, $g(b) - g(a) = \int_a^b \dot{g}(z) \, dz$.

(B3) Letting $\phi_g(z) := -\dot{g}(z)/g(z)$ and $\varphi_g(z) := z \phi_g(z) - 1$, the Fisher information $\mathcal{J}(g) := \int_{\mathbb{R}^+} \varphi_g^2(z) g(z) \, dz$ for scale exists and is finite.

Note that, under those assumptions, the gradients $\dot{\Psi}_i := \nabla \psi_i$ have finite expectations $E_{\theta, g} [\dot{\Psi}_i]$ for any $g \in \mathcal{G}$ and $\theta \in \Theta$.

Let $\mathcal{H}^{(n)}_{\theta, g}$ and $\mathcal{H}^{(n)}_{\theta, g}$ denote (for given $g \in \mathcal{G}$ and $\theta \in \Theta$) the simple hypothesis $\{P_{\theta, g}^{(n)}\}$ and the nonparametric collection $\{P_{\theta, g}^{(n)} : g \in \mathcal{G}\}$, respectively. Writing

$$Z_i(\theta) := \frac{Y_i}{\psi_{i-1}(Y_{i-1}, Y_{i-2}, \ldots, Y_1; \theta)}$$

for the residuals associated with the parameter value $\theta$, the hypotheses $\mathcal{H}^{(n)}_{\theta, g}$ and $\mathcal{H}^{(n)}_{\theta, g}$ clearly hold true iff the residuals $Z_i(\theta)$ (which coincide with $\epsilon_i$) are i.i.d., and iff they are i.i.d. with density $g$, respectively. For all $\theta$, the observation $Y_i$ can be expressed reversibly in terms of the initial values\(^4\) and the residuals $Z_i(\theta), Z_{i-1}(\theta), \ldots, Z_1(\theta)$: for each $Y_i$, a linear combination, with (exponentially decreasing, under the positivity/stationarity constraints on $\theta$) coefficients involving products of $\beta$’s and $\gamma$’s, of products of residuals, of the form $Z_i(\theta), Z_i(\theta)Z_{i-1}(\theta), Z_i(\theta)Z_{i-1}(\theta)Z_{i-2}(\theta), \ldots, Z_i(\theta)Z_{i-1}(\theta)\cdots Z_1(\theta)$. Letting $Z_i^{(n)}(\theta) := (Z_i(\theta), Z_{i-1}(\theta), \ldots, Z_1(\theta))$, write, with a slight abuse of notation, $\psi(Z_i^{(n)}(\theta); \theta)$ and $\dot{\psi}(Z_i^{(n)}(\theta); \theta)$ for $\psi(Y_i, Y_{i-1}, \ldots, Y_1; \theta)$ and $\dot{\psi}_i(Y_i, Y_{i-1}, \ldots, Y_1; \theta)$, respectively, when the $Y_i$’s are expressed as functions of the $Z_i(\theta)$’s. In case the $Z_i(\theta)$’s and $\epsilon_i$’s coincide, the same quantities will be denoted, with $\epsilon_i := (\epsilon_i, \epsilon_{i-1}, \ldots, \epsilon_1)$, as $\psi(Z_i(\theta); \theta)$ and $\dot{\psi}(Z_i(\theta); \theta)$, respectively.

4. Local asymptotic normality, semiparametric efficiency, and ranks

In this section, we introduce the main methodological tools to be used in the sequel. First, we restate the uniform local asymptotic normality (ULAN) as in Drost et al. (1997) and Drost and Werker (2004), with central sequence $\Delta^{(n)}(\theta, g)$, of the parametric fixed-$g$ submodels $P_{\theta}^{(n)} := \{P_{\theta}^{(n)} : \theta \in \Theta\}$. Then, following Hallin and Werker (2003), we project $\Delta^{(n)}(\theta, g)$ onto the $\sigma$-algebra generated by the ranks of the residuals $Z_i(\theta)$. The proofs of the results follow along the general lines developed in Hallin and La Vecchia (2017).

4.1. Uniform local asymptotic normality

Define

$$\Delta^{(n)}(\theta, g) := n^{-1/2} \sum_{i=1}^n I(Z_i^{(n)}(\theta); \theta, g)$$

and $\Gamma(\theta, g) := E_{\theta, g} \left[ I(Z_i^{(n)}(\theta); \theta, g) \dot{I}(Z_i^{(n)}(\theta); \theta, g) \right]$.

---

\(^3\) See Tsay (2005) for sufficient conditions.

\(^4\) Namely, $(Y_{i+p+1}, \ldots, Y_i)$ and $(\psi_{i-q}, \ldots, \psi_{i-1})$. 
where
\[ \hat{\theta}(Z^n; \theta, g) := \frac{\psi_g(Z^n; \theta)}{\psi_{g \hat{f}}(Z^n; \theta, \hat{g})}, \]
along with the additional assumption

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

The following ULAN property then follows as a particular case of Proposition 3 in Hallin and La Vecchia (2017) or, with some minor variations (our parametrization is slightly different), from equation (7) in Drost and Werker (2004).

**Proposition 1 (ULAN).** Let Assumptions (A)–(C) hold. For all $g \in \mathcal{G}$, the parametric model $\mathcal{P}^{(n)}_g$ is ULAN with central sequence $\Delta^{(n)}(\theta, g)$ and information matrix $\Gamma(\theta, g)$. More precisely, we have, for all $g \in \mathcal{G}$, all $\theta \in \Theta$, all $\theta^{(n)}$ such that $\theta^{(n)} - \theta = O(n^{-1/2})$, and all bounded sequence $\tau_n \in \mathbb{R}^p$, under $\mathcal{P}^{(n)}_{\theta^{(n)}, g}$ as $n \to \infty$,
\[ \Lambda_n := \frac{d\mathbb{P}^{(n)}_{\theta^{(n)}, g}}{d\mathbb{P}^{(n)}_{\theta^{(n)}, g}} = \tau_n \Delta^{(n)}(\theta^{(n)}, g) - \frac{1}{2} \tau_n \Gamma(\theta^{(n)}, g) \tau_n + o_P(1), \tag{7} \]
and $\Delta^{(n)}(\theta^{(n)}, g) \overset{\mathcal{L}}{\to} \mathcal{N}(0; \Gamma(\theta^{(n)}, g))$.

A constant difficulty with ACD models, however, is that no explicit form of $\Gamma(\theta, g)$ is available—not even for exponential $g$; that difficulty, as we shall see, extends to the cross-information matrices used in R-estimation, to be replaced with consistent estimators. Numerical values easily can be computed, though, either as estimators or via simulations. To be precise, denoting by $\varepsilon_1, \ldots, \varepsilon_N$ an i.i.d. sequence generated from density $g$, since $\psi_g(\varepsilon_i)$ is centered (under $g$) and independent of $\hat{\psi}_{g^{(n)}}(\varepsilon_i; \theta)$, $\mathcal{D}^{(n)}(\theta, g, \varepsilon_i)$ is a consistent estimator, as $N \to \infty$, of $\Gamma(\theta, g)$.

4.2. Ranks and rank-based central sequences

4.2.1. Theoretical justification

In the classical semiparametric approach, the semiparametrically efficient central sequence is the tool that one needs to conduct inference and reach semiparametric efficiency bounds. That semiparametrically efficient central sequence, denoted as $\Delta^{(n)}(\theta, g)$, is obtained by projecting the central sequence $\Delta^{(n)}(\theta, g)$ along the so-called tangent space; see Drost and Werker (2004) for the case of ACD models.

Typically, the actual computation of semiparametrically efficient central sequences for time-series models is a painful case-by-case task which proves to be quite challenging and often discourages the application of the semiparametric approach. Moreover, once $\Delta^{(n)}(\theta, f)$ has been obtained for arbitrary $f \in \mathcal{F}$, the actual density $g$ still has to be estimated by some adequate $g^{(n)}$, to be plugged into $\Delta^{(n)}(\theta, f)$, yielding $\Delta^{(n)}(\theta, g^{(n)})$. Indeed, the asymptotic distribution of $\Delta^{(n)}(\theta, g)$ under $\mathcal{P}_{\theta, g}$ is unknown whenever $f \neq g$, and, typically, $\mathbb{E}_{g, \theta}[\Delta^{(n)}(\theta, f)] \neq 0$. As a consequence, the estimators based on $\Delta^{(n)}(\theta, f)$ reach the semiparametric efficiency bounds under $\mathcal{P}_{\theta, g}$, but are no longer root-$n$ consistent under $\mathcal{P}_{\theta, f}$ for $f \neq g$. Finally, further precautions, such as sample splitting, in principle are required in order for $\Delta^{(n)}(\theta, g^{(n)})$ to be asymptotically equivalent, under $\mathcal{P}_{\theta, g}$, to $\Delta^{(n)}(\theta, g)$.

It has been shown in Hallin and Werker (2003) that, for a very broad class of models (including most time series models) where the density $g$ of some underlying white noise remains completely unspecified, the maximal-invariance properties of residual ranks offer an attractive way to achieve semiparametric efficiency at $f$ without recurring to kernel estimation nor sample splitting. More precisely, projecting $\Delta^{(n)}(\theta, f)$ onto the $\sigma$-field generated by the ranks of the residuals $Z_{t+1}(\theta), \ldots, Z_n(\theta)$ yields a rank-based, hence distribution-free, version $\mathcal{D}^{(n)}(\theta, f)$, say, of the semiparametrically efficient central sequence $\Delta^{(n)}(\theta, f)$. Namely, under a very mild condition on score functions (Assumption (D)), $\mathcal{D}^{(n)}(\theta, f) - \Delta^{(n)}(\theta, f) = o_P(1)$ under $\mathcal{P}_{\theta, f}$. Estimators (and the tests) based on $\mathcal{D}^{(n)}(\theta, f)$ thus reach the semiparametric efficiency bounds under $\mathcal{P}_{\theta, f}$ (that is, when $g = f$); however, the distribution-freeness of ranks ensures that the same estimators remain root-$n$ consistent, and the tests keep their nominal asymptotic size, under $\mathcal{P}_{\theta, f}$ for any $g \in \mathcal{G}$. Thus, they remain valid (i.e., rate-optimal) irrespective of the actual density $g$, the estimation of which therefore is not required. Hallin and La Vecchia (2017) show that the above arguments can be adapted to a fairly general class of dynamic location-scale models that includes the ACD model (2).
4.2.2. Computation of rank-based central sequences for ACD models

Let us provide some details on the computation of the rank-based central sequences $\Delta^{(n)}(\theta, f)$ just described. Let $f \in G$ be some reference density. Typical candidates are the Gamma, Weibull, Generalized-$F$, or Burr densities. Denote by $R^{(n)}(\theta)$ the vector $(R_1^{(n)}(\theta), \ldots, R_n^{(n)}(\theta))$ of residual ranks, where $R_i^{(n)}(\theta)$ is the rank of $z_i^{(n)}(\theta)$ among $z_1^{(n)}(\theta), \ldots, z_n^{(n)}(\theta)$. For notational convenience, we also write $R^{(n)}, R_i^{(n)}$ and $Z^{(n)}$, or $R(\theta), R_i(\theta)$ and $Z(\theta)$, dropping the dependence on $\theta$ and/or $n$ when no confusion is possible.

To perform the rank-based construction, simply consider the central sequence $\Delta^{(n)}(\theta, f)$ in (5), which (for given starting values) is entirely measurable with respect to the residuals $Z_i(\theta)$, and (i) replace those residuals $Z_i(\theta)$ with $F^{-1}(R_i(\theta)/(n+1))$; (ii) re-center$^5$ the resulting expression about its expected value (this centering, for ACD models, can be dispensed with, as it leads to a negligible $o_p(1)$ correction). This yields a vector of rank-based statistics which, being measurable with respect to the ranks, is distribution-free under parameter value $\theta$. Provided that the score function $\psi_f$ associated with the reference density $f$ satisfies the very mild Assumption (D),$^7$ the general results of Hallin and La Vecchia (2017) imply the asymptotic equivalence (under parameter value $\theta$) of that vector (an approximate-score rank statistic) with its exact-score counterpart $\Delta^{(n)}(\theta, f)$; we thus safely use the same notation for both.

**Assumption (D).** The mapping (from $\mathbb{R}^+$ to $\mathbb{R}$) $z \mapsto \psi_f(z)$ is monotone, or a linear combination of monotone mappings.

4.2.3. Illustration

To fix the ideas, let us illustrate the above construction of $\Delta^{(n)}(\theta, f)$ for the widely-applied ACD(1,1) model; the special case ACD(1,0) follows in a straightforward way. As in Engle and Russell (1998), we consider

$$Y_i = \epsilon_i \Psi_{i-1}, \quad \text{with} \quad \Psi_{i-1} = \alpha + \beta Y_{i-1} + \gamma \Psi_{i-2},$$

where the shocks $\epsilon_i$ are i.i.d. with density $g \in G$. Initial values have negligible asymptotic influence on the central sequence, hence can be put to zero (a more reasonable choice is the unbiased estimator of $Y_i$’s stationary unconditional expectation). It then follows from Proposition 1 that

$$\Delta^{(n)}(\theta, f) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\psi_f(Z_i(\theta))}{\alpha + \beta Y_{i-1} + \gamma \Psi_{i-2}} \Psi_{i-1}(Y_{i-1}, Y_{i-2}, \ldots, Y_1; \theta)$$

(10)

where $Z_i(\theta) = Y_i/\Psi_{i-1}(\theta)$, $\Psi_{i}$ is as in (3) with, from the second part of (9),

$$\Psi_i = \alpha + \sum_{k=1}^{i-1} \gamma^k + \beta \sum_{k=0}^{i-1} \gamma^k Y_{i-k} + \gamma^i \Psi_0.$$

Here again, the asymptotic influence of the starting value $\Psi_0$ is negligible, and for simplicity can be put to one or to $\Psi_1$’s stationary unconditional expectation as estimated by $Y$. The semiparametrically efficient central sequence$^6$ is available in Drost and Werker (2004), and reads as

$$\Delta^{(n)}(\theta, g) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\{Z_i(\theta) - 1\} \Psi_{i-1}(Y_{i-1}, Y_{i-2}, \ldots, Y_1; \theta)}{\alpha + \beta Y_{i-1} + \gamma \Psi_{i-2}} \Psi_{i-1}(Y_{i-1}, Y_{i-2}, \ldots, Y_1; \theta)$$

(11)

with $c_g(\theta) := E_g(\Psi_{i-1}/\Psi_{i-1})$. We remark that $\Delta^{(n)}(\theta, g)$ involves the unknown actual innovation density $g$, its derivative $g'$, $E_g(\epsilon_i^2)$ and $E_g(\Psi_{i-1}/\Psi_{i-1})$; estimating all these quantities would require kernel density estimations, entailing sample splitting and bandwidth selections.

Turning $\Delta^{(n)}(\theta, f)$ into a rank-based central sequence first requires expressing it in terms of the residuals $Z_i(\theta)$. Denote by $R_1^{(n)}(\theta), \ldots, R_n^{(n)}(\theta)$ the ranks of $Z_1^{(n)}(\theta), \ldots, Z_n^{(n)}(\theta)$. Letting $\Psi_0 = \Psi_0$, start, for $i = 1, \ldots, n$, the recurrence

$$X_i = \Psi_{i-1} F^{-1}(R_i^{(n)}(\theta)/(n+1)), \quad \Psi_i = \alpha + \beta X_i + \gamma \Psi_{i-1}.$$

---

5 As remarked by an anonymous referee, the transformation applied to get the new residuals is a double probability transform: first we use the empirical CDF and then we use a specified inverse CDF. This type of transformation is very common in financial time series analysis, often for diagnostic checking.

6 The $\sigma$-field generated by the residual ranks indeed also contains the constants; note that the recentering term does not depend on the unspecified actual $g$. Moreover, it follows from Section 5.1(c) of Hallin and La Vecchia (2017) that such recentering, being $o(n^{-1/2})$, can be omitted.

7 Assumption (D) actually is needed in establishing the asymptotic representations underlying all asymptotic results about serial rank statistics; essentially, it requires $\psi_f$ to have bounded variation.

8 The semiparametrically efficient score is derived through the tangent space argument, where the projection is performed onto the space generated by all possible score functions associated with $g$, such that each score function has zero expectation and $E_g(\epsilon_i^2) = 1$; see equations (8)–(10) in Drost and Werker (2004). Note that no restriction on $E_g(\epsilon_i^2)$ is imposed.
This yields, for reference density \( f \), the rank-based central sequence
\[
\Delta^{(n)}(\theta, f) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{i} \left( F^{-1}(\hat{R}^{(n)}_{i}(\theta)/(n+1)) \right)
\]
\[
\alpha + \beta \chi_{1} - \gamma \chi_{2} - \ldots - \chi_{1} \; ; \; \theta
\]
(12)
satisfying \( \Delta^{(n)}(\theta, f) = \Delta^{(n)}(\theta, f) + o_p(1) \) under \( P_{\theta, f}^{(n)} \).

The exact covariance matrix \( \Gamma^{(n)}(\theta, f) \) of \( \Delta^{(n)}(\theta, f) \) in principle can be computed as an average over the \( n! \) possible values of the vector of residual ranks. Due again to the complicated nonlinear form of ACD likelihoods, sampling those \( n! \) possible values seems preferable, and is performed on the basis of the recursive construction just described.

The rank-based central sequence \( \Delta^{(n)}(\theta, f) \) in (12) requires the specification of a reference density \( f \). In this paper we propose the use of an exponential with rate one, viz. \( f = f_{\exp} \). Thanks to this choice, the rank-based central sequence is particularly easy to compute and easy to implement.\(^9\)

Clearly, other choices for \( f \) are possible. For instance, one can use standardized Weibull or Burr densities. Unlike Weibull- or Burr-based QMLEs, and thanks to the distribution-freeness of the residual ranks, Weibull- and Burr-based R-estimators preserve root-\( n \) consistency even under misspecified \( g \); their semiparametric efficiency, however, only holds under correctly specified (Weibull or Burr) densities. Moreover, a data-driven choice \( f^{(n)} \) of the reference density \( f \) (on the model of Dodge and Jurečková, 2000 or Hallin and Mehta, 2015) is also possible. For instance, one can use for \( f \) a Weibull density \( f^{(n)} = f_n^{\alpha} \) with estimated shape parameter \( \hat{\alpha}_{n}^{*} \) in order to match some characteristics (e.g. the skewness) of the data, and then adjust the scale parameter to preserve an expected value of one; no sample splitting is needed, provided that \( \hat{\alpha}_{n}^{*} \) only involves the order statistic of residuals.\(^10\)

A relevant feature of this approach is that, contrary to the classical semiparametric estimators à la Bickel et al. root-\( n \) consistency of the R-estimator implied by the data-driven scores does not require \( f^{(n)} \) to converge to \( g \). Alternatively, if semiparametric efficiency is to be attained at the actual \( g \), choosing as a data-driven reference density \( f^{(n)} \) a kernel estimator of \( g \) is an option as well—which does not require sample splitting, since the kernel estimator only depends on the order statistic of residuals.

5. R-estimation

5.1. Method

We now turn to the specific problem of R-estimating \( \theta \). The basic idea consists in treating the rank-based central sequence \( \Delta^{(n)}(\theta, f) \) as we would, under density \( f \), treat the central sequence \( \Delta^{(n)}(\theta, f) \). The huge advantage of \( \Delta^{(n)}(\theta, f) \) over \( \Delta^{(n)}(\theta, f) \) is that, due to distribution-freeness, its expectation under any \( g \) (and parameter value \( \theta \)) is zero—the Fisher consistency condition that \( \Delta^{(n)}(\theta, f) \) and \( \Delta^{(n)}(\theta, f) \) typically only enjoy under \( g = f \). As a consequence, the resulting estimators (called R-estimators) remain root-\( n \) consistent irrespective of the actual density \( g \)—a property that does not hold for the estimator based on \( \Delta^{(n)}(\theta, f) \) or \( \Delta^{(n)}(\theta, f) \) unless \( f \) is exponential or Gamma.

The technique, described in Hallin and La Vecchia (2017), basically consists in subjecting \( \Delta^{(n)}(\theta, f) \) to Le Cam’s classical one-step method. It requires a preliminary, root-\( n \) consistent (under any \( P_{\theta, g}^{(n)} \), \( g \in G \) ), estimator \( \hat{\theta}_{n}^{(n)} \) of \( \theta \), along with a consistent (under any \( P_{\theta, g}^{(n)} \), \( g \in G \) ) estimator \( \hat{\Gamma}_{n}^{(n)} \) of the cross-information matrix
\[
\Gamma(\theta, f, g) := \lim \mathbb{E}_{\theta, g} \left[ \Delta^{(n)}(\theta, f) \Delta^{(n)}(\theta, g) \right].
\]
(13)
That matrix is the one determining, via Le Cam’s Third Lemma, the asymptotic behavior under local alternatives of the form \( P_{\theta+n^{-1/2}r, g}^{(n)} \) of \( \Delta^{(n)}(\theta, f, g) \), as well as its asymptotic linearity (see below).

Constructing such \( \hat{\Gamma}_{n}^{(n)} \) is a delicate task, since \( \Gamma(\theta, f, g) \) involves the expectation, under the actual density \( g \), which is unknown, of quantities that themselves depend on \( g \) and \( f \). The general method described in Hallin and La Vecchia (2017) does not apply here, as the factorization of \( \Gamma(\theta, f, g) \) on which it relies does not hold. As an alternative, we propose
\[
n^{-1/2} \hat{\Gamma}_{n}^{(n)} := - \left( \Delta^{(n)}(\hat{\theta}_{n}^{(n)}) + n^{-1/2} \epsilon_{1}, f, \ldots, \Delta^{(n)}(\hat{\theta}_{n}^{(n)}) + n^{-1/2} \epsilon_{p+q+1}, f \right) + \left( \Delta^{(n)}(\hat{\theta}_{n}^{(n)}), f, \ldots, \Delta^{(n)}(\hat{\theta}_{n}^{(n)}, f) \right)
\]
(14)
where \( \epsilon_{k}, k = 1, \ldots, p+q \), stands for the \( k \)th vector of the canonical basis in \( \mathbb{R}^{p+q} \). Proposition 2 establishes under which assumptions \( \hat{\Gamma}_{n}^{(n)} \) indeed is an adequate estimator of \( \Gamma(\theta, f, g) \).

The one-step (see Section 5.2) R-estimator of \( \theta \) based on reference density \( f \) then is defined as
\[
\hat{\theta}_{n}^{(n)} := \hat{\theta}_{n}^{(n)} + n^{-1/2} \left( \hat{\Gamma}_{n}^{(n)} \right)^{-1} \Delta^{(n)}(\hat{\theta}_{n}^{(n)}, f).
\]
(15)

\(^9\) Indeed, \( \psi_{i}(F^{-1}(\hat{R}^{(n)}_{i}(\theta)/(n+1))) = F_{\exp}^{-1}(F_{\exp}(\hat{R}^{(n)}_{i}(\theta)/(n+1))) - 1 \), where for \( u \in [0, 1] \) \( F_{\exp}^{-1}(u) = -\ln(1 - u) \) is the inverse of the exponential (with rate one) distribution function.

\(^10\) The ranks and the order statistic are mutually independent, and this implies that order statistic-driven scores can be treated as if they were deterministic.
Before characterizing the asymptotic behavior of \( \hat{\theta}^{(n)} \), let us summarize the assumptions to be made.

**Assumption (F).** \( (F1) \) \( \hat{\theta}^{(n)} \) is a root-\( n \) consistent and asymptotically discrete estimator of \( \theta \)—namely, (i) for any \( g \in \mathcal{G} \) and \( \theta \in \Theta \), under \( \mathbb{P}_{\theta,g}^{(n)} \), as \( n \to \infty \), \( n^{1/2} (\hat{\theta}^{(n)} - \theta) \) is \( O_p(1) \), and (ii) for any \( \varepsilon > 0 \), \( \hat{\theta}^{(n)} \) only takes a finite number of values in balls of radius \( cn^{-1/2} \) centered at \( \theta \);

\( (F2) \) \( g \) is such that (asymptotic linearity), for all \( \theta \in \Theta \) and \( \tau \in \mathbb{R}^{p+q} \),

\[
\Delta^{(n)}(\theta + n^{-1/2} \tau, f) - \Delta^{(n)}(\theta, f) = -n^1 \text{adj}(\theta, f, g) \tau + O_p(1)
\]

under \( \mathbb{P}_{\theta,g}^{(n)} \), as \( n \to \infty \).

We then have, for the one-step R-estimator (15), the following result.

**Proposition 2.** Let Assumptions (A)--(F) hold. Then, under \( \mathbb{P}_{\theta,g}^{(n)} \), as \( n \to \infty \),

\[
n^{1/2} (\hat{\theta}^{(n)} - \theta) \xrightarrow{D} \mathcal{N}\left( 0, \mathbb{E}^{-1}(\theta, f, g) \right).
\]

reducing, under \( \mathbb{P}_{\theta,f}^{(n)} \), to

\[
n^{1/2} (\hat{\theta}^{(n)} - \theta) \xrightarrow{D} \mathcal{N}\left( 0, \mathbb{E}^{-1}(\theta, f) \right).
\]

**Proof.** The proof straightforwardly follows from Proposition 4.2 in Hallin and La Vecchia (2017) (recall that \( \mathbb{E}(\theta, f) \)) coincides with the semiparametric information matrix \( \Gamma^0(\theta, f) \)) provided that \( \mathbb{P}_{\theta,f}^{(n)} \) consistently estimates \( \Gamma^0(\theta, f, g) \). This, however, is a consequence of the asymptotic linearity assumption (F2) for \( \tau = e_1, \ldots, e_{p+q} \), along with the fact (F1) that \( \hat{\theta}^{(n)} \) is locally asymptotically discrete. Indeed, for any \( \varepsilon > 0 \), \( c > 0 \) and \( n \) sufficiently large, the probability that \( \hat{\theta}^{(n)} \) lies outside a ball of radius \( cn^{-1/2} \) centered at \( \theta \) is smaller than \( \varepsilon/2 \), say. Inside the ball, \( \hat{\theta}^{(n)} \) only takes a finite number of values. At each of those (deterministic) values, (16) holds for \( \tau = e_k, k = 1, \ldots, p+q \), with an \( O_p(1) \) term the maximum of which (since there is only a finite number of them) is, for \( n \) large enough, less than \( \varepsilon \) with probability at least \( \varepsilon/2 \). Consistency of \( \hat{\theta}^{(n)} \) follows. As in Hallin and La Vecchia (2017), asymptotic linearity and local discreteness of \( \hat{\theta}^{(n)} \) entail, under \( \mathbb{P}_{\theta,f}^{(n)} \), the asymptotic equivalence

\[
n^{1/2} (\hat{\theta}^{(n)} - \theta) = \mathbb{E}^{-1}(\theta, f, g) \Delta(\theta, f) + O_p(1),
\]

of which (17) is an immediate consequence. \( \square \)

Proposition 2 and (17) imply that \( \hat{\theta}^{(n)} \) remains root-\( n \) consistent and asymptotically normal for any \( (f, g) \in \mathcal{G} \) such that Assumption (F2) holds, while (18) entails semiparametric efficiency under \( g = f \); contrary to many M- and L-estimators, and unlike non-exponential quasi-likelihood estimators, thus, our R-estimators are robust to innovation density misspecification.

**5.2. Comments**

The rank-based estimation procedure just described calls for some comments.

(i) As a rule, preliminary estimators should not aim at efficiency, and should focus on robustness: the only purpose of \( \hat{\theta}^{(n)} \) indeed is to restrict subsequent estimation to a root-\( n \) neighborhood of the actual parameter value \( \theta \). A natural candidate here is the QMLE, the consistency of which (under any \( \mathbb{P}_{\theta,g}^{(n)} \)) has been established by Drost and Werker (2004), or “robustified” versions thereof. (ii) Assumption (F1) requires \( \hat{\theta}^{(n)} \) to be locally asymptotically discrete. This is easily achieved by discretizing any root-\( n \) consistent estimator \( \hat{\theta}^{(n)} \) along a grid of mesh \( c_0 n^{-1/2} \), \( c_0 > 0 \) arbitrarily small but fixed. 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 discretizing, as \( \hat{\theta}^{(n)} \) in practice only involves a finite number of digits, and \( n \) does not go to infinity; see pages 125 and 188 of Le Cam and Yang (1990). (iii) Note that, for \( g = f \), \( \Gamma(\theta, f) \) coincides with \( \Gamma^0(\theta, f) = \Gamma^0(\theta, f, f) \). (iv) For exponential reference density \( f = f_{\exp} \), the limit \( \mathbb{E}(\theta, f_{\exp}) \) of \( \theta \) and the covariance matrix under \( f_{\exp} \) of the exponential QMLE coincide. This implies that \( \Gamma(\theta, f_{\exp}) \) can be obtained as explained at the end of Section 4.1. With this regard, for data-driven \( f \), we emphasize that the definition of R-estimators using \( f \neq f_{\exp} \) requires additional numerical efforts to estimate \( \Gamma(\theta, f^{(n)}) \), a task which increases the complexity of the algorithm performing R-estimation. See Section 7.2 for a related discussion and numerical details. (v) The one-step method (15) can be considered as a single iteration, with starting value \( \hat{\theta}^{(n)} \), of the Newton–Raphson numerical solution of

\[
\hat{\theta}_{\text{NL1}}^{(n)} := \arg\min_{\theta} \Delta^{(n)}(\theta, f) f^{-1}(\theta, f, f) \Delta^{(n)}(\theta, f).
\]

---

Please cite this article as: M. Hallin and D. La Vecchia, A Simple R-estimation method for semiparametric duration models. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.04.036.
which is the natural generalization to the present setting of the Hodges–Lehmann definition of R-estimators (Hodges and Lehmann, 1956) is restricted to univariate location parameters. The idea is to improve on the initial estimator \( \hat{\theta}^{(i)} \), which is already in a \( n^{-1/2} \)-neighborhood of the true parameter value \( \theta \); see the proof of Proposition 2. A single iteration of the Newton–Raphson scheme produces an estimator that is asymptotically equivalent to the estimator defined by (19)—root-\( n \) consistency is guaranteed by the root-\( n \) consistency of the initial value. But of course, in practice, several iterations can be performed, and are likely tp improve finite-\( n \) performance without modifying asymptotic behavior. We refer to Drost and Werker (2004) for a discussion on the one-step method in the ACD semiparametric modeling.

6. Simulation study

In this section, we illustrate the actual performances of our R-estimators. For the sake of simplicity, we work on the ACD(1,0) model as in the motivating example. We concentrate on the Burr(\( \varphi_1 = 1 \), \( \varphi_2 = 2 \)) case, and refer to Section 2 for the case of Weibull innovations with \( \zeta \geq 1 \).

We simulate the ACD process (1), for parameter value \( \beta = 0.25 \), Burr(1,2) innovation density, and four different sample sizes: \( n = 250, 500, 750, \) and 1000; the Burr(1,2) density generates asymmetric and leptokurtic innovations.

The problem here is to estimate \( \beta \). The following four estimators were computed and compared: (a) the usual QMLE, based on exponential quasi-likelihood; (b) the QMLE based on Weibull quasi-likelihood, as suggested by Engle and Russell (1998); (c) the parametric MLE based on the actual Burr likelihood; (d) the R-estimator based on exponential reference density \( f \). Out of those four estimators, only (a), (c), and (d) are root-\( n \) consistent under Burr innovations; (b) is not— the Weibull distribution does not belong to the exponential family and Weibull QMLEs therefore are not consistent M-estimators, see Gouriéroux et al. (1984) or Meitz and Teräsvirta (2006); neither (b) nor (c) qualify as semiparametric estimators. The Weibull QMLE nevertheless belongs to daily practice and, in some settings (such as the Monte Carlo design of this simulation), its bias can be relatively small.

Estimators (a)–(c) were obtained using the R routine \texttt{acdfit} (as available in the R package \texttt{ACDm}), while the R-estimates are easily derived from the recurrence (11) followed by iterating the one-step update (15) until numerical stabilization.

Results are displayed in Fig. 2. They reveal that the R-estimator outperforms both QMLEs, for all sample sizes considered. When \( n = 1000 \) (“large” sample), the boxplot of the (infeasible) Burr MLE and the boxplot of the R-estimator are pretty similar (even though the R-estimator is based on a misspecified reference density), whereas the QMLEs exhibit higher variability. When \( n = 250 \), the R-estimator still outperforms both QMLEs; inspection of the top left panel in Fig. 2 indicates that (d) has smallest bias and smallest interquartile range when compared to (a) and (b). We conjecture that this is due to the robustness/stability features of rank-based procedures. In contrast, it is well known that QMLEs (a) and (b) are, typically, much more sensitive to small-sample effects and/or observed extreme values: the combination of these issues is most likely the cause of the large number of outliers visible in the boxplot of the exponential QMLE in Fig. 2, for \( n = 250 \) and \( n = 500 \)—their impact decreases as the sample size increases, though.\(^{11}\)

7. Real data

As highlighted by Tsay (2009) and Yatigammana et al. (2016), ACD models can be fitted to any time series with non-negative observations. In this section, we illustrate the use of our R-estimators in the context of stock volatility modeling based on the daily range (see Chou, 2005 and Martens and Van Dijk, 2007), and consider the time series of daily ranges for the Swiss stock market index (SMI), over the period 02-Jan-2015 to 30-Dec-2015.\(^{12}\)

7.1. Data description, modeling and estimation methods

Data. Let \( P_i \) denote the open price of SMI at day \( i \), which we assume to be driven by a geometric Brownian motion with stochastic volatility; see Martens and Van Dijk (2007) for details. Fig. 3 (left panel) provides a plot in log scale of that series.

SMI prices are observed at high frequency—every 5 min, throughout the trading day, from 9:00 AM to 5:30 PM (Zürich time). Denoting by \( P_{id} \) the SMI value at discrete intra-day time \( d \) of day \( i \), define the observed range \( Y_i := 100 \cdot \left[ \ln(\max_{d} P_{id}) - \ln(\min_{d} P_{id}) \right] \).

Fig. 3 displays the series of observed ranges of the SMI index, along with its histogram and its autocovariance function. The plots indicate the presence of extreme values, high asymmetry in the distribution of the range and strong autocovariance. Indeed, at the beginning of the period of observation, the Swiss National Bank suddenly announced that it would no longer hold the Swiss Franc at the fixed exchange rate (as introduced in 2011) with the Euro. Then, stocks in Switzerland experienced their worst fall in 25 years—a shock that triggered wild swings in equity and currency markets.

\(^{11}\) Unreported results suggest that the R-estimator based on the exponential density yields some efficiency gains also for Weibull innovations with \( \zeta < 1 \). However, for these distributions the numerical stability of the QMLE seems to deteriorate, specially in samples with \( n < 250 \). Plots for a Weibull with \( \zeta = 0.6 \) are available upon request.

\(^{12}\) The SMI data, the values of \( \max_{d} P_{id} \) and \( \min_{d} P_{id} \) for every day, are freely available at Yahoo finance website.
Fig. 2. Boxplots for different estimators of $\beta$ in an ACD(1,0) model with $\epsilon_i \sim \text{Burr}(1,2)$ and sample sizes $n = 250, 500, 750,$ and $1000$ (clockwise). The following estimators are considered in each panel: (a) Exponential QMLE; (b) Weibull QMLE; (c) Burr MLE; (d) R-estimator with exponential reference density; the horizontal line corresponds to the true value $\beta = 0.25$. In each panel the Monte Carlo size is 1000.

**Modeling and estimation methods.** We model the range dynamics using the ACD(1,1) model as in (9) and we compare different estimates of the model parameter $\theta := (\alpha, \beta, \gamma)$. In particular, we compare (i) the exponential QMLE $\hat{\theta}_{\text{QMLE}}^{(n)}$; (ii) our R-estimator based on the exponential (with mean one) reference density $\hat{\theta}_{\text{exp}}^{(n)}$; (iii) an R-estimator based on data-driven Weibull scores $\hat{\theta}_{\text{W}}^{(n)}$ (with scale adjusted to have mean one). We conjecture that the exponential QMLE is likely to suffer from a severe efficiency loss, because of extreme observations, while our R-estimation can yield better accuracy and robustness properties.

Some numerical details. The $\hat{\theta}_{\text{QMLE}}^{(n)}$ is obtained using the R routine acdFit and the Nelder–Mead optimization algorithm; we estimate its asymptotic variance matrix using the sandwich formula in Engle (2000). The rank-based central sequence $\mathbf{A}^{(n)}(\theta, f_{\text{exp}})$ is as in (12), while the R-estimator is obtained using the one-step method as in (15)—implementing the recursion in (11) with $\psi_0 = 1$ and $Y_0$ equal to the sample mean $\bar{Y}$, and the $\hat{\theta}_{\text{QMLE}}^{(n)}$ as a preliminary estimator. The estimated cross-information matrix $\mathbf{C}^{(n)}(\theta, f_{\text{exp}})$ is obtained as in (14). The asymptotic variance matrix of the R-estimator as in (17) requires the computation of $\mathbf{C}(\theta, f_{\text{exp}})$; this can be obtained either using the permutations of the residual ranks (see the last paragraph of Section 4.2.2), or via simulation. Indeed, $\mathbf{C}(\theta, f_{\text{exp}})$ coincides with the covariance matrix of the QMLE, thus one can estimate it by simulating an ACD(1,1) with innovation density $f_{\text{exp}}$. The latter method was adopted since it is computationally much lighter and can fully exploit the availability of stable routines available in R package ACDm. As far as the data-driven R-estimator is concerned, we estimate the Weibull shape parameter using the order statistic of the residuals. Then, we adjust the Weibull scale parameter to preserve expectation equal to one and we use the resulting $f_{\text{exp}}^{(n)}$ to define the rank-based central sequence $\mathbf{C}(\theta, f^{(n)})$. This preliminary step is supposed to provide a reference density $f^{(n)}$ which is closer than the exponential one to the actual density $g$. As in the case of R-estimator based on the exponential density, the estimated cross-information matrix is obtained as in (14). Finally, the asymptotic variance matrix of the
Fig. 3. Left panel: the range of the SMI over the period 02-Jan-2015 to 30-Dec-2015. Middle panel: autocorrelation function (and its 90% confidence bands, as dotted lines) of the range time series. Right panel: histogram of the range of the SMI, along with its median (read vertical line).

Table 1

| Parameter Improving QMLE | $\alpha$ | $\beta$ | $\gamma$ | $\delta$-trace $\times 10^2$ |
|---------------------------|---------|---------|----------|-----------------------------|
| $\hat{\theta}^{[\text{n}]}_{\text{QMLE}}$ | 0.229   | 0.225   | 0.581    | $\odot$                     |
|                           | (0.114) | (0.074) | (0.159)  |                             |
| $\hat{\theta}^{[\text{n}]}_{\text{exp}}$ | 0.238   | 0.271   | 0.553    | $\checkmark$                 |
|                           | (0.091) | (0.126) | (0.127)  | (0.336)                     |
| $\hat{\theta}^{[\text{n}]}_{\text{W}}$  | 0.248   | 0.341   | 0.533    | $\checkmark$                 |
|                           | (0.143) | (0.059) | (0.098)  | (0.993)                     |

R-estimator as in (17) requires the computation of $\Gamma(\theta, \hat{f}^{[\text{n}]}_{\text{n}})$ via a simulation procedure (slightly) more complicated than the one implemented in the case of $f = f_{\text{exp}}$, a task which increases (slightly) the computational burden of the algorithm.

7.2. Inference

Estimates and asymptotic variance matrices. The estimates (and their standard errors) are reported in Table 1. In the present context, the comparison of asymptotic variance (or information) matrices in the partial order of nonnegative definite matrices is difficult, since the asymptotic variance (information) matrices are based on different estimating functions. In particular we observe that the R-estimators yield a reduction in the standard errors for some estimates, but not for all of them. For instance, $\hat{\theta}^{[\text{n}]}_{\text{exp}}$ has smaller standard errors than $\hat{\theta}^{[\text{n}]}_{\text{QMLE}}$ for the estimation of $\alpha$ and $\gamma$, while $\hat{\theta}^{[\text{n}]}_{\text{W}}$ features better accuracy than $\hat{\theta}^{[\text{n}]}_{\text{QMLE}}$ in the estimation of $\beta$ and $\gamma$.

To shed light, we replace the matrix comparison by a simpler scalar comparison which gives us a way to rank the estimators under study; see e.g. Heyde (1997) for a book-length discussion. We here adopt the customary criterion based on the comparison of the traces of the asymptotic covariance matrices: the higher the trace, the lower the accuracy. Thus, given two root-$n$ consistent estimators of $\theta$, we compute the difference between the traces of their respective asymptotic covariance matrices, and say that the second estimator improves over the first one if the difference is positive. Using obvious notation, Table 1 reports those trace differences as “$\delta$-trace”. Under this trace criterion, both R-estimators improve on $\hat{\theta}^{[\text{n}]}_{\text{QMLE}}$ and $\hat{\theta}^{[\text{n}]}_{\text{W}}$ improves on $\hat{\theta}^{[\text{n}]}_{\text{exp}}$, with a $\delta$-trace value of $= 0.657$. Thus, if one is willing to accept a slightly more complex algorithm, the use of data-driven scores in R-estimation is recommended.

Diagnostics. We also computed the autocorrelation function of the standardized residuals (in log-scale) obtained from the estimators above. In Fig. 4, we display the results. The left-hand plot indicates that $\hat{\theta}^{[\text{n}]}_{\text{QMLE}}$ leaves some significant
Fig. 4. Diagnostics via residual autocorrelation function (ACF) of log-residuals, as implied by the QMLE (left-hand panel), the $f_{lep}$ R-estimator (middle panel), and the data-driven R-estimator with Weibull reference density (right-hand panel). In each plot, the dotted lines provide the 90% ACF confidence bands.

(at 90% level) residual lag-one autocorrelation; the two R-estimators do not suffer from the same criticism. R-estimation-based residual analysis (middle and right-hand plots) conclude that the ACD$(1,1)$ model is appropriate, but QMLE-based residual analysis (left-hand plot) does not.

8. Conclusion

In this paper we develop a novel semiparametric R-estimation method for duration models. Numerical exercises illustrate the advantages of our procedure, both in simulations and in real-life data. Challenging extensions can be considered to more sophisticated models. Among them, the class of semiparametric augmented duration models introduced in Fernandes and Grammig (2006) or the class of autoregressive conditional intensity models (see Hautsch, 2011 for a book-length introduction), both representing general univariate multiplicative error models, for which no ULAN expansion and/or semiparametric estimation methods à la Bickel et al. are available. An even more challenging research topic is the derivation, in the spirit of this paper, of an R-estimation method for the class of multivariate multiplicative error models. The main problem is related to the lack of a suitable definition of ranks for multivariate time series. A possible way to tackle this problem is to make use of the recent results in Chernozhukov et al. (2017), Hallin (2017) and Hallin et al. (2020), where new concepts of ranks and signs are developed in a multivariate context. These tools are used by Hallin et al. (2019) in the construction of R-estimators for VARMA models. Potential financial applications of this methodology are the estimation of multivariate dynamic conditional correlation structures, as introduced by Engle (2002a) and widely applied in the analysis of time-varying correlations of financial market indices or the modeling of daily correlation between returns on several currencies. We are currently working on this research direction.

Acknowledgments

We thank Patrick Gagliardini, Elvezio Ronchetti, Olivier Scaillet, Mervyn Silvapulle, and Bas Werker for helpful discussions about an earlier version of this manuscript. We also thank the Editor, one Co-Editor and two anonymous Referees for their careful reading of the manuscript and helpful remarks. Davide La Vecchia acknowledges the Swiss National Science Foundation, Switzerland 100018_16955 for financial support. Part of this work was completed as Marc Hallin was visiting the Statistics Department at the Universidad Carlos 3 de Madrid on a grant funded by the Banco Santander, Spain.

References

Aït-Sahalia, Y., Jacod, J., 2014. High-Frequency Financial Econometrics. Princeton University Press.
Andrews, E., Werker, B., 2015. Residual-based rank specification tests for AR-GARCH type models. J. Econometrics 185, 305–331.
Andrews, B., 2012. Rank-based estimation for GARCH processes. Econometric Theory 28, 1037–1064.
Benghabrit, Y., Hallin, M., 1992. Optimal rank-based tests against first-order superdiagonal bilinear dependence. J. Statist. Plann. Inference 32, 45–61.
Benghabrit, Y., Hallin, M., 1996. Rank-based tests for autoregressive against bilinear serial dependence. J. Nonparametr. Stat. 6, 147–170.
Bickel, P., Klaassen, C., Ritov, Y., Wellner, J., 1993. Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins University Press, Baltimore.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econometrics 31, 307–327.
Chernozhukov, V., Galichon, A., Hallin, M., Henry, M., 2017. Monge–Kantorovich depth, quantiles, ranks and signs. Ann. Statist. 45, 223–256.
Chou, R.Y., 2005. Forecasting financial volatilities with extreme values: the conditional autoregressive range (CARR) model. J. Money Credit Bank. 37, 561–582.
Dodge, Y., Jurečková, J., 2000. Adaptive Regression. Springer, New York.
Drost, F., Klaassen, C., Werker, B., 1997. Adaptive estimation in time-series models. Ann. Statist. 25, 786–817.
Drost, F.C., Werker, B.J., 2004. Semiparametric duration models. J. Bus. Econom. Statist. 22, 40–50.

Dufour, J.-M., 1981. Rank tests for serial dependence. J. Time Series Anal. 2, 117–128.

Dufour, A., Engle, R.F., 2000. Time and the price impact of a trade. J. Finance 55, 2467–2498.

Dufour, J.-M., Hallin, M., 1987. Tests non paramétriques optimaux pour une autorégression d’ordre un. Ann. Econom. Stat. 6–7, 411–434.

Dufour, J.-M., Lepage, Y., Zidean, H., 1982. Nonparametric testing for time series: a bibliography. Canad. J. Statist. 10, 1–38.

Dufour, J.-M., Roy, R., 1985. Some robust exact results on sample autocorrelations and tests of randomness. J. Econometrics 29, 257–273.

Engle, R.F., 2000. The econometrics of ultra-high-frequency data. Econometrica 68, 1–22.

Engle, R., 2002a. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. J. Bus. Econom. Statist. 20, 339–350.

Engle, R., 2002b. New frontiers for ARCH models. J. Appl. Economometrics 17, 425–446.

Engle, R.F., Russell, J.R., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 68, 1127–1162.

Fernandes, M., Grammig, J., 2006. A family of autoregressive conditional duration models. J. Econometrics 130, 1–23.

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

Hallin, M., 2017. On distribution and quantile functions, ranks and signs in $\mathbb{R}^d$. Available at https://ideas.repec.org/p/eca/wpaper/2013-258262.html.

Hallin, M., del Barrio, E., Cuesta Albertos, J., Matran, S., 2020. On distribution and quantile functions, ranks, and signs in $\mathbb{R}^d$: a measure transportation approach. Available at https://arxiv.org/abs/1806.01238.

Hallin, M., Ingenbleek, J.-F., Puri, M., 1985. Linear serial rank tests for randomness against ARMA alternatives. Ann. Statist. 13, 1156–1181.

Hallin, M., Ingenbleek, J.-F., Puri, M., 1989. Asymptotically most powerful rank tests for multivariate randomness against serial dependence. J. Multivariate Anal. 30, 34–71.

Hallin, M., La Vecchia, D., 2017. Semiparametrically efficient R-estimation in dynamic location scale models. J. Econometrics 196, 233–247.

Hallin, M., La Vecchia, D., Liu, H., 2019. Center-outward R-estimation for semiparametric VARMA models. Available at https://arxiv.org/abs/1910.08442.

Hallin, M., Mehta, C., 2015. R-estimation for asymmetric independent component analysis. J. Amer. Statist. Assoc. 110, 218–232.

Hallin, M., Paindaveine, D., 2002. Optimal procedures based on interdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against ARMA dependence. Bernoulli 8, 787–815.

Hallin, M., Paindaveine, D., 2004a. Multivariate signed rank tests in vector autoregressive order identification. Statist. Sci. 19, 697–711.

Hallin, M., Paindaveine, D., 2004b. Rank-based optimal tests of the adequacy of an elliptic VARMA model. Ann. Statist. 32, 2642–2678.

Hallin, M., Paindaveine, D., 2005. Affine-invariant aligned rank tests for the multivariate general linear model with VARMA errors. J. Multivariate Anal. 93, 122–163.

Hallin, M., Puri, M., 1994. Aligned rank tests for linear models with autocorrelated error terms. J. Multivariate Anal. 50, 175–237.

Hallin, M., van den Akker, R., Werker, B., 2006. A class of simple semiparametrically efficient rank-based unit root tests. J. Econometrics 163, 200–214.

Hallin, M., van den Akker, R., Werker, B., 2011. Semiparametric error-correction models for cointegration with trends: pseudo-Gaussian and optimal rank-based tests of the cointegration rank. J. Econometrics 190, 46–61.

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

Hasbrouck, J., 2007. Empirical Market Microstructure: the Institutions, Economics, and Econometrics of Securities Trading. Oxford University Press.

Hautsch, N., 2011. Econometrics of Financial High-Frequency Data. Springer Science & Business Media.

Heyde, C.C., 1997. Quasi-Likelihood and its Applications. Springer-Verlag, New York.

Hodges, J., Lehmann, E.L., 1956. The efficiency of some nonparametric competitors of the t-test. Ann. Math. Stat. 27, 324–335.

Hoel, P.G., Saleh, A.K., 1999. R-estimation of the parameters of autoregressive AR(p) models. Ann. Statist. 21, 534–551.

Kunitomo, N., 1992. Improving the Parkinson method of estimating security price volatilities. J. Bus. 65, 295–302.

Le Cam, L., Yang, G., 1990. Asymptotics in Statistics. Springer-Verlag, New York.

Manganeli, S., 2005. Duration, volume and volatility impact of trades. J. Financial Mark. 8, 377–399.

Martens, M., Vande Vondel, D., 2007. Measuring volatility with the realized range. J. Econometrics 138, 181–207.

Meitz, M., Teräsvirta, T., 2006. Evaluating models of autoregressive conditional duration. J. Bus. Econom. Statist. 24, 104–124.

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

Pacurar, M., 2008. Autoregressive conditional duration models in finance: a survey of the theoretical and empirical literature. J. Econ. Rev. 22, 711–751.

Parkinson, M., 1980. The extreme value method for estimating the variance of the rate of return. J. Bus. 53, 61–65.

Ranasinghe, K., Silvapulle, M.J., 2011. Estimation under inequality constraints: Semiparametric estimation of conditional duration models. Econometric Rev. 30, 359–378.

Tsay, R., 2005. Analysis of Financial Time Series. John Wiley & Sons.

Tsai, R., 2009. Autoregressive conditional duration models. In: Palgrave Handbook of Econometrics. Springer, pp. 1004–1024.

Vaart, A., 1998. Asymptotic Statistics. Cambridge University Press, Cambridge.

Yatigammana, R.P., Choy, S., Chan, J., 2016. Autoregressive conditional duration model with an extended Weibull error distribution. In: Huynh, V.N., Kreinovich, V., Sriboonchitta, S. (Eds.), Causal Inference in Econometrics. Springer, pp. 83–107.