Efficient Estimation in Semiparametric Time Series: the ACD Model

FEIKE C. DROST and BAS J.M. WERKER *

Tilburg University

January 26, 2000

Abstract

In this paper we consider efficient estimation in semiparametric ACD models. We consider a suite of model specifications that impose less and less structure. We calculate the corresponding efficiency bounds, discuss the construction of efficient estimators in each case, and study the efficiency loss between the models. We provide a simulation study that shows the practical gain from using the proposed semiparametric procedures. We find that, although one does not gain as much as theory suggests, these semiparametric procedures definitely outperform more classical procedures. We apply the procedures to model semiparametrically durations observed on the Paris Bourse for the Alcatel stock in July and August 1996.

KEYWORDS: Adaptiveness, Durations, One-step improvement, Semiparametric efficiency.

*CentER, Department of Econometrics, and Finance group, Tilburg University, P.O.Box 90153, 5000 LE, Tilburg, The Netherlands. We thank Joann Jasiak for kindly providing the data.

1 Introduction

The last years, an enormous progress has been made in the area of semiparametric estimation from a theoretical point of view. Starting with the work of Stein (1956) about the possibility of adaptiveness in the symmetric location model, the techniques have been developed since. The work by Hájek and Le Cam is especially worth mentioning here. Traditionally, the models considered are based on i.i.d. observations. A fairly complete account on the state of the art in i.i.d. models can be found in the monograph by Bickel, Klaassen, Ritov, and Wellner (1993), henceforth BKRW (1993). Newey (1990) is an overview paper from an econometric perspective. Semiparametric efficiency considerations and adaptiveness in time series have been discussed as well, beginning with Kreiss (1987a) and Kreiss (1988b) for ARMA-type models. Koul and Schick (1997) discusses nonlinear autoregressive models with special emphasis on the initial value problem. Drost, Klaassen, and Werker (1997) considers so-called group models, covering nonlinear location-scale time series. Linton (1993) discusses linear models with ARCH errors. Drost and Klaassen (1997) particularizes to the GARCH model and Wefelmeyer (1996) calculates efficiency bounds in general Markovian type models.

The crucial ingredient in semiparametric efficiency calculations is the efficient score-function. Let us recall this concept here. For a rigorous treatment, one may consult, e.g., BKRW (1993). Consider a setup where i denotes the observation number and $\theta \in \Theta$ is a finite dimensional parameter of interest. We denote (conditional) expectations under θ by E_θ. In general, a score function $s_i(\cdot)$ is a sequence of random functions, such that

$$E_\theta \{s_i(\theta)\} = 0, \;\; \theta \in \Theta, \;\; i = 1, \ldots, n. \tag{1.1}$$

Often, the expectation in (1.1) has to be conditional on "the past" in order to get a martingale structure allowing for the derivation of limiting distributional results of estimators based on s. A Z-estimator $\hat{\theta}$ based on the score function s is defined as the solution of

$$\frac{1}{n} \sum_{i=1}^{n} s_i(\hat{\theta}) = 0. \tag{1.2}$$

Under sufficient regularity conditions, this estimator is asymptotically normal with influence function

$$D(\theta)^{-1} s_i(\theta), \tag{1.3}$$

where $D(\theta) = -E_\theta \dot{s}_i(\theta)$ and $\dot{s}$ is the derivative of s with respect to θ. Observe that we have

$$D(\theta) = E_\theta s_i(\theta) \dot{l}_i(\theta), \tag{1.4}$$

where $\dot{l}_i(\theta)$ denotes the derivative of the log-likelihood for θ. This property is well-known for $s = \dot{l}$ (in which case $D(\theta)$ is the Fisher information). An immediate consequence of the Hájek-Le Cam convolution theorem is that, for models that are regular in the sense that they satisfy the Local Asymptotic Normality property (LAN), optimal estimators have the influence function (1.3) based on $s = \dot{l}$.

The above analysis is, by construction, a parametric one. They key idea in semiparametric analysis is to reduce the semiparametric problem to a specific well-chosen parametric one. This special parametric model is called the least-favorable parametric submodel. Compare also Newey (1990). For completeness, we repeat the argument here. First, consider an arbitrary parametric submodel of the semiparametric model under consideration. Obviously, since the information for statistical inference decreases if one enlarges the model, lower bounds on the local and asymptotic behavior of estimators in the parametric submodel, are also lower bounds for the behavior of estimators in the complete semiparametric model. The supremum of the lower bounds over the class of all parametric submodels gives, hence, also a lower bound for the semiparametric model. The second problem is to prove that a given lower bound is sharp. To prove that a given bound is sharp, is most often done by providing a semiparametric estimator attaining this bound. Hence, if one finds a parametric submodel and an estimator for the semiparametric model, such that the bound of the parametric submodel is attained by the semiparametric estimator, then the bound is sharp and the estimator efficient. The particular parametric submodel is then a least-favorable parametric submodel.

In order to find the least-favorable submodel, a technique based on tangent spaces has proven to be very useful. If one passes from a parametric model (say a model where the density f of the innovations is completely known), to a semiparametric model where one supposes that f is unknown, there is usually an efficiency loss. This efficiency loss is caused by local changes in the density f that cannot be distinguished from local changes in the parameter of interest θ. Let $\dot{l}$ denote the score function for θ in the parametric model. The tangent space for f is defined as the space generated by all possible score-functions that can be obtained by changes in the nonparametric nuisance parameter f. The least-favorable parametric submodel should induce a nuisance score (i.e. an element of the tangent space) that is closest to the score induced by θ, i.e. $\dot{l}$. This element is then, by construction, the projection of $\dot{l}$ onto the tangent space. The residual of this projection defines the information left for estimating θ once f is unknown. This residual is called the efficient score-function. We will use the same technique in order to find the efficient score-function and hence the semiparametric efficiency bound.

2 Lower bounds in the ACD model

2.1 the parametric ACD model

In this paper we focus on a particular model: the Autoregressive Conditional Durations (ACD) model as introduced in Engle and Russel (1998). Suppose that we observe durations $x_1, \ldots, x_n$. These x's represent the time elapsed between two events, e.g., transactions in some asset. Let $\mathcal{F}_i$ denote the information available for modeling $x_{i+1}, x_{i+2}, \ldots$. We will set $\mathcal{F}_i = \sigma(x_i, x_{i-1}, \ldots, x_0)$, but it is very well possible to include exogenous variables in $\mathcal{F}_i$[1]. The key ingredient in the ACD model is the (conditional) mean duration time,

$$E\{x_i|\mathcal{F}_{i-1}\} = \psi_{i-1}. \tag{2.1}$$

In its most simple form, the formulation of the ACD model is completed by stipulating, e.g.,

$$\mathbf{P}\{x_i \leq x|\mathcal{F}_{i-1}\} = F(x/\psi_{i-1}), \tag{2.2}$$
$$\psi_i = \alpha + \beta x_i + \gamma\psi_{i-1}, \tag{2.3}$$

where F denotes a particular distribution function on the positive half-line. In this case, the parameter of prime interest is $\theta = (\alpha, \beta, \gamma)^T$. In a parametric setting, standard choices for F include the exponential and Weibull distributions. Often, F is normalized to have expectation one in order to identify a possible constant in the specification of ψ_i. If F is not specified, we obtain a semiparametric model. The model (2.2) is implicitly based on underlying i.i.d. innovations. It is not difficult to see that (2.2) is equivalent to saying that $\varepsilon_i = x_i/\psi_{i-1}$ defines a sequence of i.i.d. positive random variables, each with distribution function F.

2.2 the semiparametric ACD model

Often, the strong i.i.d. assumption (2.2) is considered to be unsuitable and one would like to relax this condition. In our specification, this is nothing else than allowing F to be dependent on the

[1] This is because the derivations that follow are independent of the parametric form for the conditional duration ψ_{i-1} defined in (2.1)

past as well. If it is unknown in what way F should depend on the past, a semiparametric approach seems to be the most reasonable one. We do assume that one is willing to define the set of variables that F may depend on (as is usual in econometric modeling), and we will see that the actual choice of these variables influences the semiparametric analysis.

In complete generality, we denote by $\mathcal{H}_{i-1} \subset \mathcal{F}_{i-1}$ the information set that $\mathbf{P}\{\varepsilon_i \leq \varepsilon | \mathcal{F}_{i-1}\}$ may depend on. We do not assume that $(\mathcal{H}_i)$ is a filtration. Formally, our model is now described by

$$x_i = \psi_{i-1}\varepsilon_i, \tag{2.4}$$

$$\mathcal{L}(\varepsilon_i|\mathcal{F}_{i-1}) = \mathcal{L}(\varepsilon_i|\mathcal{H}_{i-1}) \; a.s., \text{ and} \tag{2.5}$$

$$E\{\varepsilon_i|\mathcal{F}_{i-1}\} = 1. \tag{2.6}$$

The choice of $\mathcal{H}_i$ formalizes the dependence among the random variables ε_i. The model with independence can be obtained by taking $\mathcal{H}$ equal to the trivial sigma-field. There are two other important cases. If one chooses $\mathcal{H}_i = \mathcal{F}_i$, one leaves the dependence structure for the ε_i completely unrestricted. In more familiar terms, this would lead to a model that is solely characterized by the moment condition (2.1). One could also set $\mathcal{H}_i = \sigma(\varepsilon_i)$. In that case, the conditional distribution of ε_i given the past, may only depend on ε_{i-1}. This induces a Markovianity assumption on the innovations. Of course, there are many more possibilities. The theoretical derivations in the rest of this paper are based on a general specification with an arbitrary $\mathcal{H}_i$ and we will specialize to the above mentioned choices in order to point out their differences from an estimation point of view.

In order to derive efficiency bounds in the semiparametric model described by (2.4)–(2.6) with an arbitrary specification of ψ_{i-1}[2] and $\mathcal{H}_{i-1}$, we follow the steps as set out in the introduction. Let θ denote the parameter of interest and write f_{i-1} for the density associated with $\mathcal{L}(\varepsilon_i|\mathcal{H}_{i-1})$. We assume that f_{i-1} admits a Radon-Nikodym derivative f'_{i-1}. We will not further discuss any regularity conditions needed. A rigorous statement of conditions needed to obtain Local Asymptotic Normality for models of the same type as the one under consideration, can be found in Drost, Klaassen, and Werker (2000).

The score function for θ can be obtained by differentiation of the log-likelihood:

$$\dot{l}_i(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta} \log\left(\frac{1}{\psi_{i-1}} f(x_i/\psi_{i-1})\right) = -\left(1 + \varepsilon_i \frac{f'_{i-1}(\varepsilon_i)}{f_{i-1}(\varepsilon_i)}\right) \frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}). \tag{2.7}$$

In order to obtain the efficient score function in the semiparametric model in which f_{i-1} remains unspecified, we need to calculate the projection of $\dot{l}_i(\theta)$ on the tangent space generated by the nuisance function f_{i-1}. This tangent space $T_i(\theta)$ is easily seen to be given by all functions $h_{i-1}(\varepsilon_i)$ of ε_i for which:

$$h_{i-1}(\cdot) \quad \text{is} \quad \mathcal{H}_{i-1} - \text{measurable}, \tag{2.8}$$

$$0 = E\{h_{i-1}(\varepsilon_i)|\mathcal{H}_{i-1}\} = \int_\varepsilon h_{i-1}(\varepsilon)\mathrm{d}\mathbf{P}\{\varepsilon_i \leq \varepsilon|\mathcal{H}_{i-1}\}, \tag{2.9}$$

$$0 = E\{\varepsilon_i h_{i-1}(\varepsilon_i)|\mathcal{H}_{i-1}\} = \int_\varepsilon \varepsilon h_{i-1}(\varepsilon)\mathrm{d}\mathbf{P}\{\varepsilon_i \leq \varepsilon|\mathcal{H}_{i-1}\}. \tag{2.10}$$

Condition (2.8) follows from the fact that f_{i-1} is known to depend on $\mathcal{H}_{i-1}$ only, so that scores obtained by local changes in f_{i-1} depend also on $\mathcal{H}_{i-1}$ only. Condition (2.9) is the standard constraint in tangent space calculations following from the fact that densities by definition integrate to one. In more classical terms it represents the condition that expectations of score functions are always zero (compare (1.1)). Finally, condition (2.10) results from the moment restriction (2.6). The argument is as follows. Local changes h_{i-1} in f_{i-1} induce a change in the first (conditional) moment of $\int_\varepsilon \varepsilon h_{i-1}(\varepsilon)\mathrm{d}\mathbf{P}\{\varepsilon_i \leq \varepsilon|\mathcal{H}_{i-1}\}$. However, this moment is restricted by condition (2.10).

[2]We require, as usual, that ψ_{i-1} is $\mathcal{F}_{i-1}$-measurable

Therefore, this change must always be zero (since otherwise one would not remain in the specified model).

Lemma 2.1 *In the model* (2.4)-(2.6) *the projection of the score function* $\dot{l}_i(\theta)$ *in* (2.7) *on the tangent space* $T_i(\theta)$ *defined by* (2.8)-(2.10) *is given by*

$$-\left[\left(1+\varepsilon_i\frac{f'_{i-1}(\varepsilon_i)}{f_{i-1}(\varepsilon_i)}\right)+\frac{\varepsilon_i-1}{V\{\varepsilon_i|\mathcal{H}_{i-1}\}}\right]E\left\{\frac{d}{d\theta}\log(\psi_{i-1})\middle|\mathcal{H}_{i-1}\right\}. \tag{2.11}$$

PROOF: First of all, note that the proposed projection indeed satisfies conditions (2.8)–(2.10). Secondly, the residual of the proposed projection can be written as

$$\begin{aligned}&\frac{\varepsilon_i-1}{V\{\varepsilon_i|\mathcal{H}_{i-1}\}}E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\mathcal{H}_{i-1}\right\}\\ &-\left(1+\varepsilon_i\frac{f'_{i-1}(\varepsilon_i)}{f_{i-1}(\varepsilon_i)}\right)\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})-E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\mathcal{H}_{i-1}\right\}\right].\end{aligned} \tag{2.12}$$

We show that both terms are orthogonal to the tangent space. Let $h_{i-1}\in T_i(\theta)$ be arbitrary. Then, we obtain for the first term:

$$\begin{aligned}&E\left\{\frac{\varepsilon_i-1}{V\{\varepsilon_i|\mathcal{H}_{i-1}\}}E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\mathcal{H}_{i-1}\right\}h_{i-1}(\varepsilon_i)\right\}\\ &= E\left\{E\left\{(\varepsilon_i-1)h_{i-1}(\varepsilon_i)|\,\mathcal{H}_{i-1}\right\}\frac{E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\mathcal{H}_{i-1}\right\}}{V\{\varepsilon_i|\mathcal{H}_{i-1}\}}\right\}.\end{aligned}$$

From equation (2.9) and (2.10) we see that the latter term equals zero, proving the desired orthogonality.

For the second term, we obtain

$$\begin{aligned}&E\left\{\left(1+\varepsilon_i\frac{f'_{i-1}(\varepsilon_i)}{f_{i-1}(\varepsilon_i)}\right)\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})-E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\mathcal{H}_{i-1}\right\}\right]h_{i-1}(\varepsilon_i)\right\}\\ &= E\left\{E\left\{\left(1+\varepsilon_i\frac{f'_{i-1}(\varepsilon_i)}{f_{i-1}(\varepsilon_i)}\right)\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\right.\right.\right.\\ &\qquad\left.\left.\left.-\ E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\mathcal{H}_{i-1}\right\}\right]h_{i-1}(\varepsilon_i)\middle|\mathcal{F}_{i-1}\right\}\right\}\\ &= E\left\{\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})-E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\mathcal{H}_{i-1}\right\}\right]\right.\\ &\qquad\left.\times\ E\left\{\left(1+\varepsilon_i\frac{f'_{i-1}(\varepsilon_i)}{f_{i-1}(\varepsilon_i)}\right)h_{i-1}(\varepsilon_i)\middle|\mathcal{F}_{i-1}\right\}\right\}\\ &= E\left\{\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})-E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\mathcal{H}_{i-1}\right\}\right]\right.\\ &\qquad\left.\times\ E\left\{\left(1+\varepsilon_i\frac{f'_{i-1}(\varepsilon_i)}{f_{i-1}(\varepsilon_i)}\right)h_{i-1}(\varepsilon_i)\middle|\mathcal{H}_{i-1}\right\}\right\},\end{aligned}$$

where the last inequality follows from (2.5). It is easily seen that this expression equals zero. This completes the proof of the lemma.

As mentioned before, the residual of the projection in (2.12) is the efficient score function. Optimal semiparametric estimators should be based on this score-function. However, (2.12) cannot be used directly, since it depends on the unknown (conditional) density f_{i-1}. In Section 3 we discuss how to estimate f_{i-1} in order to get a semiparametrically efficient estimator for θ. In case the efficient score function (2.12) equals the parametric score function (2.7) we say that adaptiveness occurs. Thus, adaptiveness means that the projection on the tangent space (2.11) is zero. In that case, there is as much information in the semiparametric model as in the parametric model for estimating θ: the parametric score and the semiparametrically efficient score coincide. In the ACD model as we consider it, there are two ways in which adaptiveness can occur. First of all, it could be that

$$-\left(1+\varepsilon_i \frac{f'_{i-1}(\varepsilon_i)}{f_{i-1}(\varepsilon_i)}\right) - \frac{\varepsilon_i - 1}{V\{\varepsilon_i|\mathcal{H}_{i-1}\}} = 0.$$

It is easily seen that this is equivalent to, for some $c > 0$

$$f_{i-1}(\varepsilon) = \frac{c^{-1/c}}{\Gamma(1/c)} \varepsilon^{1/c-1} \exp(-\varepsilon/c). \tag{2.13}$$

Hence, adaptiveness only occurs if the conditional innovations' distribution is a Gamma distribution (rescaled to have expectation 1). Secondly, it could be that

$$E\left\{ \frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}) \middle| \mathcal{H}_{i-1} \right\} = 0.$$

However, for the specification (2.3), this condition generally does not hold.

Summarizing, adaptiveness only occurs if the true (conditional) density f_{i-1} belongs to the Gamma class. The practical consequence of such a result is, of course, limited since the bound is calculated in a model that does not make any distributional assumptions. However, it is well-known that densities for which adaptiveness occurs are generally also the densities for which the QMLE is consistent (see, e.g., Bickel (1982)). This shows that, a QMLE type estimator may be used if and only if it is based on a Gamma distribution. Since for these densities $1 + \varepsilon f'(\varepsilon)/f(\varepsilon)$ is always proportional to $1 - \varepsilon$, the obtained QMLE estimators are in fact identical. The estimator thus obtained is consistent for the full semiparametric model, but only efficient if the true density is of the Gamma type. An estimator that is always efficient will be discussed in Section 3.

The information for estimating θ in the parametric model is given by the (limiting) variance of the parametric score (2.7). Assuming stationarity this yields

$$E\left\{ J_{f_{i-1}} \left(\frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}) \right)^2 \right\},$$

where J_f denotes the Fisher information for scale:

$$J_f = \int \left(1 + \varepsilon \frac{f'(\varepsilon)}{f(\varepsilon)}\right)^2 f(\varepsilon)\mathrm{d}\varepsilon.$$

The information loss, with respect to the parametric model, is given by the variance of (2.11):

$$E\left\{ \left(J_{f_{i-1}} - \frac{1}{V\{\varepsilon_i|\mathcal{H}_{i-1}\}} \right) \left(E\left\{ \frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}) \middle| \mathcal{H}_{i-1} \right\} \right)^2 \right\}. \tag{2.14}$$

Note that the information loss is indeed zero (adaptiveness) if the (conditional) density f_{i-1} belongs to the Gamma class, since we have, by the Cauchy-Schwarz inequality,

$$J_{f_{i-1}} \int (\varepsilon - 1)^2 f_{i-1}(\varepsilon)\mathrm{d}\varepsilon \geq 1 = \left[\int_\varepsilon (\varepsilon - 1) \left(1 + \varepsilon \frac{f'_{i-1}(\varepsilon)}{f_{i-1}(\varepsilon)}\right) f_{i-1}(\varepsilon)\mathrm{d}\varepsilon \right]^2,$$

with equality if and only if f_{i-1} is of the form (2.13). The information in the semiparametric model is given by the variance of the residual of the projection which, by the Pythagorean theorem, equals

$$E\left\{J_{f_{i-1}}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\right)^2\right\}$$
$$-E\left\{\left(J_{f_{i-1}}-\frac{1}{V\{\varepsilon_i|\mathcal{H}_{i-1}\}}\right)\left(E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\mathcal{H}_{i-1}\right\}\right)^2\right\}. \tag{2.15}$$

2.3 examples

We consider the efficiency calculations in more detail for four specific models.

Example 2.1 [iid innovations] In case $\mathcal{H}_i$ is the trivial sigma-field, we obtain that

$$E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\mathcal{H}_{i-1}\right\}$$

is a vector of constants. This implies that all components of the projection (2.11) generate the same direction in the tangent space $T_i(\theta)$. Therefore, one may consider that adaptiveness occurs if one chooses, e.g., the constant α in (2.3) as a parametric nuisance parameter. This is a particular presentation of a phenomenon that is more generally observed in time series based on i.i.d. innovations (see Drost, Klaassen, and Werker (1997)).

Example 2.2 [Markovian innovations I] In a true Markovian setting for the innovations, one would take $\mathcal{H}_i = \sigma(\varepsilon_i)$. The efficient score (2.12) does not simplify in this Markovian case. General statements are therefore difficult to make in this setting.

Example 2.3 [Markovian innovations II] Another possibility would be to require that the innovations' density may only depend upon the past through ψ_{i-1}, i.e. $\mathcal{H}_i = \sigma(\psi_i)$. In that case, the second factor in (2.11) reduces to $\mathrm{d}\log\psi_{i-1}/\mathrm{d}\theta$ and the efficient score becomes

$$\frac{\varepsilon_i - 1}{V\{\varepsilon_i|\mathcal{H}_{i-1}\}}\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1}).$$

In this expression, the (conditional) density f_{i-1} enters only through $V\{\varepsilon_i|\psi_{i-1}\}$. This shows that the semiparametrically efficient estimator for θ is the moment estimator based on (2.6) with (optimal) instrument, compare Wefelmeyer (1996),

$$V\{\varepsilon_i|\psi_{i-1}\}^{-1}\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1}).$$

Note that our general semiparametric approach shows that the optimal semiparametric estimator is a moment estimator. We did not limit attention a priori to moment estimators.

Example 2.4 [moment condition] Consider the case where $\mathcal{H}_i = \mathcal{F}_i$. Then, the efficient score function is as in the previous example, except that in this case $V\{\varepsilon_i|\mathcal{H}_{i-1}\} = V\{\varepsilon_i|\mathcal{F}_{i-1}\}$. Again, the optimal estimator is a moment estimator with above mentioned instruments and weighting matrix. The efficient score does not alter if one enlarges the Markovian model of Example 2.3 to a model in which no additional structure is imposed. One may also turn this argument around. Starting from a model which is solely characterized by the relation (2.6), no statistical information is added if one imposes that the conditional distribution of the innovations given the past $\mathcal{F}_{i-1}$ is determined by ψ_{i-1} alone. In that sense, adaptiveness occurs between these two situations. From a (locally and asymptotically) statistical point of view, it makes no sense to "risk" possible misspecifation and consider Example 2.3.

2.4 changing the information structure

It is worth investigating the effects of changing the information structure $\mathcal{H}_i$ from a statistical point of view. First of all, note that, of course, restricting the information structure, i.e., considering $\mathcal{G}_i \subset \mathcal{H}_i$ implies restricting the model. Therefore, such an operation will generally increase the information for estimating θ. This can be seen directly from the information loss (2.14), since, by Jensen's inequality,

$$\begin{aligned}
&E\left\{\left(J_{f_{i-1}} - \frac{1}{V\{\varepsilon_i|\mathcal{G}_{i-1}\}}\right)\left(E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\,\mathcal{G}_{i-1}\right\}\right)^2\right\} \\
&= E\left\{\left(J_{f_{i-1}} - \frac{1}{V\{\varepsilon_i|\mathcal{G}_{i-1}\}}\right)\left(E\left\{E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\,\mathcal{H}_{i-1}\right\}\middle|\,\mathcal{G}_{i-1}\right\}\right)^2\right\} \\
&\leq E\left\{\left(J_{f_{i-1}} - \frac{1}{V\{\varepsilon_i|\mathcal{G}_{i-1}\}}\right)E\left\{\left(E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\,\mathcal{H}_{i-1}\right\}\right)^2\middle|\,\mathcal{G}_{i-1}\right\}\right\} \\
&= E\left\{\left(J_{f_{i-1}} - \frac{1}{V\{\varepsilon_i|\mathcal{G}_{i-1}\}}\right)\left(E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\,\mathcal{H}_{i-1}\right\}\right)^2\right\} \\
&= E\left\{\left(J_{f_{i-1}} - \frac{1}{V\{\varepsilon_i|\mathcal{H}_{i-1}\}}\right)\left(E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\,\mathcal{H}_{i-1}\right\}\right)^2\right\}.
\end{aligned}$$

Note that the first factor in (2.14) is unaltered since we must, obviously, compare DGP's that belong to the smallest model in order for these calculations to make sense. Thus, both $J_{f_{i-1}}$ and $V\{\varepsilon_i|\mathcal{H}_{i-1}\} = V\{\varepsilon_i|\mathcal{G}_{i-1}\}$ have to $\mathcal{G}_{i-1}$-measurable in the derivation above.

From a theoretical point of view, one might want to consider relaxing the assumptions in a semiparametric model, as long as this has no repercussions on efficiency bounds for the parameters of interest. From the semiparametric information matrix in (2.15), we find that no information loss occurs if one enlarges the filtration $\mathcal{H}_i$ so as not to influence the conditional expectation of $(\mathrm{d}/\mathrm{d}\theta)\log(\psi_{i-1})$.

3 Construction of efficient estimators

We follow standard lines for the construction of efficient estimators in the models under consideration. The idea is to apply a one-step improvement of a given $\sqrt{n}$-consistent estimator. Let $\dot{l}_i$ denote the efficient score as obtained above, for a given model, i.e., for a given specification of $\mathcal{H}_i$. The approach is as follows:

1. Construct a $\sqrt{n}$-consistent initial estimator $\tilde{\theta}_n$.
2. Construct a new estimator from

$$\hat{\theta}_n = \tilde{\theta}_n + \left(\frac{1}{n}\sum_{i=1}^{n}\dot{l}_i(\tilde{\theta}_n)\dot{l}_i(\tilde{\theta}_n)^T\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\dot{l}_i(\tilde{\theta}_n). \tag{3.1}$$

Unknown (conditional) densities are consistently estimated by kernel methods, and the Nadaraya-Watson estimator is a consistent estimator of the conditional variance of the innovations.

The idea of such an approach is rather old. Intuitively, the estimator $\tilde{\theta}_n$ brings you in a $\sqrt{n}$-neighborhood of the true value θ_0. Then, in order to obtain a locally and asymptotically efficient estimator, we need to achieve an influence function

$$\left(E\dot{l}_i(\theta_0)\dot{l}_i(\theta_0)^T\right)^{-1}\dot{l}_i(\theta_0).$$

The expectation in this efficient influence function is estimated by the corresponding sample mean. Alternatively, the approach may be explained as follows. The local Gaussian behavior of the model (following from the LAN property), implies that the log-likelihood is approximately quadratic. The estimator $\hat{\theta}_n$ is then the maximum likelihood estimator obtained from maximizing the approximate quadratic log-likelihood following from the initial estimator $\tilde{\theta}_n$. For more details on the construction, we refer the reader to, e.g., BKRW (1993).

4 Simulation results

In this section, we consider the behavior of the estimators discussed above in a simulation study. We use the specification (2.3) with parameters $(\beta, \gamma) = (0, 0)$, $(\beta, \gamma) = (.9, 0)$, or $(\beta, \gamma) = (.1, .8)$ with in all cases the constant standardized as $\alpha = 1 - \beta - \gamma$.

In this first draft we present a small simulation study where the innovations are simulated as i.i.d. random variables with either an exponential distribution or a Weibull distribution with parameter .75[3]. We present results for several estimators. QMLE refers to the estimator based on the assumption that the innovations follow a Gamma distribution. MLE refers to the (parametric) maximum likelihood estimator based on the true density for the innovations. The semiparametric estimators under the i.i.d. assumption (see Example 2.1) is denoted as SP1. The semiparametric estimators SP2 and SP3 refer to Examples 2.3 and 2.4. As argued in Section 2, the semiparametric lower bound will not be influenced by the exact specification of $\mathcal{H}_{i-1}$. For SP2, the variance is estimated on a Nadaraya-Watson estimator using past errors, while the variance estimator in SP3 uses past observations. All simulation results are based on 1000 observations and 500 replications.

The first conclusions are stimulating. For the exponential errors, we know that the QMLE estimator is optimal, and hence the semiparametric lower bounds are equal to the bound induced by the QMLE=MLE estimator. This is reflected by the approximations in square brackets of the (simulated) inverses of the theoretical Fisher information matrix. The realized variances of the estimators are quite close, implying that more advanced estimators based on kernel procedures do not suffer from small sample biases in the current set-up.

For the Weibull errors, the situation is slightly different since the QMLE estimator is not the optimal one. As we can see from the figures between square brackets in Table 2, the theoretical differences between the different estimators are small. Just as for the exponential case, this is reflected by the realized variances of the estimators.

Since the small-sample behavior of the different estimators is quite promising in the classical i.i.d. framework, we expect the semiparametric procedures also to be reliable in more complex, non-i.i.d. situations. This topic will be included in an updated version of this paper.

5 Empirical illustration

We illustrate our results using durations observed at the Paris Bourse for transactions in Alcatel. The observations cover July and August 1996. During this period all transaction are observed. The trading system at the opening of the Paris Bourse differs from that during the day. In order to avoid problems caused by this, we deleted observations with trades within 15 minutes of the opening, compare Gourieroux and Jasiak (1999). Simultaneous trades where aggregated, so that there are no zero durations in our dataset. These simultaneous trades are usually due to large orders on one side of the market that are matched against several orders on the other side.

In order to get an idea of the data, Figure 5.1 shows the trading intensity for five consecutive days. Note that the is a clear flattening of the cumulative intensity around lunch time. There is an

[3] Other simulations including Gamma specifications are available on request

Exponential innovations									
	$(\beta,\gamma)=(0,0)$			$(\beta,\gamma)=(.9,0)$			$(\beta,\gamma)=(.1,.8)$		
	α	β	γ	α	β	γ	α	β	γ
QMLE=MLE	1.0002	-0.0007	n.i.	0.1000	0.8995	-0.0012	0.1129	0.0996	0.7872
QMLE=MLE	(2.01)	(1.06)		(0.084)	(2.87)	(0.37)	(2.67)	(0.62)	(4.43)
QMLE=MLE	[2.00]	[0.99]		[0.075]	[2.67]	[0.21]	[1.65]	[0.63]	[3.25]
SP1	0.9999	-0.0005	n.i.	0.0999	0.8985	-0.0010	0.1129	0.0996	0.7872
SP1	(1.96)	(1.03)		(0.085)	(2.84)	(0.37)	(2.67)	(0.62)	(4.43)
SP1	[1.94]	[0.99]		[0.074]	[2.66]	[0.21]	[1.64]	[0.63]	[3.25]
SP2	1.0053	-0.0153	n.i.	0.1007	0.8693	0.0020	0.1003	0.0864	0.8104
SP2	(1.98)	(1.05)		(0.097)	(3.34)	(0.44)	(2.26)	(0.59)	(3.94)
SP2	[1.92]	[0.98]		[0.073]	[2.59]	[0.23]	[1.64]	[0.63]	[3.25]
SP3	1.0052	-0.0150	n.i.	0.1037	0.8557	-0.0018	0.1048	0.0859	0.8064
SP3	(2.03)	(1.09)		(0.193)	(2.97)	(0.42)	(2.37)	(0.61)	(4.10)
SP3	[1.92]	[0.98]		[0.073]	[2.59]	[0.23]	[1.64]	[0.63]	[3.25]

Table 1: Simulation results for the QMLE, MLE, and semiparametric estimators. Simulates standard errors are between parentheses. In square brackets we report the theoretical standard errors for the estimators based on (simulated) inverses of the theoretical Fisher information matrix. See text for a description of the simulation setup.

Weibull innovations									
	$(\beta,\gamma)=(0,0)$			$(\beta,\gamma)=(.9,0)$			$(\beta,\gamma)=(.1,.8)$		
	α	β	γ	α	β	γ	α	β	γ
QMLE	1.0018	-0.0010	n.i.	0.1015	0.8886	-0.0006	0.1131	0.0990	0.7876
QMLE	(3.12)	(1.05)		(0.090)	(3.81)	(0.33)	(2.68)	(0.86)	(5.18)
QMLE	[2.72]	[0.86]		[0.070]	[5.22]	[0.14]	[1.70]	[0.73]	[3.45]
MLE	1.0019	0.0000	n.i.	0.1008	0.8942	0.0021	0.1123	0.0992	0.7885
MLE	(3.02)	(1.05)		(0.091)	(5.48)	(0.31)	(2.55)	(0.85)	(5.05)
MLE	[2.71]	[0.94]		[0.077]	[5.27]	[0.14]	[1.79]	[0.78]	[3.67]
SP1	1.0004	0.0010	n.i.	1.0008	0.8937	0.0021	0.1126	0.0994	0.7878
SP1	(3.03)	(0.96)		(0.092)	(5.47)	(0.31)	(2.50)	(0.83)	(4.93)
SP1	[2.68]	[0.94]		[0.077]	[5.33]	[0.14]	[1.78]	[0.78]	[3.67]
SP2	0.9873	-0.0118	n.i.	0.1026	0.8213	0.0038	0.0935	0.0796	0.8204
SP2	(2.99)	(0.93)		(0.104)	(6.43)	(0.44)	(2.15)	(0.84)	(4.64)
SP2	[2.72]	[0.97]		[0.078]	[5.39]	[0.15]	[1.82]	[0.80]	[3.77]
SP3	0.9867	-0.0119	n.i.	0.1048	0.8299	-0.0017	0.0992	0.0796	0.8148
SP3	(2.90)	(0.93)		(0.104)	(5.47)	(0.40)	(2.05)	(0.78)	(4.31)
SP3	[2.72]	[0.97]		[0.078]	[5.39]	[0.15]	[1.82]	[0.80]	[3.77]

Table 2: Simulation results for the QMLE, MLE, and semiparametric estimators. Simulates standard errors are between parentheses. In square brackets we report the theoretical standard errors for the estimators based on (simulated) inverses of the theoretical Fisher information matrix. See text for a description of the simulation setup.

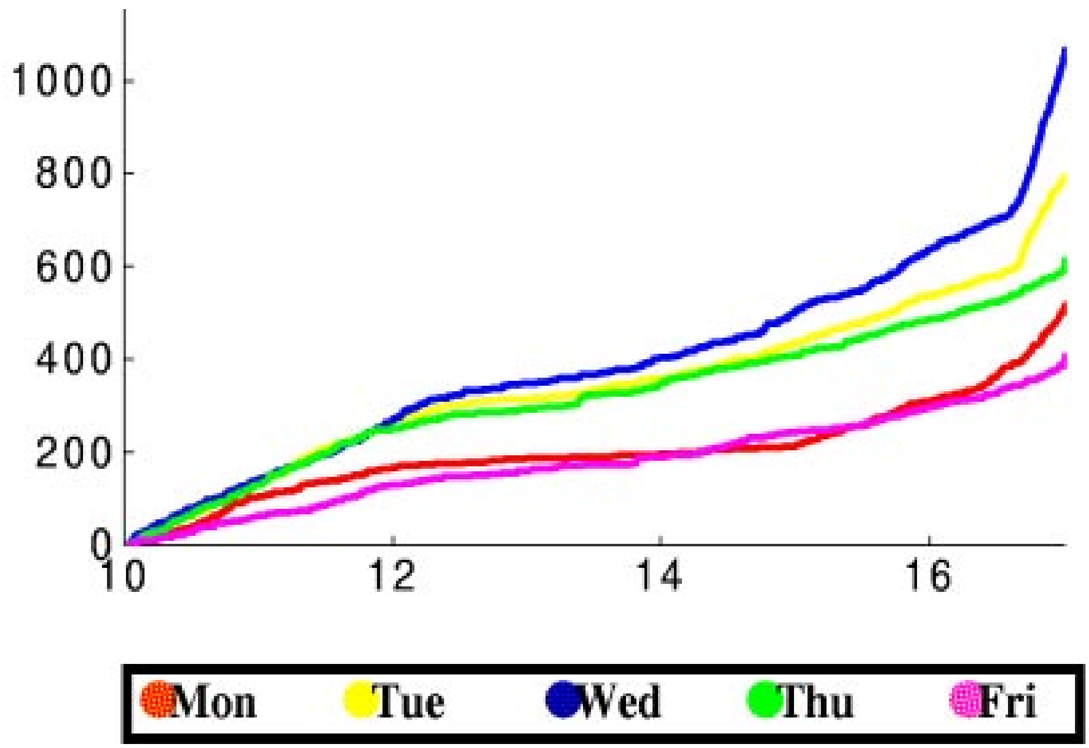

Figure 5.1: Cumulative trading intensities per day for July 22-26, 1996.

	α		β		γ	
QMLE	0.054	(0.0137)	0.142	(0.0085)	0.806	(0.0192)
IID=SP1	0.054	(0.0155)	0.153	(0.0096)	0.796	(0.0216)
MART1=SP2	0.047	(0.0138)	0.118	(0.0077)	0.828	(0.0190)
MART2=SP3	0.049	(0.0140)	0.112	(0.0068)	0.829	(0.0189)

Table 3: Estimates of the parameters in the ACD model (2.3) based on the four procedures described in the main text. Standard errors are reported in parentheses.

increase in intensity during the late afternoon as US markets open. Similar graphs for other days show comparable patterns.

We estimated the ACD model using the QMLE method and the three semiparametric methods as described before. As mentioned before, the QMLE method is efficient in case the innovations are i.i.d. exponential or gamma random variables. The first semiparametric estimator relaxes the distributional assumption an on the innovations, while still imposing independent errors. The final two semiparametric estimators also relay the independency assumption. In order to standardize the parameters, we normalized all durations to have unit mean. Estimation results are presented in Table 3. The empirical results underline the conclusions from the simulation study. The semiparametric procedures not relying on independence outperform for the estimation of β and have similar performance for the estimation of α and β. The gain with respect to β is about 30% in terms of number of observations. This means that, in order to achieve comparable confidence intervals one would need 30% more observations if using the QMLE procedure. Note that the point estimates for α do shift as well, which could indicate a form of model misspecification. We did not perform a formal test for this.

Another point of concern is stability of the parameters. One might argue that parameters should depend on the clock-time during the day. This will be investigated in the near future.

6 Concluding remarks

In this paper we consider several semiparametric ACD models. The models differ with respect to the conditional distribution of the innovations. Models with i.i.d. innovations or models described by only a moments condition are obtained as extreme examples of our more general setup. For all specifications, we derive efficient scores and semiparametric lower bounds. We discuss adaptiveness between the models. We derive semiparametric efficient estimators along standard lines. Simulation results show that the semiparametric procedures definitely outperform QMLE based procedures, but do not gain as much as is suggested by theoretical consideration. One has to keep in mind, however, that finite sample behavior of the semiparametric estimators may be further improved using more refined nonparametric estimation procedures for the nuisance functions. We illustrate our results using trade durations as observed for the Alcatel stock on the Paris Bourse.

References

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

Drost, F.C., and C.A.J. Klaassen (1997), Efficient estimation in semiparametric GARCH models, *Journal of econometrics*, **81**, 193–221 .

Drost, F.C., C.A.J. Klaassen, and B.J.M. Werker (1997), Adaptive estimation in time-series models, *The Annals of Statistics*, **25**, 786–818.

Drost, F.C., C.A.J. Klaassen, and B.J.M. Werker (1998), Semiparametric Lower Bounds in Time Series Models with Non-I.I.D. Error Structures, work in process.

Engle, R.F. and J.R. Russell (1998), Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data, *Econometrica*, **66**, 1127–1162.

Gourieroux, C. and J. Jasiak (1999), Nonlinear Autocorrelograms; an Application to Intertrade Durations, working paper.

Koul, H.L., and A. Schick (1997), Efficient estimation in nonlinear autoregressive timeseries models, *Bernoulli*, **3**, 247-277.

Kreiss, J.-P., (1987a), On adaptive estimation in stationary ARMA processes, *The Annals of Statistics*, **15**, 112–133.

Kreiss, J.-P., (1987b), On adaptive estimation in autoregressive models when there are nuisance functions, *Statistics & Decisions*, **5**, 59-76.

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

Newey, W.K., (1990), Semiparametric efficiency bounds, *Journal of Applied Econometrics*, **5**, 99–135.

Steigerwald, D.G., (1992), Adaptive estimation in time series regression models, *Journal of Econometrics*, **54**, 251–275.

Stein, C. (1956), Efficient nonparametric testing and estimation, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 187–195.

Wefelmeyer, W. (1996), Quasi-likelihood models and optimal inference, *The annals of statistics*, **24**, 405–422.