# Semiparametric duration models

Drost, F.C.; Werker, B.J.M.

*Published in:*
Journal of Business and Economic Statistics

*Publication date:*
2004

# Semiparametric Duration Models

**Feike C. Drost and Bas J. M. Werker**
CentER, Department of Econometrics, and Department of Finance, Tilburg University, P.O. Box 90153, 5000 LE, Tilburg, The Netherlands

In this article we consider semiparametric duration models and efficient estimation of the parameters in a non-iid environment. In contrast to classical time series models where innovations are assumed to be iid we show that in, for example, the often-used autoregressive conditional duration (ACD) model, the assumption of independent innovations is too restrictive to describe financial durations accurately. Therefore, we consider semiparametric extensions of the standard specification that allow for arbitrary kinds of dependencies between the innovations. The exact nonparametric specification of these dependencies determines the flexibility of the semiparametric model. We calculate semiparametric efficiency bounds for the ACD parameters, discuss the construction of efficient estimators, and study the efficiency loss of the exponential pseudolikelihood procedure. This efficiency loss proves to be sizeable in applications. For durations observed on the Paris Bourse for the Alcatel stock in July and August 1996, the proposed semiparametric procedures clearly outperform pseudolikelihood procedures. We analyze these efficiency gains using a simulation study confirming that, at least at the Paris Bourse, dependencies among rescaled durations can be exploited.

KEY WORDS: Adaptiveness; Durations; One-step improvement; Semiparametric efficiency.

## 1. INTRODUCTION

Over the last decade, the availability of financial data at a tick-by-tick level has greatly increased. The irregularly spaced data require new econometric techniques to extract the economic information contained in such data. This article concentrates on the durations between transactions on financial markets. To that extent, we base ourselves on the autoregressive conditional duration (ACD) model of Engle and Russell (1998). For the data at hand, the traditional assumption of independently and identically distributed (iid) innovations seems to be inappropriate. Therefore, we need to extend the traditional semiparametric time series models in which innovations are iid. We consider a sequence of semiparametric models imposing less and less structure on the innovations (with iid innovations on the one end of the specification and martingale innovations on the other end). To obtain efficient estimators in these semiparametric models, we must extend the semiparametric results available from the emerging literature on semiparametrics.

During recent years, enormous progress has been made in the area of semiparametric estimation. Starting with the work of Stein (1956) on the possibility of adaptiveness in the symmetric location model, the techniques have been further developed ever since. The work by Hájek and Le Cam is especially worth mentioning here. Traditionally, the models considered are based on iid observations. A fairly complete account on the state of the art in iid models can be found in the monograph by Bickel, Klaassen, Ritov, and Wellner (1993). Newey (1990) provided an overview from an econometric perspective. Semiparametric efficiency considerations and adaptiveness in time series have also been discussed, beginning with Kreiss (1987a, 1987b) for autoregressive moving average–type models. In this stream of literature, the innovations are assumed to be iid. Koul and Schick (1997) discussed nonlinear autoregressive location models, with special emphasis on the initial value problem. Drost, Klaassen, and Werker (1997) considered so-called group models, covering nonlinear location-scale time series. Steigerwald (1992) studied linear regression models in a time series context. Linton (1993) discussed linear models with autoregressive conditional hetero-scedasticity (ARCH) errors. Drost and Klaassen (1997) particularized to the generalized ARCH (GARCH) model, and Wefelmeyer (1996) calculated efficiency bounds in models with general Markov-type transitions. In this article we discuss the ACD model, which is, probabilistically, closely related to the ARCH-type models. However, all previous work on semiparametric efficient estimation for ARCH-type models assumed the innovations to be iid, an assumption that we relax significantly.

In this article we drop the iid assumption on the innovations. The semiparametric techniques mentioned earhier are used and extended to build an adequate model for durations between transactions on financial markets. Therefore, we consider semiparametric specifications in which the innovations may have dependencies of unknown functional form. As shown in Section 2, such a specification leads to a nontrivial analysis of semiparametric efficiency. The empirical results in Section 4 show that the gains from considering these more complicated semiparametric procedures may be important, at least for the present dataset and under the imposed hypotheses. Whether sizeable gains are available in other situations remains an empirical issue. Possible efficiency gains are important because they allow for much more precise parameter estimates and predictions. Also, in financial applications, where the number of observations is typically large, this may lead to a more precise empirical analysis.

The crucial ingredient in semiparametric efficiency calculations is the *efficient score function*. Let us recall this concept here. (For a rigorous treatment, consult, e.g., Bickel et al. 1993 or Drost et al. 1997.) Consider a setup where $i$ denotes the observation number and $\theta \in \Theta$ is a finite-dimensional parameter of interest. Denote (conditional) expectations under $\theta$ by $E_\theta$. In general, a score function $s_i(\cdot)$ is a random function of the parameter $\theta$, such that

$$E_{\theta_0}\{s_i(\theta_0)\} = 0, \qquad \theta_0 \in \Theta, \qquad i = 1, \ldots, n. \qquad (1)$$


Journal of Business & Economic Statistics
???? 2003, Vol. 0, No. 0
DOI 10.1198/00

Generally, the expectation in (1) must be conditional on "the past" to get a martingale structure allowing for the derivation of limiting distributional results of estimators based on $s_i$. A $Z$-estimator $\hat{\theta}$ based on the score function $s_i$ is subsequently defined as the solution of

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

In parametric settings, the optimal score function is given by the derivative of the conditional log-likelihood for $\theta$. An estimator based on the parametric score function is clearly infeasible in a semiparametric situation. However, the key idea in a semiparametric setting is to reduce the problem to a specific well-chosen parametric one. This special parametric model is called the *least-favorable parametric submodel* (cf. Newey 1990). For completeness, we repeat the argument here. First, consider an arbitrary parametric submodel of the semiparametric model under consideration. Obviously, because the information for statistical inference decreases if one enlarges the model, a lower bound (evaluated at distributions within the parametric submodel) on the asymptotic variance of estimators in the parametric submodel is also a lower bound for the behavior of estimators in the semiparametric model. Because this holds for any parametric submodel, the lower bound on the asymptotic variance of semiparametric estimators must be larger than each of these parametric lower bounds. Thus the supremum of the lower bounds over the class of all parametric submodels also gives a lower bound for the semiparametric model. The particular parametric submodel for which this supremum is attained (if it exists) is called the *least-favorable parametric submodel*. The second problem is to prove that a given lower bound is sharp. Usually, sharpness of a given bound is proved by providing a semiparametric estimator attaining this bound. Hence, if one finds a parametric submodel and an estimator in 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 is efficient.

To find the least-favorable submodel, a technique based on tangent spaces has proved to be very useful (see, e.g., Bickel et al. 1993; Van der Vaart 1998). If one passes from a parametric model (say a model in which the density $f$ of the innovations is completely known) to a semiparametric model where one supposes that $f$ is unknown, then 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 $\theta$. Let $\dot{l}$ denote the score function for $\theta$ in the parametric model. The tangent space for $f$ is defined as the space generated by all possible score functions for the nuisance parameter, that is, those score functions that can be obtained by changes in the nonparametric nuisance parameter, $f$. The least-favorable parametric submodel induces a nuisance score (i.e., an element of the tangent space) that is closest to the score $\dot{l}$ induced by $\theta$. This nuisance element is, by construction, the projection of $\dot{l}$ onto the tangent space. The residual of this projection defines the information left for estimating $\theta$ once $f$ is unknown. This residual is called the *efficient score function*. In this article we extend this idea to the situation where innovations are not likely to be iid (as in duration models). The known procedure for time series models with iid innovations is adapted to cover several forms of dependencies. In Section 2 we develop the necessary theory leading to the relevant tangent spaces and efficient score functions of the parameters of interest.

The article is organized as follows. In Section 2 we discuss duration models in their general form and develop the semiparametric theory as discussed earlier for the non-iid setting at hand. Examples (Sec. 2.3) show how common specifications may be obtained. These specifications include different assumptions on the innovations, like iid-ness or a Markov-type assumption. We consider the estimation problem in Section 3. We consider the consistency and efficiency of pseudolikelihood procedures and a construction generally leading to efficient semiparametric estimators. These semiparametric procedures prove superior to pseudolikelihood procedures. In Section 4 we discuss the properties of the durations observed on the Paris Bourse for the Alcatel stock in July–August 1996. We choose this sample because it has been considered previously in the literature (see, e.g., Ghysels, Gourieroux, and Jasiak 2001; Gourieroux and Jasiak 2000). To give a possible explanation for the semiparametric efficiency gains observed in Section 4, we study some parametric extensions of the basic ACD model in Section 5. These extensions are chosen such that they exhibit similar dependencies as we find in the Alcatel data. The simulation study in Section 5 confirms the empirical findings of Section 4. Finally, in Section 6 we provide some concluding remarks.

## 2. THE AUTOREGRESSIVE CONDITIONAL DURATION MODEL

### 2.1 The Parametric Autoregressive Conditional Duration Model

In this article we focus on the autoregressive conditional duration (ACD) model as introduced by Engle and Russell (1998). Suppose that we observe durations $x_1, \ldots, x_n$. These $x$'s represent the time elapsed between two events, for example, transactions of some asset. Let $\mathcal{F}_i$ denote the information available for modeling $x_{i+1}, x_{i+2}, \ldots$. We set $\mathcal{F}_i = \sigma(x_i, x_{i-1}, \ldots, x_0)$, but it is very well possible to include exogenous variables in $\mathcal{F}_i$. This is because the derivations that follow are independent of the parametric form of the conditional duration $\psi_{i-1}$ defined in (2). Such extra exogenous variables would allow other observable factors to influence the distribution of future durations.

The key ingredient in the ACD model is the (conditional) mean duration time,

$$E\{x_i|\mathcal{F}_{i-1}\} = \psi_{i-1}. \tag{2}$$

In its simplest form, the formulation of the ACD model is completed by stipulating, for example,

$$\mathbb{P}\{x_i \le x|\mathcal{F}_{i-1}\} = F(x/\psi_{i-1}) \tag{3}$$

and

$$\psi_i = \alpha + \beta x_i + \gamma\psi_{i-1}, \tag{4}$$

where $F$ denotes a particular distribution function (or a parametric set of distribution functions) on the positive half-line. In this case, the parameter of prime interest is $\theta = (\alpha, \beta, \gamma)^T$. In its original parametric setting, standard choices of $F$ include the exponential distribution and, as an extension, the gamma and lognormal or Weibull distributions. The distribution $F$ has to be normalized to have expectation one to identify the constant in the specification of $\psi_i$. If $F$ is not specified parametrically, we obtain a semiparametric model. The model (3) is implicitly based on underlying iid innovations. It is not difficult to see that (3) is equivalent to saying that

$$\varepsilon_i = x_i/\psi_{i-1} \tag{5}$$

defines a sequence of iid-positive random variables, each with distribution function $F$. Moreover, note that the ACD model is closely related to ARCH-type models. Rewriting (5), with $\eta_i^2 = \varepsilon_i$, $y_i^2 = x_i$, and $\sigma_{i-1}^2 = \psi_{i-1}$, yields $y_i = \sigma_{i-1}\eta_i$, the standard ARCH formulation. Our results are thus easily adapted to ARCH-type models. Finally, note that all of our results rely on the assumption that the conditional mean equation is correctly specified, as in other works in the literature.

The foregoing ACD model, including various extensions, was introduced by Engle and Russell (1998) and also studied by Engle (2000) together with a modeling of prices. These authors explicitly recognized the fact that the independence assumption in (3) implies that all temporal dependence between durations is supposed to be captured by the conditional mean duration function $\psi_i$. In that case, several parametric and nonparametric specifications of the distribution of the innovations $F$ have been studied. Zhang, Russell, and Tsay (2001) relaxed the independence assumption on the innovations by introducing a parametric regime-switching model. In this article we relax the assumption of independent innovations to semiparametric alternatives. We specify the general model in the next section.

## 2.2 The Semiparametric ACD Model

Often, the strong iid assumption (3) is considered to be unsuitable and one would like to relax it. In our specification, this is equivalent to allowing $F$ to be dependent on the 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 assume that one is willing to define a set of variables that may influence $F$, and we see that the actual choice of these variables influences the semiparametric analysis. In complete generality, we assume that $\mathbb{P}\{\varepsilon_i \le \varepsilon | \mathcal{F}_{i-1}\}$ is $\mathcal{H}_{i-1}$ measurable, where $\mathcal{H}_{i-1} \subset \mathcal{F}_{i-1}$. So the restricted information set $\mathcal{H}_{i-1}$ (of the full information set $\mathcal{F}_{i-1}$) defines the relevant past variables to be used as parameters in the conditional distribution of the innovations $\varepsilon_i$. As we show, the situation where $\mathcal{H}_{i-1}$ is strictly smaller than $\mathcal{F}_{i-1}$ is both common and relevant. We do not assume that $(\mathcal{H}_i)$ forms a filtration; that is, $\mathcal{H}_{i-1}$ is not necessarily included in $\mathcal{H}_i$. This allows for, for example, semiparametric Markov models, see below.

Formally, our semiparametric model is now described by (2) and

$$\mathcal{L}(x_i/\psi_{i-1}|\mathcal{F}_{i-1}) = \mathcal{L}(x_i/\psi_{i-1}|\mathcal{H}_{i-1}), \quad \text{a.s.} \tag{6}$$

One may choose the specification (4) of $\psi_i$, but other choices (like the ones in Engle 2000) do not change the arguments presented later. Writing $\varepsilon_i = x_i/\psi_{i-1}$, we clearly have from (2) that $E\{\varepsilon_i|\mathcal{F}_{i-1}\} = 1$. We do not make other assumptions on the innovation's distribution, although later some assumptions are needed for the nonparametric estimation. It is known that symmetry of the density sometimes helps in semiparametric estimation. In the present case, given the positiveness of duration, symmetry could be imposed for the distribution of the log-innovations. We do not make such an assumption, because its empirical foundation is unclear at the moment. Note that in our specification (6), the choice of the restricted information set $\mathcal{H}_i$ formalizes the dependence among the innovations $\varepsilon_i$.

A model with independent innovations can be obtained by taking $\mathcal{H}_i$ equal to the trivial sigma field, that is, $\mathcal{H}_i = \{\emptyset, \Omega\}$. There are two other important cases. By choosing $\mathcal{H}_i = \mathcal{F}_i$, one leaves the dependence structure of the $\varepsilon_i$ completely unrestricted. In more familiar terms, this would lead to a model characterized solely by the moment condition (2). One could also set $\mathcal{H}_i = \sigma(\varepsilon_i)$. In that case, the conditional distribution of $\varepsilon_i$, given the past, may only depend on $\varepsilon_{i-1}$. This induces a first-order Markov assumption on the innovations. In a similar manner, one can study the effect of a $K$-order Markov assumption by taking $\mathcal{H}_i = \sigma(\varepsilon_i, \ldots, \varepsilon_{i+1-K})$. Of course, there are many more possibilities. The theoretical derivations in the rest of this article are based on a general specification with an arbitrary choice of $\mathcal{H}_i$, and we specialize to the abovementioned choices to point out their differences from an estimation standpoint in Section 2.3.

To derive efficiency bounds in the semiparametric model described by (2) and (6) with an arbitrary specification of the conditional expected duration $\psi_{i-1}$ and $\mathcal{H}_{i-1}$, we follow the steps as set out in Section 1. Let $\theta$ denote the Euclidean parameter of interest describing the functional form of the conditional mean duration $\psi_{i-1}$, for example $\theta = (\alpha, \beta, \gamma)^T$ in (4). 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 $f'_{i-1}$, i.e. $f_{i-1}$ can be written as

$$f_{i-1}(\varepsilon) = \int_0^{\varepsilon} f'_{i-1}(u)\mathrm{d}u.$$

Note that this rules out, for instance, a uniform innovation distribution. Regularity conditions under which the results presented later hold are standard in the semiparametric literature (see, e.g., Bickel et al. 1993, sec. 2.1; Drost et al. 1997, sec. 2).

The score function for $\theta$ can be obtained by differentiation of the log-likelihood,

$$\begin{aligned} \dot{l}_i(\theta) &= \frac{\mathrm{d}}{\mathrm{d}\theta} \log\left(\frac{1}{\psi_{i-1}} f_{i-1}(x_i/\psi_{i-1})\right) \\ &= -\left(1 + \varepsilon_i \frac{f'_{i-1}(\varepsilon_i)}{f_{i-1}(\varepsilon_i)}\right) \frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}). \end{aligned} \tag{7}$$

To obtain the efficient score function in the semiparametric model in which the conditional density $f_{i-1}$ remains unspecified, we need to calculate the projection of the score $\dot{l}_i(\theta)$ on the tangent space generated by the nuisance function $f_{i-1}$. As we argue along the general lines of, for example, Bickel et al. (1993), this tangent space $T_i(\theta)$ is generated by all observation $i$ score functions $h_{i-1}(\cdot)$ for which

$$h_{i-1}(\cdot) \in \mathcal{H}_{i-1}, \tag{8}$$

$$0 = E\{h_{i-1}(\varepsilon_i)|\mathcal{H}_{i-1}\} = \int_\varepsilon h_{i-1}(\varepsilon)\mathrm{d}\mathbb{P}\{\varepsilon_i \le \varepsilon|\mathcal{H}_{i-1}\}, \tag{9}$$

and

$$0 = E\{\varepsilon_i h_{i-1}(\varepsilon_i)|\mathcal{H}_{i-1}\} = \int_\varepsilon \varepsilon h_{i-1}(\varepsilon)\mathrm{d}\mathbb{P}\{\varepsilon_i \le \varepsilon|\mathcal{H}_{i-1}\}. \tag{10}$$

Because we have only two conditions on $f_i$, we have two conditions on the scores in the tangent space. Condition (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 $f_{i-1}$ also depend on $\mathcal{H}_{i-1}$ only. Condition (9) is the standard constraint in tangent space calculations, following from the fact that densities by definition integrate to 1. In more classical terms, this represents the condition that expectations of score functions are always 0 [cf. (1)]. Finally, condition (10) results from the moment restriction $E\{\varepsilon_i|\mathcal{H}_{i-1}\} = E\{\varepsilon_i|\mathcal{F}_{i-1}\} = 1$. The argument is as follows. Local changes in $f_{i-1}$ represented by the score $h_{i-1}$ induce a change in the first (conditional) moment of $\int_\varepsilon \varepsilon h_{i-1}(\varepsilon)\mathrm{d}\mathbb{P}\{\varepsilon_i \le \varepsilon|\mathcal{H}_{i-1}\}$. However, this moment is restricted to be 1 by condition (2). Therefore, the change must always be 0—otherwise, one would not remain in the specified model. With these ingredients, we are ready to state the key proposition providing the lower bound for estimation of the parameters in $\psi_i$ of the general semiparametric model described by (2) and (6).

*Proposition 1.* In the semiparametric model described by (2) and (6), the projection of the score function $\dot{l}_i(\theta)$ in (7) on the tangent space $T_i(\theta)$ defined by (8)–(10) is given by $h^*_{i-1}(\varepsilon_i)$ with

$$h^*_{i-1}(\varepsilon) = -\left(1 + \varepsilon\frac{f'_{i-1}(\varepsilon)}{f_{i-1}(\varepsilon)} + \frac{\varepsilon - 1}{\mathrm{var}\{\varepsilon_i|\mathcal{H}_{i-1}\}}\right) \times E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\Big|\mathcal{H}_{i-1}\right\}. \tag{11}$$

*Proof.* First, note that the proposed projection (11) indeed belongs to the tangent space $T_i(\theta)$ because it satisfies conditions (8)–(10). Second, the residual of the proposed projection of $\dot{l}_i(\theta)$ can be written as

$$\begin{aligned}\dot{l}^*_i(\theta) &\equiv \dot{l}_i(\theta) - h^*_{i-1}(\varepsilon_i)\\ &= \frac{\varepsilon_i - 1}{\mathrm{var}\{\varepsilon_i|\mathcal{H}_{i-1}\}}E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\Big|\mathcal{H}_{i-1}\right\}\\ &\quad - \left(1 + \varepsilon_i\frac{f'_{i-1}(\varepsilon_i)}{f_{i-1}(\varepsilon_i)}\right)\\ &\quad \times \left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1}) - E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\Big|\mathcal{H}_{i-1}\right\}\right].\end{aligned} \tag{12}$$

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

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

From (9) and (10), we see that the latter term equals 0, proving the desired orthogonality.

For the second term in (12), we obtain

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

where the last equality follows from (6). It is easily seen that this expression equals 0, by first conditioning on $\mathcal{H}_{i-1}$. This completes the proof.

The proof of the foregoing proposition is indirect. Only very few constructive arguments for obtaining efficient score functions are known in the semiparametric literature. It is important to note that the efficient score functions is, as a projection, unique (see, Newey 1990 for a more general discussion).

As mentioned before, the residual (12) of the projection (11) is the efficient score function, which we denote by $\dot{l}^*_i(\theta)$. Optimal semiparametric estimators must be based on this score function. However, (12) cannot be used directly, because it depends on the unknown density $f_{i-1}$ and on $E\{(\mathrm{d}/\mathrm{d}\theta)\log(\psi_{i-1})|\mathcal{H}_{i-1}\}$. In Section 3.2 we discuss how to estimate $f_{i-1}$ and $E\{(\mathrm{d}/\mathrm{d}\theta)\log(\psi_{i-1})|\mathcal{H}_{i-1}\}$ to get a semiparametrically efficient estimator of $\theta$.

Adaptiveness occurs (by definition) in the case where the efficient score function (12) equals the parametric score function (7). Thus adaptiveness means that the projection of the parametric score on the tangent space is 0. In that case, there is (asymptotically) as much information in the semiparametric model as in the parametric model for estimating $\theta$; the parametric score and the semiparametrically efficient score coincide. In the ACD model (4), we have $\psi_{i-1} > 0$. Therefore, using $(\mathrm{d}/\mathrm{d}\theta)\log(\psi_i) = \psi_i^{-1}(\mathrm{d}/\mathrm{d}\theta)\psi_i$ and $(\mathrm{d}/\mathrm{d}\theta)\psi_i = \gamma(\mathrm{d}/\mathrm{d}\theta)\psi_{i-1} + (1, x_{i-1}, \psi_{i-1})^T$, that specification implies $(\mathrm{d}/\mathrm{d}\theta)\log(\psi_{i-1}) > 0$. Hence adaptiveness occurs if and only if

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

It is easily seen that for some positive $c_{i-1}$, this is equivalent to

$$f_{i-1}(\varepsilon) = \frac{c_{i-1}^{-1/c_{i-1}}}{\Gamma(1/c_{i-1})}\varepsilon^{1/c_{i-1}-1}\exp(-\varepsilon/c_{i-1}), \qquad \varepsilon > 0. \tag{13}$$

Hence adaptiveness occurs if and only if the conditional innovation's distribution is of the gamma type (rescaled to have expectation 1). Note that the free parameter $c_{i-1}$ may be time-varying and thus that innovations need not be iid for adaptiveness to occur. A similar result has been obtained for location models where adaptiveness occurs for the normal distribution and symmetrized square roots of chi-squared distributions (see González-Rivera 1997). In our scale case, we have adaptiveness for the exponential and gamma distributions. The practical consequence of such a result is, of course, limited, because the bound is calculated in a model that does not make any distributional assumptions.

It is well known that densities at which adaptiveness occurs also are often the densities for which the pseudo–maximum likelihood estimator (PMLE) is consistent (see, e.g., Bickel 1982). This shows that a PMLE-type estimator is consistent if and only if it is based on a gamma distribution. Because for these densities $1+\varepsilon f'(\varepsilon)/f(\varepsilon)$ is always proportional to $1-\varepsilon$, the PMLE estimators obtained are in fact identical, and the resulting PMLE is based purely on the moment condition (2). The estimator thus obtained is consistent in the full semiparametric model. In Section 3.1 we explain that the PMLE is only semiparametrically efficient under very restrictive conditions. We give an alternative estimator that is semiparametrically efficient in the model under consideration in Section 3.2.

The information for estimating $\theta$ in the parametric model is given by the variance of the parametric score (7). Assuming stationarity, this yields

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

where $J_f$ denotes the Fisher information for scale, that is,

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

The information loss of the semiparametric model, with respect to the parametric model, is given by the variance of (11),

$$E\left\{\left(J_{f_{i-1}}-\frac{1}{\operatorname{var}\{\varepsilon_i|\mathcal{H}_{i-1}\}}\right)E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\Big|\mathcal{H}_{i-1}\right\} \times E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\Big|\mathcal{H}_{i-1}\right\}^T\right\}.$$

Note that the information loss is indeed 0 (adaptiveness) if and only if the (conditional) density $f_{i-1}$ belongs to the Gamma class. This follows, because we have, by the Cauchy–Schwarz inequality,

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

with equality if and only if $f_{i-1}$ is of the form (13). The information in the semiparametric model is given by the variance of the residual of the projection that, by the Pythagorean theorem, equals

$$E\left\{J_{f_{i-1}}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\right)\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\right)^T\right\} - E\left\{\left(J_{f_{i-1}}-\frac{1}{\operatorname{var}\{\varepsilon_i|\mathcal{H}_{i-1}\}}\right)E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\Big|\mathcal{H}_{i-1}\right\} \times E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\Big|\mathcal{H}_{i-1}\right\}^T\right\}.$$

## 2.3 Examples

We consider the efficiency calculations in more detail in three specific models.

*Example 1* (IID Innovations). In the case where the restricted information set $\mathcal{H}_i$ is the trivial sigma-field, we obtain that $\dot\psi=E\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})|\mathcal{H}_{i-1}\}$ is a vector of constants. This implies that all components of the projection (11) generate the same direction in the tangent space $T_i(\theta)$. Moreover, in this case, $f_{i-1}=f$. Adaptiveness in such models is well studied (see Drost et al. 1997). The efficient score function becomes

$$\dot l_i^*(\theta)=\frac{\varepsilon_i-1}{\operatorname{var}\{\varepsilon_i\}}\dot\psi-\left(1+\varepsilon_i\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}\right)\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})-\dot\psi\right]. \quad (14)$$

*Example 2* (Markov Innovations). In a true Markov setting of the innovations, one would take $\mathcal{H}_i=\sigma(\varepsilon_i)$. The efficient score (12) does not simplify in this Markov case,

$$\dot l_i^*(\theta)=\frac{\varepsilon_i-1}{\operatorname{var}\{\varepsilon_i|\varepsilon_{i-1}\}}E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\Big|\varepsilon_{i-1}\right\} -\left(1+\varepsilon_i\frac{f'_{i-1}(\varepsilon_i)}{f_{i-1}(\varepsilon_i)}\right) \times\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})-E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\Big|\varepsilon_{i-1}\right\}\right]. \quad (15)$$

General statements are difficult to make in this setting. Clearly, the first-order Markov case is easily generalized to higher-order Markov settings.

*Example 3* (Martingale Condition). Consider the case where $\mathcal{H}_i=\mathcal{F}_i$. In that case, the second factor in (11) reduces to $\mathrm{d}\log\psi_{i-1}/\mathrm{d}\theta$, and the efficient score becomes

$$\dot l_i^*(\theta)=\frac{\varepsilon_i-1}{\operatorname{var}\{\varepsilon_i|\mathcal{H}_{i-1}\}}\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1}). \quad (16)$$

In this expression, the (conditional) density $f_{i-1}$ enters only through $\operatorname{var}\{\varepsilon_i|\mathcal{H}_{i-1}\}$. This shows that the semiparametrically efficient estimator of $\theta$ is the moment estimator based on (2) with (optimal) instrument

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

Note that our general semiparametric approach shows that in the present example, the optimal semiparametric estimator is a moment estimator. We did not limit attention to moment estimators a priori. Wefelmeyer (1996) obtained similar results in more general models specified in terms of conditional moments conditions only. Note that the same efficient score would be obtained in any model where $\mathcal{H}_{i-1}$ contains $\mathrm{d}\log\psi_{i-1}/\mathrm{d}\theta$, that is, $\mathcal{H}_{i-1}\supset\sigma(\mathrm{d}\log\psi_{i-1}/\mathrm{d}\theta)$. The efficient score does not change if one enlarges such a model to the martingale model with $\mathcal{H}_i=\mathcal{F}_i$. One may also turn this argument around. Starting from a model that is solely characterized by the relation (2), no statistical information is added if one imposes the condition that the conditional distribution of the innovations given the past $\mathcal{F}_{i-1}$ is determined by $\mathrm{d}\log\psi_{i-1}/\mathrm{d}\theta$ alone. In that sense, adaptiveness occurs between these two situations. However, the construction of efficient estimators is much simpler in cases where the restricted information set $\mathcal{H}_{i-1}$ is not too large. Therefore, from a practical standpoint, alternative specifications of the restricted information set $\mathcal{H}_{i-1}$, like the one that we use in Section 4, are relevant.

## 3. ESTIMATION IN SEMIPARAMETRIC AUTOREGRESSIVE CONDITIONAL DURATION MODELS

### 3.1 Pseudolikelihood Procedures

The most basic ACD models assume that innovations are iid and exponentially distributed. This assumption has been easily rejected in several studies (see, e.g., Engle and Russell 1998; Engle 2000), but it can be used in a pseudolikelihood procedure. The score function in the ACD model with iid exponential innovations is given by

$$(\varepsilon_i-1)\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1}). \tag{17}$$

In view of (2), this score clearly satisfies the score condition (1). Consequently, the pseudolikelihood estimator based on iid exponential innovations yields consistent estimators under standard regularity conditions. However, this estimator is only efficient under fairly restrictive conditions, which are discussed at the end of this section.

One might consider enlarging the distributional class of the innovations to accommodate the misspecification in the exponential density. However, such an enlargement may have undesirable consequences, as we will see shortly. Two classes are widely used in the literature: the gamma and the log-normal distributions. In both specifications, one added parameter makes the exponential distribution more flexible.

Let $f_\lambda$ denote the density of a normalized gamma distribution, denoted by $\Gamma(\lambda,\lambda)$, that is,

$$f_\lambda(x)\propto x^{\lambda-1}\exp(-\lambda x),$$

then we have

$$-\left(1+\varepsilon_i\frac{f'_\lambda(\varepsilon_i)}{f_\lambda(\varepsilon_i)}\right)=\lambda(\varepsilon_i-1).$$

Thus a pseudolikelihood procedure based on this gamma distribution yields a score function that is proportional to (17). Therefore, the estimator obtained is identical to the one obtained from an exponential pseudolikelihood procedure. The "extension" to gamma distributions is thus void as far as a pseudolikelihood procedure is concerned. Of course, in a parametric setting, a gamma distribution provides a more flexible way to fit the residuals than the exponential distribution.

A second popular class of distributions is the lognormal class. The density of the normalized lognormal distribution, denoted by $LN(-\frac{1}{2}\sigma^2,\sigma^2)$, is given by

$$f_{\sigma^2}(x)\propto(\sigma x)^{-1}\exp\left(-\frac{1}{2}\left(\log(x)+\frac{1}{2}\sigma^2\right)^2/\sigma^2\right).$$

In this class, the scale score function is given by

$$-\left(1+\varepsilon_i\frac{f'_{\sigma^2}}{f_{\sigma^2}}(\varepsilon_i)\right)=-\frac{1}{2}-\frac{\log(\varepsilon_i)}{\sigma^2}. \tag{18}$$

However, the score function (18) does not satisfy the score condition (1) in the full semiparametric model as defined by (2) and (6). Therefore, pseudolikelihood estimators in the ACD model based on lognormal distributions will be inconsistent. Similarly, pseudolikelihood procedures based on other parametric classes of distributions (like the Weibull distributions) will generally yield inconsistent estimates. For the Weibull distributions, this result may seem counterintuitive, because the exponential distribution is in the Weibull class. However, the inconsistency of the Weibull-based PMLE follows from the fact that the score condition (1) does not hold for the full semiparametric model.

Summarizing, the exponential distribution is essentially the only pseudodistribution for which the PMLE provides consistent estimates of the ACD parameters in semiparametric settings. However, this exponential PMLE is only semiparametrically efficient under very restrictive assumptions. Indeed, the exponential PMLE is semiparametrically efficient if and only if (17) is proportional to the efficient score (12). Because to achieve general efficiency, this must hold at all $f_{i-1}$, we find that the exponential PMLE is efficient if and only if $(\mathrm{d}/\mathrm{d}\theta)\log\psi_{i-1}$ belongs to $\mathcal{H}_{i-1}$ and $\mathrm{var}\{\varepsilon_i|\mathcal{H}_{i-1}\}$ is degenerate. Relaxing the pseudodistributional assumptions in a PMLE setting may spoil the consistency of the exponential pseudolikelihood procedure. This holds even if the relaxation includes the exponential as a special case. Although there are many other examples of this effect in the literature, it is often overlooked. These considerations confirm the adaptiveness results given after (13).

### 3.2 Construction of Efficient Semiparametric Estimators

As we have seen, the often-used PMLE does not produce efficient estimators in the semiparametric ACD model. If one does not use an exponential pseudodensity, then the PMLE may not even be consistent. To obtain semiparametrically efficient estimators, we follow standard arguments that we briefly outline here. (The interested reader is referred to Bickel et al. 1993, thm. 7.8.1, prop. 7.8.1; Drost et al. 1997, thm. 3.1, for more details.)

The idea is to improve an arbitrary given $\sqrt{n}$-consistent estimator toward an efficient estimator. Let $\tilde{\theta}_n$ denote this arbitrary $\sqrt{n}$-consistent estimator, for example, the exponential PMLE of Section 3.1. In a parametric context, where the functional form of $f_{i-1}$ is known, an efficient estimator is obtained from a one-step Newton–Raphson improvement,

$$\hat{\theta}_n=\tilde{\theta}_n+\left(\frac{1}{n}\sum_{i=1}^n\dot{l}_i(\tilde{\theta}_n)\dot{l}_i(\tilde{\theta}_n)^T\right)^{-1}\frac{1}{n}\sum_{i=1}^n\dot{l}_i(\tilde{\theta}_n). \tag{19}$$

Indeed, the estimator $\hat{\theta}_n$ is easily seen to have influence function

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

A similar procedure is followed in the semiparametric model. The parametric score function $\dot{l}_i$ in (19) must be replaced by the semiparametrically efficient score functions $\dot{l}_i^*$, outlined in Section 2.3. Here the unknown (conditional) densities and expectations need to be consistently estimated by nonparametric methods. The exact estimation procedure is irrelevant, as long as the estimators are consistent in integrated mean-squared sense.

The idea of a one-step improvement using an estimated efficient score function is rather old. Intuitively, the estimator $\tilde{\theta}_n$ brings you in a $\sqrt{n}$ neighborhood of the true value $\theta_0$. Then, to obtain a locally and asymptotically efficient estimator, we need to construct an estimator with influence function

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

The local Gaussian behavior of the model implies that the log-likelihood is approximately quadratic. The estimator $\hat{\theta}_n$ is then the MLE obtained from maximizing the approximate quadratic log-likelihood following from the initial estimator $\tilde{\theta}_n$.

In Sections 4 and 5 we use kernel estimators to estimate unknown densities and their derivatives and Nadaraya–Watson regression estimators for the conditional moments and variances that appear in the efficient score function. The density of the residuals is, generally speaking, approximately gamma-shaped. Therefore, we decided to use local bandwidth choices in the kernel estimators for the densities and their derivatives. To be precise, at a given point we started by choosing the $k$ nearest neighbors to the left of that point and the $k$ nearest neighbors to the right of that point. Here $k = n^{4/5}/\sqrt{2}$ for the model with independent innovations.The local bandwidth is chosen as the standard deviation of these $2k+1$ points. The factor $1/\sqrt{2}$ is included to enforce that the traditional bandwidth choice is obtained under a uniform distribution. The nearest-neighbor rule guarantees that the bandwidth will be smaller in regions where the density is larger. We use these bandwidths (without further constants) in Section 4 for density estimation. For bivariate densities, however, the rate $n^{-1/5}$ is replaced by $n^{-1/6}$. Conditional expectations and variances are based on nearest-neighbor estimates with the same choices as for density estimation.

## 4. PARIS BOURSE: ALCATEL

We illustrate the applicability of the proposed semiparametric techniques using durations observed at the Paris Bourse for transactions in Alcatel. The observations cover July and August 1996, comprising 43 trading days (the Paris Bourse was closed on August 15 and 16). During this period, all transactions are observed. Durations vary from 1 second to 1720 seconds. The first- and second-order raw autocorrelations are 24 and 22. The Paris Bourse opens at 10:00 A.M. and closes and 17:00 P.M. At the opening, buy and sell orders that arrived before opening are matched in a call auction. During the day, the market operates as a continuous auction. We delete trades occurring within 15 minutes of the opening, to focus on durations during the day. Simultaneous trades are aggregated, so 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. On July 4, the market opened late, but we did not exclude this date from our dataset. The average number of trades per day is 458, with a standard deviation of 184. The minimum number of trades on a date is 238; the maximum, 1,022. Ghysels et al. (2001) provided some more information on the Paris Bourse structure.

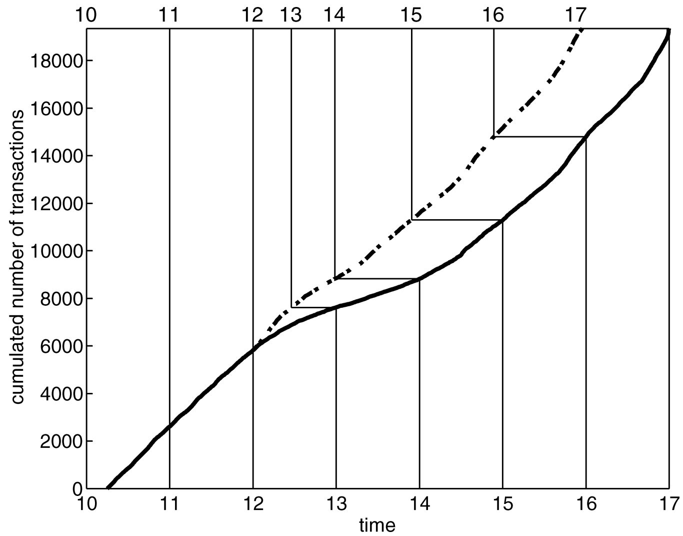


*Figure 1. Cumulative Number of Trades, by Daytime (—) and Transformed Time (- - -) for the Complete Sample.*

The mean duration in our sample is 53.2 seconds, with a standard deviation of 84.8 seconds. For each time between 10:15 A.M. and 17:00 P.M., the continuous line in Figure 1 plots the cumulative number of trades over all days. Hence the slope of the line reflects the average trading intensity (over all days) at a certain moment during the day. From Figure 1, it is clear that the average trading intensity is almost constant during the day, with lunchtime as an important exception. During lunchtime, there is a clear flattening of the average trading intensity. The lower market activity is pronouncedly present in our dataset; therefore, we must consider a mean duration function that is slightly more complicated than (4). We use the following specification:

$$\psi_i = \alpha + \delta d_i + \beta x_i + \gamma \psi_{i-1}, \tag{20}$$

where $d_i$ is an indicator for lunchtime. This extension seems to be sufficient, for the case at hand, because the trading intensity is almost constant before noon and after 2:30 P.M. We set $d_i = 1$ for transactions that occur between noon and 1:15 P.M. Note that the exponential smoothing parameter $\gamma$ will take care of a smooth transition of the "normal" intensity to the lower lunchtime intensity. By the same effect, the intensity will increase again after 1:15 P.M. This gradual change is seen in Figure 1 as the S-shaped form of the cumulative intensity around lunchtime. Engle (2000), considering IBM data, adopted a nonparametric specification of the constant in the conditional mean duration equation. There the expected durations fluctuate in a more pronounced way over the day, and the simple approach (20) would fail. As long as one is interested in the parameters $\beta$ and $\gamma$, this nonparametric approach could also be adopted in our current setup.

*Table 1. Estimates of the Parameters in the ACD Model (20) for the Alcatel Data Based on the Four Procedures Described in the Main Text*

| | $\gamma$ | | $\beta$ | | $\alpha$ | | $\delta$ | |
|---|---|---|---|---|---|---|---|---|
| PMLE | .812 | (.022) | .112 | (.016) | 3.59 | (.47) | 6.14 | (1.05) |
| Martingale | .827 | (.019) | .107 | (.016) | 3.18 | (.37) | 6.01 | (.88) |
| Markov | .810 | (.020) | .155 | (.012) | 2.33 | (.54) | 4.55 | (.90) |
| IID | .782 | (.022) | .153 | (.013) | 3.22 | (.45) | 5.39 | (1.10) |

NOTE: "PMLE" refers to the pseudolikelihood method based on an exponential innovation distribution. "Martingale" is the estimator where conditional distributions are assumed to be based on past innovations. "Markov" is the efficient semiparametric estimator in case the innovation's distribution is only affected by the last innovation. "IID" refers to the optimal semiparametric estimator in the model where the innovations are assumed to be independent over transactions. Robust standard errors are reported in parentheses; see the text for details.

As mentioned before, our theoretical results rely on a correctly specified mean duration $\psi_i$ in (20). To assess the accuracy of our specification informally, the dotted line in Figure 1 shows the cumulative number of trades against the transformed time axis in the top of the figure. The time transformation is based on estimated expected duration calculated according to (20). In particular, given the estimated values of $\alpha$, $\beta$, $\gamma$, and $\delta$, we calculate the expected value of $\psi_i$ using the recursion

$$E\psi_0 = \frac{\alpha}{1-\beta-\gamma}$$

and

$$E\psi_i = \alpha + \delta d_i + (\beta+\gamma)E\psi_{i-1},$$

where $d_i = 1$ if the $i$th trade lies between noon and 1:15 P.M. and $d_i = 0$ otherwise. Given these estimated expected durations, rescaled durations $x_i^* = x_i \frac{E\psi_0}{E\psi_i}$ are calculated from the durations $x_i$ for each day. The dotted line in Figure 1 is obtained as the cumulative number of trades based on these rescaled durations $x_i^*$. The top time axis gives the corresponding time change, which can be informally written as $\mathrm{d}t^* = \frac{E\psi_0}{E\psi_i}\mathrm{d}t$. The transformed intensity estimate is almost constant. This shows that our specification of the expected duration picks up the salient features of the data at hand. It is important to note that, except for the introduction of the lunchtime dummy $d_i$, we do not enhance the specification of the conditional mean duration $\psi_i$ or apply any preanalysis transformation to the data.

We estimated the ACD model using the PMLE method and three semiparametric methods. The first estimator (indicated by "Martingale"), uses the score (16) where the conditional variance of the innovations is estimated by a Nadaraya–Watson nonparametric regression of the squared innovations on the previous innovation $\varepsilon_{i-1}$ only, using the procedure outlined in Section 3.2. Such an approach is often followed in practice, even if theoretically the conditional variance in the optimal generalized method of moments (GMM) score (16) depends on the whole past, that is, $\varepsilon_{i-1}, \varepsilon_{i-2}, \ldots$. The second semiparametric estimator is based on a Markov assumption for the innovations $\varepsilon_i$; see Example 2. Again, unknown conditional densities in the efficient score function, in this case (15), are estimated using kernel techniques, and this estimation does not affect the asymptotic semiparametric efficiency of the estimator. The estimator thus obtained is denoted "Markov." The third semiparametric estimator imposes independence of the innovations, that is $\mathcal{H}_i = \{\emptyset, \Omega\}$, without specifying the exact distribution; see the efficient score (14) in Example 1. This estimator is denoted by "IID," and its theoretical properties in the general non-iid semiparametric model are unknown, but there is no reason to expect that even an elementary property as consistency is preserved. Because the analysis of the residuals later in this section clearly shows that the innovations are unlikely to be independent, the "IID" estimator is only given for comparison and not discussed further. Results of all four estimators for the Alcatel data are presented in Table 1.

Table 1 shows that the semiparametric procedures Martingale and Markov provide smaller standard errors than the pseudolikelihood estimator. Generally speaking, the gain is equivalent to an increase in the number of observations by about 30%. This number is obtained as the average relative efficiency of the Martingale estimator and the Markov estimator with respect to the PMLE. The GMM-type estimator Martingale and the efficient semiparametric estimator in the Markov model Markov behave similarly for the data at hand. A concern is a possible bias in the Markov estimates for the long-term levels of the durations as measured by $\alpha$ and $\delta$ in (20).

It is known that estimates for the Fisher information in semiparametric models often have weak convergence properties. Therefore, we do not base the standard errors in Table 1 on the estimated Fisher information directly; rather, we apply a resampling technique. For each day, estimates of the parameters are separately obtained. The estimates and standard errors are based on the location and dispersion of the daily estimates. Assuming that the model innovations are independent over different days, this gives consistent estimates for the standard errors. Whether the true independence in the data is sufficient to apply this technique is an empirical issue that lies outside the scope of the present investigation. We use the median and the median absolute deviation as measure for location and dispersion, to prevent a dominating effect of outlying daily estimates. The median absolute deviation is standardized such that in cases of normality, the standard deviation is obtained. Note that, on average, the daily estimates are based on approximately 500 observations. Of course, an alternative would be to use a bootstrap-type procedure, but the theoretical properties of such an approach would be difficult to establish in our non-iid situation, and the computational effort involved would be enormous.

To assess the source of the gain of the semiparametric procedures over the pseudolikelihood procedure, we study the residuals from the pseudolikelihood procedure. Figure 2 plots an estimate of the unconditional density of the innovations, together with a standard exponential density. Engle (2000) found a similar graph (see his fig. 1) and the data suggest a nonexponential marginal distribution. To study the dependencies between the innovations, Figure 3 shows the autocorrelation function of the residuals, the squared centered residuals, and the log-residuals.

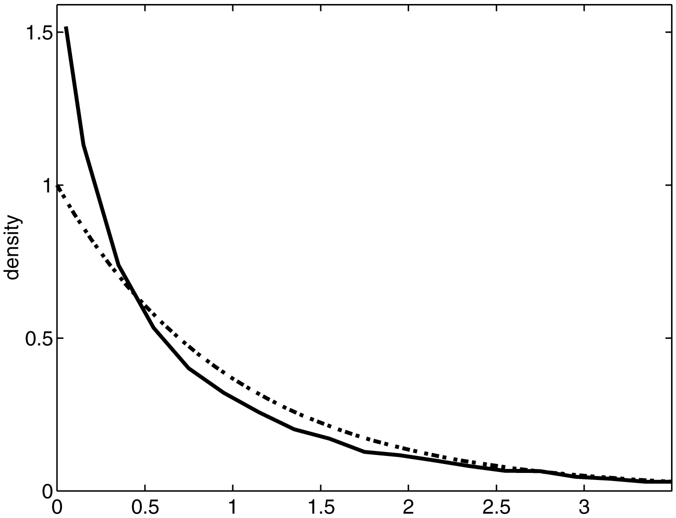


*Figure 2. Comparison of the Estimated Unconditional Density (—) and the Standard Exponential Density (- - -) of Innovations.*

We clearly see that both residuals and their squares are (almost) uncorrelated, whereas the log-residuals show a small, but significantly positive autocorrelation. The proposed semiparametric procedure effectively takes such dependencies into account. The first-order autocorrelation of log-residuals is only about .094. Apparently, such low dependencies still show up in the efficiency gains for the semiparametric estimators. Of course, another possibility could be that the conditional mean is misspecified. In that case, a nonparametric specification of the conditional mean could resolve the problem.

## 5. SIMULATION RESULTS

The results in the previous section indicate that even small dependencies in the innovations as measured by, for example, the autocorrelation of the log-residuals can induce sizeable efficiency gains of efficient semiparametric procedures over pseudolikelihood procedures. In this section we investigate this phenomenon in more detail. We consider parametric models that mimic the most salient features of the Alcatel data. We do not advocate to use these parametric models as an alternative to the semiparametric models introduced in Section 2.2, because misspecification is quite likely and may adversely affect the estimators. The parametric models in this section are used merely to confirm the properties of the semiparametric estimators in realistic settings.

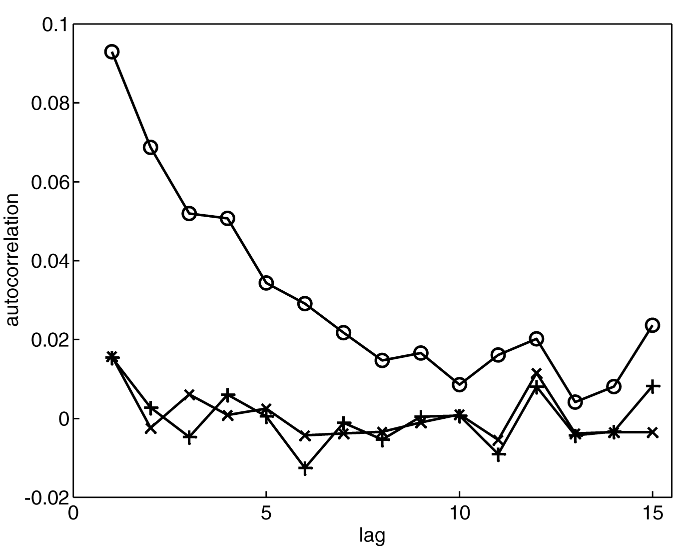


*Figure 3. ACF of Residuals (+), Squared Centered Residuals (×), and Log-Residuals (○).*

The residuals of the Alcatel durations in Section 4 show some delicate dependencies. Clearly, the model specification requires that the residuals be uncorrelated. Squared residuals also appear uncorrelated, whereas logarithmic residuals show some weak, but significant first-order autocorrelation. An extension of the classical gamma (including the exponential) or log-normal specifications incorporating these stylized features is obtained by making the parameters of those distributions time-varying. As an example, consider the possible specification

$$\varepsilon_i \sim \Gamma(\sigma_{i-1}^{-2}, \sigma_{i-1}^{-2}) \tag{21}$$

or

$$\varepsilon_i \sim LN\left(-\frac{1}{2}\log(1+\sigma_{i-1}^2), \log(1+\sigma_{i-1}^2)\right), \tag{22}$$

with

$$\sigma_{i-1}^2 = .10 + .90\varepsilon_{i-1}.$$

Note that for both specifications, the conditional variance of the innovations is indeed given by $\sigma_{i-1}^2$. Clearly, the foregoing specifications are not the only parametric ones that generate dependence structures comparable to those found in the Alcatel data. Therefore, we advocate using a semiparametric technique for econometric analysis of the structural parameters in the specification of the conditional expected duration $\psi_i$. This seems all the more reasonable because a parametrically misspecified model of the innovation's distribution does not produce consistent pseudolikelihood estimates in general. As has been pointed out before, this holds also if the parametric specification includes the iid exponential specification for which pseudolikelihood procedures are consistent.

We present results for the same four estimators as used in the analysis of the Alcatel data. The first estimator ("PMLE") is the exponential PMLE. For the second estimator ("Martingale"), the conditional variance of the innovations may depend in an arbitrary way on the past. The third estimator ("Markov") is based on the efficient semiparametric score (15) and assumes that the innovations follow a Markov process with unknown transition density. The final estimator ("IID") is the efficient semiparametric estimator in case the innovations are iid (see Example 1). The true values in (4) are $\alpha = 4.50$, $\beta = .10$, and $\gamma = .80$, and we consider both the gamma specification (21) and the lognormal specification (22). The daily number of observations is, in accordance with the average in the Alcatel data, fixed at 500. The computational effort in the simulations is substantial. Therefore, the number of replications is limited to 2,500. Again, we present location and dispersion estimates that are based on robust estimates, that is, the median and the median absolute deviation. The reported standard errors are multiplied by $\sqrt{2{,}500/43}$ to make them comparable with the empirical results of Section 4.

Table 2. Simulation Results for the ACD Model (4) With Innovation Structure (21), Where $\sigma^2_{i-1} = .1 + .9\varepsilon_{i-1}$

| | Exact scores | | | | | | Estimated scores | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\gamma$ | | $\beta$ | | $\alpha$ | | $\gamma$ | | $\beta$ | | $\alpha$ | |
| PMLE | .804 | (.0140) | .088 | (.0073) | 4.48 | (.367) | .804 | (.0140) | .088 | (.0073) | 4.48 | (.367) |
| Martingale | .800 | (.0113) | .093 | (.0052) | 4.60 | (.328) | .807 | (.0110) | .090 | (.0052) | 4.37 | (.315) |
| Markov | .800 | (.0113) | .093 | (.0052) | 4.60 | (.328) | .813 | (.0131) | .086 | (.0059) | 4.18 | (.372) |
| IID | | | | | | | .803 | (.0143) | .090 | (.0068) | 4.57 | (.399) |

NOTE: See Table 1 for an explanation of the terminology used.

The simulation results for the gamma and the lognormal specification are presented in Tables 2 and 3. For reference, the results in case the innovations are independently and identically exponentially distributed are given in Table 4. We present both the estimation results based on exact scores (i.e., as appropriate for the data-generating process at hand) and the estimation results based on estimated scores. The exact scores are calculated from the relevant formulas in Section 2.3 using the specified densities. From these exact scores, we can then infer the theoretical semiparametric efficiency gain and the theoretical ranking of the various semiparametric estimators. The effect of the nonparametric density and regression function estimates follows from comparing the results with exact and estimated scores. Table 4 presents the results in the ideal situation of iid exponential innovations $\varepsilon_i$. As discussed in Section 3, all four estimators are efficient (even adaptive) in this case. Indeed, the scores used by all estimators are the same, and, consequently, when using exact scores, the estimators are identically equal to the PMLE. In cases where the score functions are estimated, the estimators theoretically still behave the same. The simulation results confirm this, because as there is little variation with respect to standard errors. A slight increase in the variation for the semiparametric Markov estimator may be noted, which is caused by the nonparametric conditional density estimation therein. The somewhat better behavior of the Martingale estimator over the IID estimator using estimated scores is due to sampling error.

To examine the effect of dependencies on the performance of the estimators, we first consider the conditional gamma innovations in (21). In this case, the PMLE and the Martingale and Markov semiparametric estimators provide consistent estimates. There is no guarantee (known to us) that the IID semiparametric estimator is consistent in this setting with dependent innovations. Of course, calculations with exact scores cannot be performed for this estimator. The results based on exact scores, show that the theoretical standard errors of the PMLE are larger than those of the two consistent semiparametric estimators. This confirms the results of Section 3, because the conditions under which the PMLE provides efficient estimates are not met in the present simulation where $\operatorname{var}\{\varepsilon_i|\mathcal{H}_{i-1}\}$ is nondegenerate. Note, however, that since the innovations are conditionally gamma distributed, the Martingale and Markov semiparametric estimators are theoretically equal. This follows immediately from plugging in the theoretical conditional gamma density in the efficient score functions in Examples 2 and 3. However, the density estimation required in the implementation of the Markov semiparametric estimator increases its variability to the level of the PMLE, whereas the Martingale estimator retains its theoretical variability.

The former two simulations are still quite specific since, asymptotically, the Martingale and Markov semiparametric estimators coincide. Therefore, we also conducted the analysis using conditionally lognormally distributed innovations as in (22). As noted before, the lognormal distribution is not suited as a pseudodistribution in a PMLE procedure, because such an estimator generally would be inconsistent. However, it is informative to investigate the effect of lognormal innovations on the simulation results; see Table 3. Indeed, as can be seen from the bottom line in Table 3, the last semiparametric estimator (based on an iid-ness assumption on the innovations) does not produce consistent estimates in this case. As before, the Martingale and Markov semiparametric estimators show efficiency gains over the exponential PMLE; however, these estimators are no longer asymptotically equivalent. The table shows an improvement of the Markov estimator, but there also seems to be a bias–variance trade-off. The gains of the efficient semiparametric procedures over the standard exponential PMLE are, as for the Alcatel data, roughly on an order of magnitude of 30% of the number of observations.

Note that the standard errors for all simulations differ somewhat from those found for the Alcatel data. This suggests that in the Alcatel data, even more complicated dependencies than those studied in this section play a role. Clearly, the use of semiparametric techniques avoids misspecification problems inherently present when using parametric models. Note that the simulation results for the Martingale and Markov semiparametric estimators are quite similar in all cases. Apparently, for the

Table 3. Simulation Results for the ACD Model (4) With Innovation Structure (22), Where $\sigma^2_{i-1} = .1 + .9\varepsilon_{i-1}$

| | Exact scores | | | | | | Estimated scores | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\gamma$ | | $\beta$ | | $\alpha$ | | $\gamma$ | | $\beta$ | | $\alpha$ | |
| PMLE | .806 | (.0141) | .088 | (.0070) | 4.47 | (.379) | .806 | (.0141) | .088 | (.0070) | 4.47 | (.379) |
| Martingale | .795 | (.0137) | .097 | (.0064) | 4.65 | (.371) | .810 | (.0121) | .087 | (.0058) | 4.26 | (.339) |
| Markov | .796 | (.0112) | .098 | (.0051) | 4.66 | (.316) | .817 | (.0112) | .085 | (.0056) | 4.06 | (.324) |
| IID | | | | | | | .912 | (.0084) | .035 | (.0039) | 2.32 | (.242) |

NOTE: See Table 1 for an explanation of the terminology used.

*Table 4. Simulation Results for the ACD Model (4) With iid Exponential Innovations*

| | Exact scores | | | | | | Estimated scores | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\gamma$ | | $\beta$ | | $\alpha$ | | $\gamma$ | | $\beta$ | | $\alpha$ | |
| PMLE | .788 | (.0121) | .099 | (.0055) | 4.86 | (.368) | .788 | (.0121) | .099 | (.0055) | 4.86 | (.368) |
| Martingale | .788 | (.0121) | .099 | (.0055) | 4.86 | (.368) | .788 | (.0125) | .098 | (.0056) | 4.82 | (.386) |
| Markov | .788 | (.0121) | .099 | (.0055) | 4.86 | (.368) | .791 | (.0135) | .096 | (.0058) | 4.72 | (.424) |
| IID | .788 | (.0121) | .099 | (.0055) | 4.86 | (.368) | .790 | (.0132) | .098 | (.0058) | 4.78 | (.397) |

NOTE: See Table 1 for an explanation of the terminology used.

specifications chosen in this section, the respective scores (15) and (16) are close.

Summarizing, the simulations confirm that significant efficiency gains may be obtained from the use of semiparametric procedures. We prefer the theoretically optimal semiparametric estimators. Even if large numbers of observations are available for the study of intraday durations, the semiparametric procedures allow for much more precise empirical analysis and prediction. Moreover, with large datasets the distortions induced by the nonparametric density estimation are likely to disappear. Recall that our results are based on a moderate sample of only 500 observations.

## 6. CONCLUDING REMARKS

We have discussed optimal estimation in semiparametric duration models. The models differ in the specification of the possible dependencies between the innovations. These specifications range from the case where innovations are iid with unknown density to completely arbitrary dependencies that only impose an identifying martingale restriction. For these specifications, we derived the efficient score functions for the parameters of interest that govern the conditional expected duration. We also showed that the often-used exponential PMLE is only efficient under very restrictive conditions and that the other PMLEs (e.g., based on the lognormal or Weibull distribution) are not consistent. We showed that an easily implementable semiparametric estimator allows for significant (comparable to 30% of the observations) efficiency gains. To find a possible explanation for this phenomenon, we set up a simulation experiment with time-varying parameters in the innovation's distribution. The stylized features of the Alcatel data for our observation period are mimicked in this experiment. These simulations confirm the fact that the semiparametric procedures outperform pseudolikelihood procedures.

## ACKNOWLEDGMENTS

We thank Joann Jasiak for kindly providing the data. Several helpful remarks have been given by Rob Engle and Neil Shephard at the $(EC)^2$ meeting in Madrid (1999). Finally, two referees provided valuable comments. The first author acknowledges support from the European Commission through the Human Potential Programme, Statistical Methods for Dynamical Stochastic Models (DYNSTOCH).

*[Received May 2001. Revised April 2002.]*

## REFERENCES

BICKEL, P. J. (1982), "On Adaptive Estimation," *The Annals of Statistics*, 10, 647–671.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y., and WELLNER, J. A. (1993), *Efficient and Adaptive Statistical Inference for Semiparametric Models*, Baltimore: Johns Hopkins University Press.

DROST, F. C., and KLAASSEN, C. A. J. (1997), "Efficient Estimation in Semiparametric GARCH Models," *Journal of Econometrics*, 81, 193–221.

DROST, F. C., KLAASSEN, C. A. J., and WERKER, B. J. M. (1997), "Adaptive Estimation in Time-Series Models," *The Annals of Statistics*, 25, 786–818.

ENGLE, R. F. (2000), "The Econometrics of Ultra-High-Frequency Data," *Econometrica*, 68, 1–22.

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

GHYSELS, E., GOURIEROUX, C., and JASIAK, J. (2001), "Stochastic Volatility Duration Models," preprint.

GONZÁLEZ-RIVERA, G. (1997), "A Note on Adaptation in GARCH Models," *Econometric Reviews*, 16, 55–68.

GOURIEROUX, C., and JASIAK, J. (2000), "Nonlinear Autocorrelograms; An Application to Intertrade Durations," unpublisher manuscript.

KOUL, H. L., and SCHICK, A. (1997), "Efficient Estimation in Nonlinear Autoregressive Time Series Models," *Bernoulli*, 3, 247–277.

KREISS, J.-P. (1987a), "On Adaptive Estimation in Stationary ARMA Processes," *The Annals of Statistics*, 15, 112–133.

——— (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.

VAN DER VAART, A. (1998), *Asymptotic Statistics*, Cambridge, U.K.: Cambridge University Press.

WEFELMEYER, W. (1996), "Quasi-Likelihood Models and Optimal Inference, *The Annals of Statistics*, 24, 405–422.

ZHANG, M. Y., RUSSELL, J. R., and TSAY, R. S. (2001), "A Nonlinear Autoregressive Conditional Duration Model With Applications to Financial Transactions Data," *Journal of Econometrics*, 104, 179–207.