Semiparametric estimation of duration models when the parameters are subject to inequality constraints and the error distribution is unknown

By Kulan Ranasinghe and Mervyn J. Silvapulle

Department of Econometrics and Business Statistics, Monash University, P.O. Box 197, Caulfield East, Melbourne, Australia 3145.

kulan.ranasinghe@buseco.monash.edu.au, mervyn.silvapulle@buseco.monash.edu.au,

Abstract

The parameters in duration models are usually estimated by a Quasi Maximum Likelihood Estimator [QMLE], which is available in standard econometric software and is widely used in practice. This estimator is efficient if the errors are iid and exponentially distributed. If the error distribution is unknown, then it may not be the most efficient. Motivated by this, a class of semiparametric estimators have been introduced recently by Drost and Werker (2004) to improve upon the QMLE. Their method is based on the theory of efficient semiparametric estimation. A thorough evaluation of the finite sample properties of this method is not yet available. Further, although some parameters in several standard duration models are known to be nonnegative, the aforementioned semiparametric estimator does not incorporate such nonnegativity constraints. The purpose of this paper is to address these two issues. In particular, we propose a new semiparametric estimator for the case when there are inequality constraints on parameters, and report the results of simulation studies to evaluate the aforementioned two semiparametric estimators. The results lead us to conclude the following when the error distribution is unknown: (i) If the model does not impose inequality constraints on parameters then the Drost-Werker estimator is better than the QMLE, and (ii) if the model does impose inequality constraints on parameters then the estimator proposed in this paper is better than the Drost-Werker estimator and the QMLE.

1 Introduction

The availability of intraday tick-by-tick financial data increased substantially during the past two decades, which in turn has had a phenomenal impact on research in financial market microstructure. Such high frequency data are usually analyzed using essentially two classes of models: generalized autoregressive conditional heteroscedasticity [GARCH] models and duration models. In GARCH type models, the response variable is observed at equally spaced time points. An example is the hourly Dow-Jones index. By contrast, in duration models, the duration between two consecutive events, such as financial transactions, is the response variable. A range of econometric models has been proposed and studied in the literature to model the data generating process of durations. The objective of this paper is to evaluate a recently developed method of estimating duration models and propose an improvement when there are inequality constraints on some parameters, for example the parameters may be nonnegative.

To introduce the basics of the duration model, let X_i denote the duration between $(i-1)^{th}$ and the i^{th} events, $\mathcal{F}_i$ denote the information generated by the observations up to time i and $\psi_i = E(X_i \mid \mathcal{F}_{i-1})$, the expected duration. A duration model is usually expressed as $X_i = \psi_i \varepsilon_i$ where $E(\varepsilon_i) = 1$ and ε_i is referred to as the error term. The main objective of duration analysis is to model ψ_i as a function of $\{\ldots, X_{i-2}, X_{i-1}; \ldots, \psi_{i-2}, \psi_{i-1}\}$. For example, a special case of the well-known linear autoregressive conditional duration[ACD] model of Engle and Russell (1998) is the following ACD(1,1) model:

$$\psi_i = \alpha + \beta X_{i-1} + \gamma \psi_{i-1}. \tag{1}$$

More generally, ψ_i may take the form $g(\ldots, X_{i-1}; \ldots, \psi_{i-1}; \boldsymbol{\theta})$ where g is a given function and $\boldsymbol{\theta}$ is an unknown parameter. Further, g may also depend on exogenous variables.

For simplicity, let us temporarily assume that the error terms, $\varepsilon_1, \ldots, \varepsilon_n$, are independently and identically distributed with f denoting their common probability density function[pdf]. If f is known then the model can be estimated by maximum likelihood (for example, see Bauwens and Giot (2000)). On the other hand, if f is unknown, as is usually the case in practice, the quasi maximum likelihood estimator[QMLE] based on the likelihood corresponding to exponential distribution for the error terms, is consistent and asymptotically normal. This amounts to setting $f(t) = \exp(-t)$ for $t > 0$ in the likelihood function. However, such a QMLE is not necessarily the most efficient if f deviates from the exponential distribution and/or the error terms are not independent. This is important because the time-series nature of $\{X_i\}$ suggests that the error terms $\{\varepsilon_i\}$ are unlikely to be independent and identically distributed with a known density function.

Recently, Drost and Werker (2004) proposed an efficient estimator of $\boldsymbol{\theta}$ when the error distribution is unknown and $\varepsilon_1, \ldots, \varepsilon_n$ may not be independent. Their development is based on the general theory of *efficient semiparametric inference*. Detailed accounts of this topic are given in Bickel *et al.* (1993) and Tsiatis (2006). While Drost and Werker (2004) provided the theoretical derivations of their estimator, a detailed evaluation of the finite sample properties of the proposed estimator is not yet available. One of the objectives of this paper is to carry out such an evaluation.

In efficient semiparametric inference, the objective is to achieve high efficiency for inference on a finite dimensional parameter in the presence of an unknown function such as the distribution of the error terms. The theory on this topic is elegant and powerful. However, there is a significant gap between this elegant theory and its implementation for use in empirical studies.

Motivated by these considerations, we conducted a large scale simulation study to evaluate the performance of the Drost-Werker estimator[DW-estimator] for several duration models under a range of scenarios and experimented with different methods of implementation. Our results suggest that the DW-estimator is better than the usual QMLE overall, except when the true parameter is restricted by inequality constraints, such as $\theta_1 \geq 0, \theta_2 \geq 0$, and their true values are close to a certain boundary of the parameter space. This is indicated briefly in the next paragraph in the context of the model (1).

By definition, duration X_i is nonnegative, and hence $\psi_i \geq 0$. Consequently, the parameters α, β and γ in (1) must be nonnegative as well. Further, we also have $\alpha + \beta \leq 1$. However, the DW-estimator is not specifically designed to incorporate such inequality constraints and hence it may turn out to be negative even when the true parameter is known, *a priori*, to be nonnegative. If the DW-estimator $\hat{\beta}$ of β turns out to be negative, one may be tempted to simply truncate it and redefine it as $\hat{\beta} = 0$. Such a method of truncating an estimator is crude, particularly because there is already a well-developed body of statistical theory for incorporating such inequality constraints. A book-length treatment of this is given in Silvapulle and Sen (2005). In this paper, we propose an inequality constrained estimator $\bar{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. A feature of our constrained estimator is that if the DW-estimator satisfies the inequality constraints on the parameters, then the two estimators are the same. Otherwise, the constrained estimator is the point on the boundary of the parameter space that is "closest" to $\hat{\boldsymbol{\theta}}$ in some sense. A theoretical result provides the asymptotic distribution of the inequality constrained estimator $\bar{\boldsymbol{\theta}}$ and shows that it is likely to be closer to the true value than the unconstrained DW-estimator $\hat{\boldsymbol{\theta}}$. The simulation results show that if the true value of the parameter is far from the boundary of the parameter space, $\hat{\boldsymbol{\theta}}$ tends to be an interior point of the parameter space and consequently there is hardly any difference between $\hat{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\theta}}$. On the hand, if the true value is close to the boundary of $\{\boldsymbol{\theta}\}$ then our proposed constrained estimator $\bar{\boldsymbol{\theta}}$ performs better than the unconstrained DW-estimator $\hat{\boldsymbol{\theta}}$, as expected.

This paper makes two significant contributions: (i) It provides an extensive evaluation of the semiparametrically efficient DW-estimator, and (ii) it develops a new semiparametric estimator when some parameters are known to be non-negative, or more generally when there are constraints of the form $\boldsymbol{h}(\boldsymbol{\theta}) \geq \mathbf{0}$ where $\boldsymbol{h}$ is a vector function. The main findings of this paper may be summarised as follows:

1. *The errors are iid and the common distribution is exponential:* The QMLE is equal to the MLE and hence one would expect that the QMLE would be the best. The simulation results are consistent with this, but the differences between QMLE and the semiparametric estimators [SPE] tend to be generally small.

2. *There are no constraints on parameters and the errors are not iid with error distribution being exponential:* Overall, the DW-estimator performed better compared to the QMLE.

3. *There are inequality constraints on parameters:* The constrained semiparametric estimator introduced in this paper is better than the unconstrained DW-estimator.

4. *There are inequality constraints on parameters and the errors are not iid with error distribution being exponential:* In this case, the QMLE was obtaineded by maximizing the quasi-likelihood over the constrained parameter space leading to a constrained-QMLE. We considered the linear ACD model (1) and the *Power ACD* model defined by $\psi_i^\lambda = \alpha + \beta X_{i-1}^\lambda + \gamma \psi_{i-1}^\lambda$, for which α, β and γ are nonnegative. We observed that the performance of the constrained semiparametric estimator introduced in this paper relative to the constrained QMLE, depends on the the ratio β/α. If this ratio is too close to zero, then none of the estimators is uniformly best because the constrained semiparametric estimator is better for β but not for α and γ. For all other scenarios, we observed that the constrained estimator $\bar{\boldsymbol{\theta}}$ is at least as good as, and often better than, the constrained QMLE. Overall, the constrained semiparametric estimator $\bar{\boldsymbol{\theta}}$ is better than the unconstrained DW-estimator and the constrained QMLE.

We conclude that the semiparametric estimator of Drost and Werker (2004) and the inequality constrained estimator proposed in this paper are better than the QMLE that is available in standard econometric software and is widely used in practice.

The plan of the paper is as follows. Section 2 deals with the methodological aspects. In subsection 2.1, we recall some known results, and in subsection 2.2 we define the inequality constrained semiparametric estimator. Section 3 provides the results of a simulation study, section 4 provides an empirical example to illustrate the new constrained semiparametric estimator, and section 5 concludes.

2 Semiparametric Estimation of Duration Models

As in the previous section, X_i denotes the i^{th} observation of a duration variable X, $\mathcal{F}_i$ denotes the information generated by the observations up to and including X_i, $\psi_i = E(X_i \mid \mathcal{F}_{i-1})$ and

$\varepsilon_i = X_i/\psi_i$. Fernandes and Grammig (2006) provided a survey of such duration models. A simple example of each of the five main types that they studied, is given below.
1. Log-ACD Type I Model: $\log \psi_i = \alpha + \beta \log X_{i-1} + \gamma \log \psi_{i-1}$
2. Log-ACD Type II Model: $\log \psi_i = \alpha + \beta \varepsilon_{i-1} + \gamma \log \psi_{i-1}$
3. Box-Cox ACD Model: $\log \psi_i = \alpha + \beta \varepsilon_{i-1}^{v} + \gamma \log \psi_{i-1}$
4. Linear ACD Model: $\psi_i = \alpha + \beta X_{i-1} + \gamma \psi_{i-1}$
5. Power ACD Model: $\psi_i^{\lambda} = \alpha + \beta X_{i-1}^{\lambda} + \gamma \psi_{i-1}^{\lambda}$

Let $\boldsymbol{\theta}$ denote the unknown parameter; for example, $\boldsymbol{\theta} = (\alpha, \beta, \gamma)^T$ for the linear ACD(1,1) model in (1). Within the framework of this paper we do not assume that the error distribution belongs to any known parametric family. Hence $\boldsymbol{\theta}$ does not include parameters of the error distribution. To ensure that the parameters are identified, we assume that $E(\varepsilon_i \mid \mathcal{F}_{i-1}) = 1$. Usually, the errors are assumed to be independently and identically distributed [iid] for simplicity. However, the nature of the durations in practice suggests that this is unlikely to be the case in most practical situations and hence it would be desirable for the method of inference to be robust against the violation of the assumption of *iid* errors. To this end, let $\mathcal{H}_{i-1} \subset \mathcal{F}_{i-1}$ and assume that the conditional distribution of ε_i given the past depends only on the information in the set $\mathcal{H}_{i-1}$. Thus, the smaller information set $\mathcal{H}_{i-1}$ contains the relevant past variables that are assumed to affect the distribution of ε_i given the past. Now, with $\psi_i = E(X_i \mid \mathcal{F}_{i-1})$, the semiparametric[SP] model is defined formally by

$$X_i = \psi_i \varepsilon_i,\ \psi_i = g(\ldots, X_{i-1}; \ldots, \psi_{i-1}; \boldsymbol{\theta}),\ \text{and } \mathcal{L}(\varepsilon_i \mid \mathcal{F}_{i-1}) = \mathcal{L}(\varepsilon_i \mid \mathcal{H}_{i-1})$$

where g is a known function and $\mathcal{L}(\varepsilon_i \mid \mathcal{F}_{i-1})$ refers to the distribution of ε_i given $\mathcal{F}_{i-1}$. The special case of independently and identically distributed errors is obtained by setting $\mathcal{H}_i$ equal to the trivial field $\{\phi, \Omega\}$.

2.1 Semi-parametric Estimation

This subsection provides the essentials to formulate the inference problem, and states the relevant semiparametric results in a concise form for convenience, but does not contain new theoretical results. Let f_i denote the probability density function [pdf] corresponding to $\mathcal{L}(\varepsilon_i \mid \mathcal{H}_{i-1})$. We shall assume that f_i is smooth, for example, it has continuous first derivative. It follows that the conditional pdf of X_i given $\mathcal{F}_{i-1}$ is $\psi_i^{-1} f_i(x/\psi_i)$ and hence the loglikelihood $\ell(\boldsymbol{\theta})$ is given by $\ell(\boldsymbol{\theta}) = \sum \ell_i(\boldsymbol{\theta})$, where $\ell_i(\boldsymbol{\theta}) = \ln\{\psi_i^{-1} f_i(X_i/\psi_i)\}$. If f_i were known, then the maximum likelihood estimator [MLE] of $\boldsymbol{\theta}$ would be $argmax_{\boldsymbol{\theta}}\ \ell(\boldsymbol{\theta})$ and it would be asymptotically efficient. In practice, f_i is usually unknown. In this setting, the model is semiparametric and $\boldsymbol{\theta}$ can be estimated consistently by a quasi maximum likelihood estimator[QMLE] obtained by choosing the quasi likelihood equal to the loglikelihood when f_i is the exponential distribution with unit mean (see Bauwens and Giot (2001)). Efficient estimation in general semiparametric models has a specialized but a growing literature. Comprehensive accounts are given in Bickel *et al.* (1993) and Tsiatis (2006). An important result in this area is that a desirable estimator of an unknown finite dimensional parameter $\boldsymbol{\theta}$ in semiparametric models is the so called, *semiparametrically efficient estimator*, which essentially means that the estimator of $\boldsymbol{\theta}$ is efficient in some sense for the model with the density function of errors treated as unknown nuisance functions. Detailed discussions of these estimators are given in Tsiatis (2006) and Newey (1990). Here we shall state the main relevant results, without the technical details.

To introduce the semiparametrically efficient estimator of Drost and Werker (2004), first let us suppose that the error density function is known. Let $\dot{g}(\boldsymbol{\theta})$ denote $(\partial/\partial\boldsymbol{\theta})g(\boldsymbol{\theta})$ for any function g, and let $\tilde{\boldsymbol{\theta}}$ denote a $n^{1/2}$-consistent estimator of $\boldsymbol{\theta}$, for example it could be the QMLE introduced earlier. The estimator, $\{\tilde{\boldsymbol{\theta}}_n + \{n^{-1}\Sigma_{i=1}^n \dot{\ell}_i(\tilde{\boldsymbol{\theta}}_n)\dot{\ell}_i(\tilde{\boldsymbol{\theta}}_n)^T\}^{-1}n^{-1}\Sigma_{i=1}^n \dot{\ell}_i(\tilde{\theta}_n)\}$, is called the *one-step estimator*. It is asymptotically equivalent to the MLE, and is obtained by applying the Newton-Raphson iteration once, starting from any $n^{1/2}$-consistent estimator (see Bickel *et al.* (1993)).

Now, let us relax the assumption that the error density function is known. Consequently, $\dot{\ell}_i$ in the foregoing expression for the one-step estimator is also unknown. Results on semiparametrically efficient estimation suggests to replace $\dot{\ell}_i$ by $\tilde{\dot{\ell}}_i^*$, a suitable estimator of $\dot{\ell}_i^*$ which is given by

$$\dot{\ell}_i^*(\boldsymbol{\theta}) = \frac{\varepsilon_i - 1}{var\{\varepsilon_i|\mathcal{H}_i\}} E\left[\frac{\partial}{\partial\theta} log(\psi_i)|\mathcal{H}_i\right] - \left(1 + \varepsilon_i \frac{f_i'(\varepsilon_i)}{f_i(\varepsilon_i)}\right)\left[\frac{\partial}{\partial\theta} log(\psi_i) - E\left[\frac{\partial}{\partial\theta} log(\psi_i)|\mathcal{H}_i\right]\right].$$

This is the *semiparametrically efficient score function*, which corresponds to the efficient score function in classical parametric inference with finite dimensional nuisance parameters. This result is due to Drost and Werker (2004). Let $\tilde{\dot{\ell}}_i^*$ denote a 'suitable' estimator of $\dot{\ell}_i^*$. This essentially means that the former converges to the latter with respect to integrated mean squared error.

For our computations in the next section, we adopted the following method. First compute the residuals as $\tilde{\varepsilon}_i = X_i/\psi_i(\tilde{\boldsymbol{\theta}})$, $(i = 1, \ldots, n)$, and then apply the nearest neighbor method to the residuals for estimating unknown densities. For the local bandwidth at x, choose the standard deviation of the $2k + 1$ points near x, where $k = n^{4/5}/\sqrt{2}$ and the neighbourhood is chosen so that k points are on each side of x. The conditional moments and variances appearing in the foregoing expression for $\dot{\ell}_i^*(\boldsymbol{\theta})$ can be estimated using Nadaraya-Watson estimator. For example, to estimate $E[\partial/\partial\theta \log(\psi_i) \mid \mathcal{H}_i]$, we regress $(\partial/\partial\theta)\log(\tilde{\psi_i})$ on $\tilde{\psi}_i$. These steps lead to the following DW-estimator:

$$\hat{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}_n + \left(n^{-1}\sum_{i=1}^n \tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}}_n)\tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}}_n)^T\right)^{-1} n^{-1}\sum_{i=1}^n \tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}}_n) \tag{2}$$

We close this section with three special cases of the set $\mathcal{H}_i$ and the corresponding expressions for $\tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}}_n)$. The cases $\mathcal{H}_i$ equal to $\{\phi, \Omega\}$, $\sigma(\varepsilon_i)$ and $\mathcal{F}_i$ correspond to *iid*, Markov and Martingale errors. For these three cases, $\dot{\ell}_i^*(\theta)$ is given by the following three expressions respectively:

$$\{\varepsilon_i - 1/var(\varepsilon_i)\}\dot{\psi}_i - \{1 + \varepsilon_i f_i'(\varepsilon_i)/f_i(\varepsilon_i)\}(\partial/\partial\theta)log(\psi_i) - \dot{\psi}_i \tag{3}$$

$$\frac{\varepsilon_i - 1}{var\{\varepsilon_i|\varepsilon_{i-1}\}} E\left[\frac{\partial}{\partial\theta} log(\psi_i)|\varepsilon_{i-1}\right] - \left(1 + \varepsilon_i \frac{f_i'(\varepsilon_i)}{f_i(\varepsilon_i)}\right)\left[\frac{\partial}{\partial\theta} log(\psi_i) - E\left[\frac{\partial}{\partial\theta} log(\psi_i)|\varepsilon_{i-1}\right]\right] \tag{4}$$

$$\{(\varepsilon_i - 1)/\mathrm{var}(\varepsilon_i|\mathcal{H}_i)\}(\partial/\partial\theta)\log(\psi_i) \tag{5}$$

where $\dot{\psi}_i = E[(\partial/\partial\boldsymbol{\theta})\log(\psi_i)|\mathcal{H}_i]$.

The estimator $\hat{\boldsymbol{\theta}}$ in (2) with $\tilde{\dot{\ell}}_i^*$ being an estimate of $\dot{\ell}_i^*$ given in the foregoing three cases will be denoted by $\hat{\boldsymbol{\theta}}_{iid}$, $\hat{\boldsymbol{\theta}}_{Mark}$ and $\hat{\boldsymbol{\theta}}_{Mart}$ respectively. The estimator $\hat{\boldsymbol{\theta}}$ corresponding to these these three cases will be evaluated in the simulation study discussed later in this paper.

2.2 Estimation subject to inequality constraints

In the linear ACD(1,1) model $\psi_i = \alpha + \beta X_{i-1} + \gamma\psi_{i-1}$, the parameters α, β and γ are nonnegative because $\psi_i \geq 0$ and $X_i \geq 0$ for every i. However, their estimators in (2) may not satisfy such nonnegativity constraints. Therefore, it would be essential to modify the approach in Drost and Werker (2004) to ensure that such constraints are satisfied. To this end we adopt ideas that underlie constrained statistical inference; for a detailed account see Silvapulle and Sen (2005). There is no unique way to define suitable constrained estimators. We propose the following.

Let Θ denote the parameter space of $\boldsymbol{\theta}$. We shall assume that Θ is convex. Some of the results presented here would hold even if Θ is not convex, but is *Chernoff Regular* (for example, see Silvapulle and Sen 2005). However, we will not consider such general shapes for Θ here. For the linear ACD(1,1) model in (1), we have $\Theta = \{\boldsymbol{\theta} : \boldsymbol{\theta} = (\alpha, \beta, \gamma)^T, \alpha \geq 0, \beta \geq 0, \gamma \geq 0, \beta + \gamma \leq 1\}$, which is convex. We make the mild assumption that $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \boldsymbol{Z}$ where $\boldsymbol{Z} \sim N(\mathbf{0}, V)$ for some positive definite matrix V. To motivate the ideas underlying the constrained estimator to be introduced, let us temporarily suppose that $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is distributed exactly as $N(\mathbf{0}, V)$. Therefore, $\hat{\boldsymbol{\theta}}$ is distributed exactly as $N(\boldsymbol{\theta}_0, n^{-1}V)$ and we may interpret $\hat{\boldsymbol{\theta}}$ as one observation from the population $N(\boldsymbol{\theta}_0, n^{-1}V)$ with $\boldsymbol{\theta}_0 \in \Theta$. The log likelihood based on this single observation from $N(\boldsymbol{\theta}_0, n^{-1}V)$ is $(-1/2)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T V^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ and hence the corresponding MLE of $\boldsymbol{\theta}_0$ is

$$\bar{\boldsymbol{\theta}}^* = \arg\min_{\boldsymbol{\theta} \in \Theta} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T V^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}). \tag{6}$$

Therefore, $\bar{\boldsymbol{\theta}}^*$ is the projection of $\hat{\boldsymbol{\theta}}$ onto Θ with respect to the inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_V = \boldsymbol{x}^T V^{-1} \boldsymbol{y}$. The left panel in Figure 1 illustrates this for the simple case of two-dimensions and Θ equal to the first quadrant $\{\theta_1 \geq 0, \theta_2 \geq 0\}$.

Now, let us relax the assumption that $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is distributed exactly as $N(\mathbf{0}, V)$ and assume that the latter is only the limiting distribution and that V is unknown. Then, motivated by $\bar{\boldsymbol{\theta}}^*$, a natural constrained semiparametric estimator is

$$\bar{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \Theta} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T W_n^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \tag{7}$$

where W_n is positive definite. Ideally, W_n should be a consistent estimator of V, the limiting covariance matrix of $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$. Even if this were not true, $\bar{\boldsymbol{\theta}}$ would still be a consistent estimator of $\boldsymbol{\theta}_0$ as will be seen later.

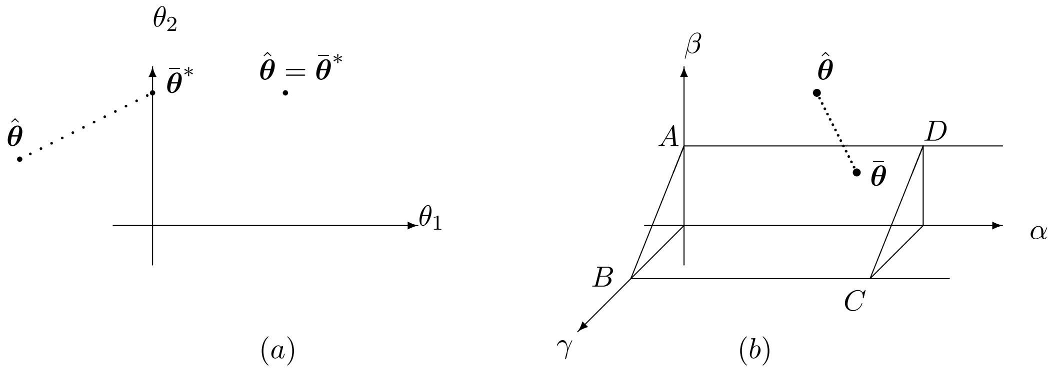

Figure 1: (a) The unconstrained estimator $\hat{\boldsymbol{\theta}}$ and the constrained estimator $\bar{\boldsymbol{\theta}}^*$ of $\boldsymbol{\theta}_0$ subject to $\boldsymbol{\theta} \in \Theta = \{(\theta_1, \theta_2) : \theta_1 \geqslant 0, \theta_2 \geqslant 0\}$, when $V = (1, 0.5 \mid 0.5, 1)$ for two possible values of $\hat{\boldsymbol{\theta}}$, one in Θ and the other outside Θ. (b) The unconstrained estimator $\hat{\boldsymbol{\theta}}$ and the constrained estimator $\bar{\boldsymbol{\theta}}$ subject to $\boldsymbol{\theta} \in \Theta = \{(\alpha, \beta, \gamma) : \alpha \geq 0, \beta \geq 0, \gamma \geq 0, \beta + \gamma \leq 1\}$ with $\hat{\boldsymbol{\theta}}$ lying outside Θ and $\bar{\boldsymbol{\theta}}$ lying on the face spanned by the rectangle $ABCD$ of the wedge-shaped Θ.

Let $\mathcal{T}(\Theta; \boldsymbol{\theta}_0)$ denote the *tangent cone* of Θ at $\boldsymbol{\theta}_0$ which is defined by

$$\mathcal{T}(\Theta; \boldsymbol{\theta}_0) = \{\boldsymbol{v} : \exists t_n \downarrow 0, \exists \boldsymbol{\theta}_n \in \Theta \text{ such that } \boldsymbol{\theta}_n \to \boldsymbol{\theta}_0 \text{ and } t_n^{-1}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) \to \boldsymbol{v}\}.$$

Intuitively, the tangent cone $\mathcal{T}(\Theta; \boldsymbol{\theta}_0)$ is constructed as follows: First approximate the boundaries of Θ at $\boldsymbol{\theta}_0$ by tangents, and then approximate Θ by the cone formed by these tangents. This is

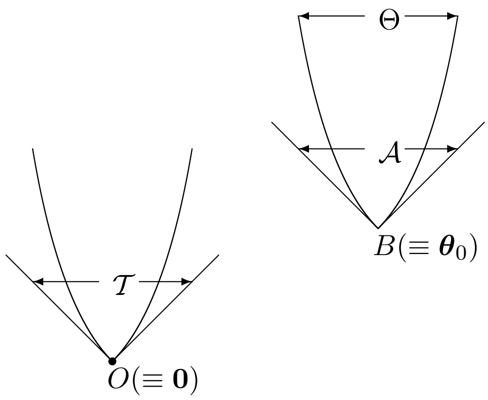

Figure 2: The *tangent cone,* $\mathcal{T}$, and the *approximating cone,* $\mathcal{A}$, of Θ at B.

called the *approximating cone* of Θ at $\boldsymbol{\theta}_0$. Now, translate the parameter space so that $\boldsymbol{\theta}_0$ moves to the origin, and hence the cone has its vertex at the origin (see Silvapulle and Sen 2005, section 4.7). These are illustrated in Figure 2.

For any $\boldsymbol{x} \in \mathbb{R}^p$, a $p \times p$ positive definite matrix W and a set $\mathcal{C}$, let

$$\|\boldsymbol{x}\|_W = \{\boldsymbol{x}^T W^{-1} \boldsymbol{x}\}^{1/2} \qquad \text{and} \qquad \Pi_W\{\boldsymbol{z} \mid \mathcal{C}\} = \arg\min_{\boldsymbol{\theta} \in \mathcal{C}} \|(\boldsymbol{z} - \boldsymbol{\theta})^T\|_W.$$

Thus, $\Pi_W\{\boldsymbol{z} \mid \mathcal{C}\}$ denotes the projection of $\boldsymbol{z}$ onto $\mathcal{C}$. A simple illustration of $\Pi(\hat{\boldsymbol{\theta}} \mid \mathbb{R}^{+2})$, which is equal to $\bar{\boldsymbol{\theta}}^*$, is given is given in Figure 1 when $\mathcal{C}$ is the positive orthant in two dimensions. Now, we provide a result about the distribution of $\bar{\boldsymbol{\theta}}$.

Proposition 1. *Suppose that* Θ *is convex and that* $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \boldsymbol{Z}$ *where* $\boldsymbol{Z} \sim N(\mathbf{0}, V)$ *for some positive definite matrix* V *and* $W_n \xrightarrow{p} W$ *where* W *and* W_n *are positive definite. Then*

$$n^{1/2}(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \Pi_W\{\boldsymbol{Z} \mid \mathcal{T}(\Theta; \boldsymbol{\theta}_0)\}. \tag{8}$$

Further, $\bar{\boldsymbol{\theta}}$ *is closer to the true value* $\boldsymbol{\theta}_0$ *than* $\hat{\boldsymbol{\theta}}$ *in the following sense:*

$$pr\{n^{1/2}\|\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_W < \delta\} \geq pr\{n^{1/2}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_W < \delta\} + o(1) \tag{9}$$

for any $\delta > 0$.

Proof. A proof of (8) when the observations are independently and identically distributed is given in (Silvapulle and Sen, 2005, section 4.9). The proof therein is not directly applicable to prove (8), but the essentials of the approach are applicable. Here, we indicate the main steps. The technical details of the proof of (8) uses the result that the parameter space Θ can be approximated by its approximating cone at the true value for the purposes of deriving the first order asymptotic properties. For example, the projections of $\hat{\boldsymbol{\theta}}$ onto Θ and onto the approximating cone $\mathcal{A}(\boldsymbol{\theta}_0)$ of Θ at θ_0 are asymptotically equivalent: $n^{1/2}(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^\dagger) = o_p(1)$ where $\boldsymbol{\theta}^\dagger = \Pi_{W_n}(\hat{\boldsymbol{\theta}} \mid \mathcal{A}(\boldsymbol{\theta}_0))$. Now treating $\boldsymbol{\theta}_0$ as the origin, we have

$$n^{1/2}(\boldsymbol{\theta}^\dagger - \boldsymbol{\theta}_0) = \Pi_{W_n}\{n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \mid \mathcal{A}(\boldsymbol{\theta}_0) - \boldsymbol{\theta}_0\} \xrightarrow{d} \Pi_W(\boldsymbol{Z} \mid \mathcal{T}(\boldsymbol{\theta}_0)),$$

the last step follows because $\Pi_W(\boldsymbol{z} \mid \mathcal{T})$ is a continuous function of $(\boldsymbol{z}, W)$.

Now, applying Proposition 3.12.3 on page 114 in Silvapulle and Sen (2005)) for the inner product defined by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^T W_n^{-1} \boldsymbol{y}$, we have that $(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T W_n^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq 0$. Therefore, $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{W_n} \geq \|\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{W_n}$. Since $W_n \xrightarrow{p} W$ and $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_p(n^{-1/2})$, we have, by Lemma 4.10.2 on page 216 in Silvapulle and Sen (2005) that $n^{1/2}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{W_n} = n^{1/2}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{W} + o_p(1)$ and $n^{1/2}\|\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{W_n} = n^{1/2}\|\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{W} + o_p(1)$. Now, the proof of (9) follows. □

The general approach to constructing a constrained estimator exploits the fact that one needs to use only the local behaviour of the objective function in an $n^{-1/2}$-neighbourhood of the true value $\boldsymbol{\theta}_0$. The foregoing $\bar{\boldsymbol{\theta}}$ adopts this approach. It is also possible to construct other estimators in such local neighbourhoods. For example, another estimator may be defined as $\hat{\boldsymbol{\theta}}(\lambda_0)$ where $\hat{\boldsymbol{\theta}}(\lambda) = \left[\tilde{\boldsymbol{\theta}}_n + \lambda\left(n^{-1}\sum_{i=1}^n \tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}}_n)\tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}}_n)^T\right)^{-1} n^{-1}\sum_{i=1}^n \tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}}_n)\right]$ for $0 \leq \lambda \leq 1$ and λ_0 is the maximum value of λ in $[0,1]$ for which $\hat{\boldsymbol{\theta}}(\lambda)$ lies in Θ. This says that the iteration moves from $\tilde{\boldsymbol{\theta}}_n$ in the direction suggested by the DW-estimator but stops before crossing the boundary of Θ.

Another estimator may be defined as $\arg\max_{\boldsymbol{\theta}\in\Theta}\ q(\boldsymbol{\theta})$ where

$$q(\boldsymbol{\theta}) = (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_n)^T n^{-1}\sum_{i=1}^n \tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}}_n) - 2^{-1}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_n)^T\Big(n^{-1}\sum_{i=1}^n \tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}}_n)\tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}}_n)^T\Big)(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_n),$$

which may be seen as a pseudo likelihood with score function $n^{-1}\sum_{i=1}^n \tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}}_n)$ and information $\left(n^{-1}\sum_{i=1}^n \tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}}_n)\tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}}_n)^T\right)$. Since the unconstrained maximum of $q(\boldsymbol{\theta})$ is the DW-estimator $\hat{\boldsymbol{\theta}}$, the foregoing estimator $\arg\max_{\boldsymbol{\theta}\in\Theta}\ q(\boldsymbol{\theta})$ can be seen as a constrained version of the DW-estimator.

3 Simulation Study

In this section, we report the results of a simulation study conducted to evaluate and compare the semiparametric estimators, $\hat{\boldsymbol{\theta}}$ and the constrained semiparametric estimator $\bar{\boldsymbol{\theta}}$ with the standard QMLE for duration models, namely the one that corresponds to $f(t) = \exp(-t)$, $t > 0$.

Design of the study:

We studied the five duration models introduced at the beginning of section 2. For each of these models, the following error distributions were studied:

$$(a)\ \varepsilon_i \sim exp(1), \qquad (b)\ \varepsilon_i \sim \Gamma(\lambda_i^{-2}, \lambda_i^2) \quad \text{and} \quad (c)\ \varepsilon_i \sim LN(-2^{-1}log(1+\lambda_i^2), log(1+\lambda_i^2)),$$

where $\Gamma(a, b)$ is the Gamma distribution with parameters (a, b), and $LN(\mu, \sigma^2)$ is the lognormal distribution. For the gamma and lognormal error distributions in the foregoing settings (b) and (c), we set $\lambda_i^2 = 0.1 + 0.9\varepsilon_{i-1}$. The estimation methods that are compared in this paper do not require the exact form of dependence of λ_i on other variables. This would enable us to evaluate the robustness of the estimators to departures from the usual assumption that the errors are *iid*.

Without loss of generality, the unconditional mean of X_i was set equal to 1. All the computations were programmed in MATLAB, and the optimizations were carried out using the optimization toolbox in MATLAB.

Number of values of $\boldsymbol{\theta}_0$: (i) Linear ACD models : 15 different values of $\boldsymbol{\theta}_0$, with some values close to the boundary of the parameter space. (ii) Linear Power ACD Model: same as for the linear ACD model. (iii) Linear ACD Type 1 : 6 values. (iv) Linear ACD Type 2: 7 values. (v) Box-Cox ACD model: 4 values.

Table 1: MSE-efficiency of $\bar{\boldsymbol{\theta}}$ relative to QMLE for the linear ACD model

True value			$\varepsilon \sim EXP$			$\varepsilon \sim NG$			$\varepsilon \sim LN$		
α_0	β_0	γ_0	α	β	γ	α	β	γ	α	β	γ
0.05	0.30	0.65	103	96	97	179	182	182	153	147	151
0.05	0.05	0.90	99	96	95	156	193	162	143	194	149
0.10	0.20	0.70	106	99	101	174	188	173	144	164	148
*0.25	0.05	0.70	58	96	61	78	162	86	65	212	76
0.10	0.15	0.75	109	99	103	169	195	170	148	174	151
0.05	0.10	0.85	102	97	97	238	207	209	181	184	174
0.20	0.20	0.60	104	101	99	149	168	145	127	155	132
*0.20	0.05	0.75	76	95	76	89	170	98	79	215	91
0.30	0.10	0.60	76	98	78	86	166	89	78	166	85
0.10	0.10	0.80	104	98	98	147	196	151	138	184	143
0.70	0.20	0.10	87	103	90	107	153	103	111	139	114
0.70	0.25	0.05	88	104	94	150	156	147	122	145	127
0.80	0.10	0.10	82	100	83	106	172	98	101	174	97
0.80	0.12	0.08	86	103	87	120	166	110	106	171	103
0.80	0.15	0.05	89	103	91	143	165	131	112	164	110

MSE-efficiency for θ_i is defined as MSE(QMLE)/MSE($\bar{\boldsymbol{\theta}}$).

Since our main objective is to compare the QMLE with the semiparametric estimators, we shall report estimates of Relative MSE Efficiency which we define as {MSE of QMLE/ MSE of the estimator}.

The results of the simulation study are based on sample size $n = 500$ and 500 repeated samples, for the linear ACD and the power ACD models with nonnegative parameters. For the other models, $n = 2000$ and 500 repeated samples. Typically with tick-by-tick data, the number of observations is usually large and hence $n = 2000$ is quite realistic.

Results:

The simulation was carried out for $\hat{\boldsymbol{\theta}}_{iid}$, $\hat{\boldsymbol{\theta}}_{Mark}$ and $\hat{\boldsymbol{\theta}}_{Mart}$. We observed that $\hat{\boldsymbol{\theta}}_{Mart}$ performed better. Therefore, in the rest of this section, we shall present the results for $\hat{\boldsymbol{\theta}}_{Mart}$ only and write $\hat{\boldsymbol{\theta}}$ for $\hat{\boldsymbol{\theta}}_{Mart}$. The results for the other estimators are available in an working paper.

The histograms of relative MSE of $\hat{\boldsymbol{\theta}}$ are shown in Figures 4 - 8. Each figure has three diagrams: the one on left, middle and right correspond to ε_i being exp(1), $\Gamma(\lambda_i^{-2}, \lambda_i^2)$ and $LN(-2^{-1}log(1 + \lambda_i^2), log(1 + \lambda_i^2))$, respectively.

The errors are iid with common error distribution exp(1):

Recall that the QMLE is equal to the MLE in this case. Since this setting is ideal for QMLE, we would expect the QMLE to perform at least as well as, if not better than, the semiparametric estimators [SPE]. The diagram on the left of each of Figures 4 - 8 show that, as expected, the

Table 2: MSE-efficiency of $\bar{\boldsymbol{\theta}}$ relative to QMLE for the linear Power ACD model

	$\varepsilon \sim EXP$				$\varepsilon \sim NG$				$\varepsilon \sim LN$			
$\boldsymbol{\theta}_0 = (\alpha_0, \beta_0, \gamma_0, \lambda_0)$	α	β	γ	λ	α	β	γ	λ	α	β	γ	λ
0.05, 0.30, 0.65, 2	121	94	93	91	1197	136	204	142	498	120	165	110
0.05, 0.05, 0.9, 2	72	84	96	69	352	136	176	101	1108	139	216	93
0.1, 0.2, 0.70, 1.5	107	96	96	92	226	149	200	207	165	128	171	119
*0.25, 0.05, 0.70, 1.5	81	86	82	47	83	117	92	53	90	127	98	59
0.1, 0.15,0.75, 2	104	89	95	85	579	132	211	136	221	122	179	105
0.05, 0.1, 0.85, 2	73	90	95	83	893	123	189	123	350	129	167	106
0.20, 0.2, 0.60, 1.5	110	97	98	90	182	144	191	198	127	125	141	120
*0.20, 0.05, 0.75,1.5	89	91	88	56	106	123	115	76	102	146	110	66
0.3, 0.1, 0.6, 0.5	94	97	95	90	92	123	95	153	82	125	89	92
0.1, 0.1, 0.8, 0.5	115	95	110	85	136	160	140	164	142	150	150	140
0.7, 0.2, 0.1, 0.5	91	99	95	89	107	115	108	136	110	114	113	129
0.7, 0.25, 0.05, 1.5	91	100	96	87	136	111	129	117	111	116	114	99
0.8, 0.1, 0.1, 0.5	91	99	92	90	99	88	93	120	110	119	107	82
0.05, 0.05, 0.9, 0.5	97	92	99	84	158	177	157	123	130	194	142	85
0.8, 0.15, 0.05, 0.5	91	104	92	104	119	97	113	155	113	112	114	101

MSE-efficiency for θ_i is defined as MSE(QMLE)/MSE($\bar{\theta}_i$).

QMLE performed at least as well as the the semiparametric estimator. However, the differences were small in most cases.

Log ACD-Type I (Figure 3):

When $\varepsilon \sim \exp(1)$, the QMLE performs at least as well as $\hat{\boldsymbol{\theta}}$, as expected, but the differences between QMLE and $\hat{\boldsymbol{\theta}}$ are small. When the error distribution is $\Gamma(\lambda_i^{-2}, \lambda_i^2)$ or $LN(-2^{-1}log(1+\lambda_i^2), log(1+\lambda_i^2))$, $\hat{\boldsymbol{\theta}}$ perform significantly better than the QMLE. These results show that for the Log ACD-Type I model, the semiparametric estimator is better than the QMLE.

Log ACD-Type II (Figure 4):

When $\varepsilon \sim \exp(1)$, the MSE-efficiency of $\hat{\boldsymbol{\theta}}$ is less than 100%. The reduction in efficiency is not negligible, but not very large. When the error distribution is $\Gamma(\lambda_i^{-2}, \lambda_i^2)$ or $LN(-2^{-1}log(1+\lambda_i^2), log(1+\lambda_i^2))$, $\hat{\boldsymbol{\theta}}$ performs significantly better than QMLE. These results show that for the Log ACD-Type II model, $\hat{\boldsymbol{\theta}}$ is better than QMLE overall.

Box-Cox ACD Model (Figure 5):

When $\varepsilon \sim \exp(1)$, the MSE-efficiency of the two semiparametric estimators fell to about 70% for some parameter values. When the error distribution is $\Gamma(\lambda_i^{-2}, \lambda_i^2)$ or $LN(-2^{-1}log(1+\lambda_i^2), log(1+\lambda_i^2))$, $\hat{\boldsymbol{\theta}}$ performs significantly better than QMLE. Overall $\hat{\boldsymbol{\theta}}$ performs better than QMLE.

Linear ACD and Power ACD Models (Figures 6-9):

In these models, $\Theta = \{(\alpha, \beta, \gamma) : \alpha \geq 0, \beta \geq 0, \gamma \geq 0, \gamma + \beta \leq 1\}$. Figures 6 and 7 show that constrained estimator $\bar{\boldsymbol{\theta}}$ performed at least well as the unconstrained DW-estimator $\hat{\boldsymbol{\theta}}$ for all true parameter values and significantly better when the true parameter value is near the boundary of Θ. The cases for which the relative efficiencies are equal to 100% or slightly higher, correspond to the case when the parameter value is away from the boundary and lie well in the interior of the parameter space. Similarly, relative efficiencies that are higher than 100% correspond to the case when the parameter value is close to the boundary. Therefore, as expected, the constrained estimator $\bar{\boldsymbol{\theta}}$ is better than the unconstrained estimator $\hat{\boldsymbol{\theta}}$.

If the true value of $\boldsymbol{\theta}$ is not in the set A, where

$$A = \{(\alpha, \beta, \gamma) : \beta \text{ and } (\beta/\alpha) \text{ are close to zero, and } \alpha \text{ and } \gamma \text{ are not close to zero } \}$$

then $\bar{\boldsymbol{\theta}}$ performs better than QMLE. Even if the true parameter lies in the set A (eg., see the rows marked with '*' in Tables 1 and 2), QMLE does not dominate $\bar{\boldsymbol{\theta}}$ as is clear from Tables 1 and 2 that QMLE is better than $\bar{\boldsymbol{\theta}}$ for (α, γ) but not for β.

In several empirical studies reported in the literature, for example Engle and Russell (1998), Engle and Russell (1997), Fernandes and Grammig (2006) and Zhang *et al.* (2001), the estimated value of $\boldsymbol{\theta}$ turned out to be away from A. Therefore, it appears that $\bar{\boldsymbol{\theta}}$ performs better than QMLE in the part of the parameter space that is of practical relevance.

Finally, let us make the following comment. It may appear surprising that even though errors were generated to have Markov structure, $\hat{\boldsymbol{\theta}}_{Mart}$ performed better than $\hat{\boldsymbol{\theta}}_{Mark}$. A possible explanation is that $\tilde{\ell}_i^*(\boldsymbol{\theta})$ for Martingale error has a much simpler form compared to that for the Markov error. Hence, it is likely that the various Nadaraya-Watson estimators for the conditional moments to compute $\hat{\boldsymbol{\theta}}_{MARK}$ may be not be very good. This together with the fact that the $\hat{\boldsymbol{\theta}}_{Mart}$ is derived under much weaker assumption lead us to recommend $\hat{\boldsymbol{\theta}}_{Mart}$ over $\hat{\boldsymbol{\theta}}_{Mark}$.

Summary of the results:

For Log ACD Types I and II models, the semiparametric DW-estimator $\hat{\boldsymbol{\theta}}$ performed better than the QMLE. For the Box-Cox ACD model, $\hat{\boldsymbol{\theta}}$ appears to be a better estimator overall. For the Linear ACD and Power ACD models, for which α, β and γ must be nonnegative and $\beta + \gamma \leq 1$, the constrained estimator $\bar{\boldsymbol{\theta}}$ performed better than the unconstrained estimator $\hat{\boldsymbol{\theta}}$ and also better than QMLE in the part of the parameter space that appears to be relevant based on past empirical studies.

4 An empirical example

In this section we use the IBM transaction data for November 1990, to illustrate the constrained estimator $\bar{\boldsymbol{\theta}}$. In this example, we do not plan to model the data in order to draw substantive conclusions about IBM transactions, therefore we do not carry out diagnostics to evaluate goodness of fit. Such issues for these data have been discussed in other studies, including Engle and Russell (1998). We estimated the parameters in the linear ACD(2,2) model

$$\psi_i = \alpha + \beta_1 X_{i-1} + \beta_2 X_{i-2} + \gamma_1 \psi_{i-1} + \gamma_2 \psi_{i-2}, \tag{10}$$

by QMLE and the semiparametric methods. The parameter space Θ is given by

$$\Theta = \{\boldsymbol{\theta} : \boldsymbol{\theta} = (\alpha, \beta_1, \beta_2, \gamma_1, \gamma_2)^T; \alpha \geq 0; 0 \leq \beta_1, \beta_2, \gamma_1, \gamma_2, \beta_1 + \beta_2 + \gamma_1 + \gamma_2 \leq 1\}.$$

Table 3: Estimates of parameters for the ACD(2,2) model for the IBM transaction data

	α	β_1	β_2	γ_1	γ_2
	Unconstrained Estimators				
QMLE	0.561	0.098	0.018	0.375	0.492
$\hat{\boldsymbol{\theta}}_{MART}$	0.321	0.108	-0.041	1.005	-0.082
$\hat{\boldsymbol{\theta}}_{MARK}$	0.423	0.108	-0.048	0.984	-0.059
$\hat{\boldsymbol{\theta}}_{IID}$	0.613	0.095	-0.026	0.806	0.103
	Constrained Estimators				
$\hat{\boldsymbol{\theta}}_{MART}$	0.471	0.099	0.000	0.616	0.270
$\hat{\boldsymbol{\theta}}_{MARK}$	0.609	0.096	0.000	0.547	0.336
$\hat{\boldsymbol{\theta}}_{IID}$	0.668	0.088	0.000	0.568	0.320

The computed values are given in Table 3, where $\hat{\boldsymbol{\theta}}_{MART}, \hat{\boldsymbol{\theta}}_{MARK}$ and $\hat{\boldsymbol{\theta}}_{IID}$ are the estimators corresponding to the three cases in (3)- (5). To compute the QMLE, we maximized the log likelihood corresponding to the assumption $\varepsilon_i \sim exp(\lambda)$. Since the unconstrained QMLE, given in Table 3, is an interior point of Θ, it is also equal to the QMLE under the constraint $\boldsymbol{\theta} \in \Theta$.

Although the unconstrained QMLE satisfies the constraint $\boldsymbol{\theta} \in \Theta$, the DW-estimator $\hat{\boldsymbol{\theta}}$ corresponding to Martingale, Markov and iid errors, are outside their allowed ranges. This is an example of the type of settings where a constrained estimator such as $\bar{\boldsymbol{\theta}}$ would be essential. Since constrained estimator $\bar{\boldsymbol{\theta}}$ is not asymptotically normal when the true parameter lies on the boundary of the parameter space, it is not particularly meaningful to provide standard errors for $\bar{\boldsymbol{\theta}}$. If a measure of variability is desired, a confidence region can be constructed by inverting an inequality constrained test based on $\bar{\boldsymbol{\theta}}$. This is not a trivial task, but possible to do.

Note that, based on the semiparametric estimators corresponding to Martingale errors, the estimate of β_2 has now moved from -0.041 to its boundary $\beta_2 = 0$, the estimate of γ_2 moved from $-.082$ to 0.27 a value that is interior to its allowed range, and the estimate of γ_1 moved from 1.005 to 0.616, a value that is also interior to its allowed range. This example illustrates that when there are several estimates that are outside their allowed range, the constrained estimation method introduced in this paper offers a methodologically sound way of constructing an efficient estimator of $\boldsymbol{\theta}$.

5 Conclusion

We studied estimation of parameters in a large class of duration models. Our work is centered around the semiparametrically efficient estimator of Drost and Werker (2004) for situations where the error distribution is unknown and the errors themselves may not be independent either. Since such situations are expected to be common in practice, this semiparametric method of estimation is of significant practical importance. To evaluate this estimator, we carried out a large scale simulation study.

Using the theoretical results of Drost and Werker (2004) as building blocks, we proposed a new semiparametric estimator for duration models for cases when some parameters are known to satisfy inequality constraints, for example nonnegativity constraints as in the standard linear ACD model of Engle and Russell (1998). We showed that our proposed constrained estimator is asymptotically better than the unconstrained DW-estimator when there are inequality constraints on parameters. We carried out a simulation study to compare our estimator with the DW-estimator and the QMLE.

For the Log ACD Models of types I and II and the Box-Cox ACD models, for which there are no inequality constraints on parameters, the DW-estimator performed better than the QMLE overall. For the Linear ACD and Power ACD Models, in which some parameters are known to be nonnegative, the inequality constrained estimator proposed in this paper performed better than the DW-estimator. Further, in these models, the constrained estimator $\bar{\boldsymbol{\theta}}_{Mart}$ performed better than the QMLE in most cases of empirical interest.

In summary, the DW-estimator is better than the QMLE when there are no inequality constraints, such as nonnegativity constraints. If there are inequality constraints, then the constrained estimator proposed in this paper is better.

Since statistical inferences based on the semiparametric estimators $\hat{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\theta}}$ are valid under quite weak and very realistic assumptions, but they perform better than QMLE overall, and only marginally worse than QMLE even under the ideal conditions for QMLE, these semiparametric estimators are serious competitors to QMLE, and the indications are that they are better than QMLE.

Acknowledgments

This research was supported under Australian Research Council's Discovery Projects funding scheme (project number DP0665710). We are grateful to Feike Drost and Bas Werker for providing a copy of their computer program, Heather Anderson and Param Silvapulle for constructive comments and suggestions, and Gunky Kim for help with the use of MATLAB and implementation of the DW-approach.

References

Bauwens, L. and Giot, P. (2000). The logarithmic acd model: An application to the bid-ask quote process of three nyse stocks. *Annales d'Economie et de Statistiquie*, **60**, 117–150.

Bauwens, L. and Giot, P. (2001). *Econometric Modelling of Stock Market Intraday Activity.* Kluwer Academic Publishers.

Bickel, P., Klaassen, C., Ritov, Y., and Wellner, J. (1993). *Efficient and Adaptive Estimation in Semiparametric Models.* The Johns Hopkins University Press, Baltimore, Maryland.

Drost, F. C. and Werker, B. J. M. (2004). Semiparametric duration models. *J. Bus. Econom. Statist.*, **22**(1), 40–50.

Engle, R. F. and Russell, J. R. (1997). Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model. *Journal of Emprical Finance*, **4**, 187–212.

Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, **66**(5), 1127–1162.

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

Newey, W. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics*, **5**(2), 99–135.

Silvapulle, M. J. and Sen, P. K. (2005). *Constrained Statistical Inference: inequality, order, and shape restrictions*. John Wiley and Sons, New York.

Tsiatis, A. A. (2006). *Semiparametric theory and missing data*. Springer Series in Statistics. Springer, New York.

Zhang, M. Y., Russel, J. R., and Tsay, R. S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics*, **104**, 179–207.

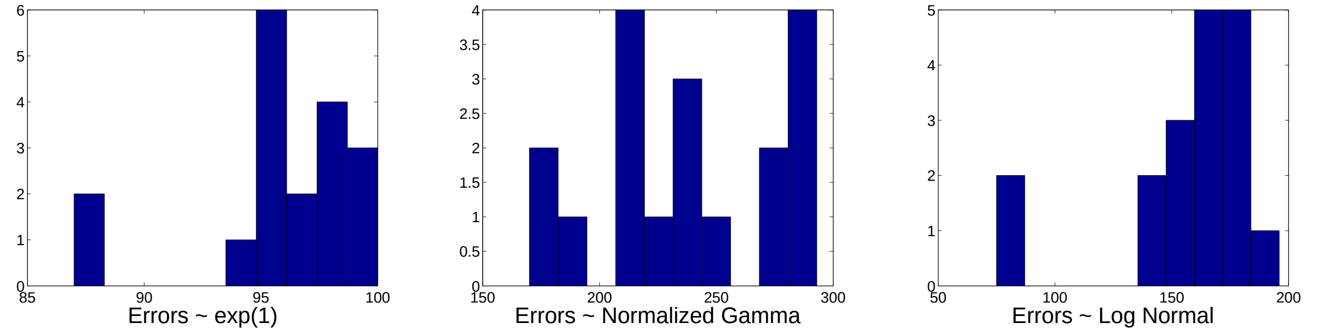

Figure 3: MSE of $\hat{\boldsymbol{\theta}}$ relative to QMLE for the LACD-1 model.

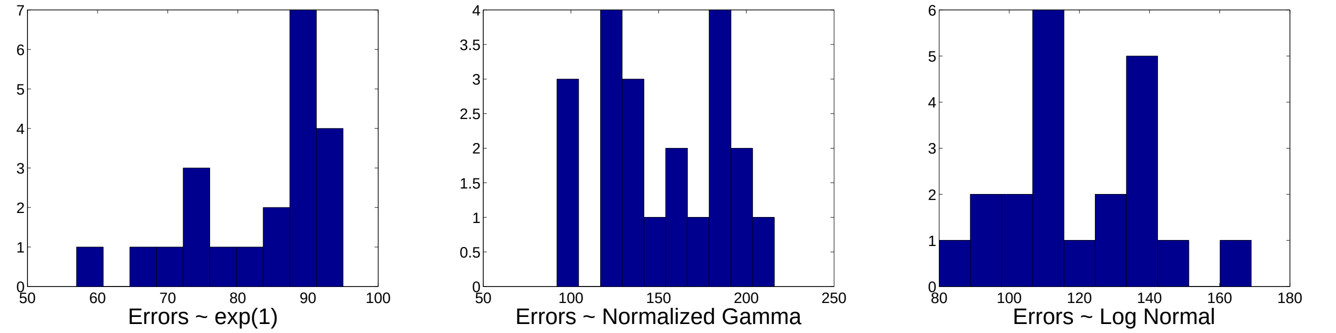

Figure 4: MSE of $\hat{\boldsymbol{\theta}}$ relative to QMLE for the LACD-2 model.

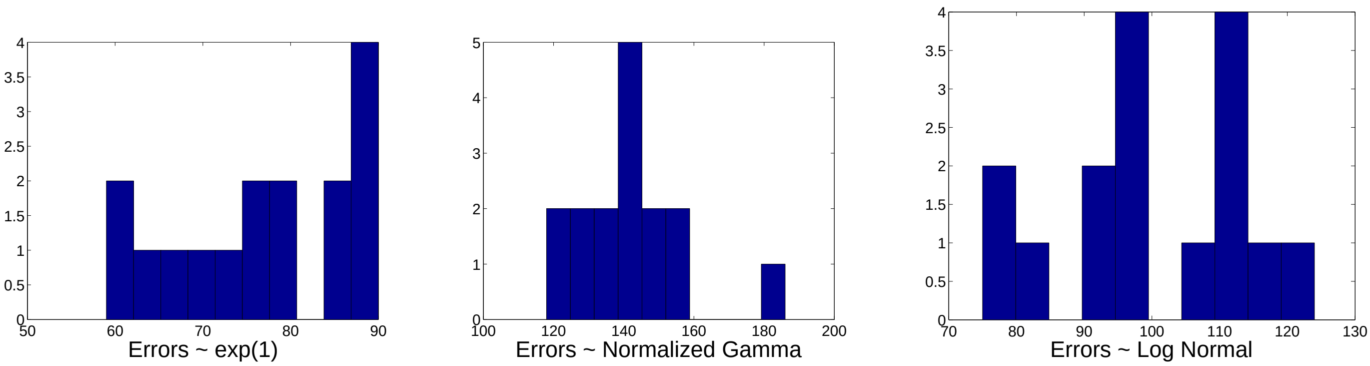

Figure 5: MSE of $\hat{\boldsymbol{\theta}}$ relative to QMLE for the BCACD model

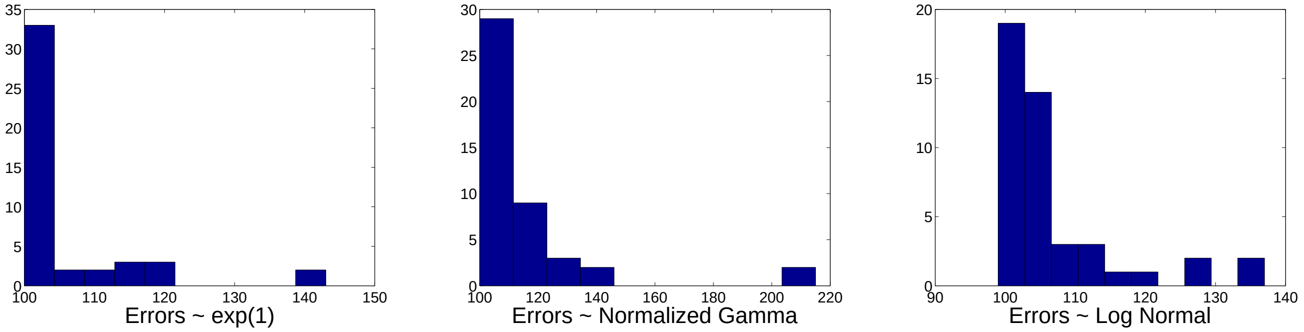

Figure 6: MSE of $\bar{\boldsymbol{\theta}}$ relative to $\hat{\boldsymbol{\theta}}$ for the ACD model.

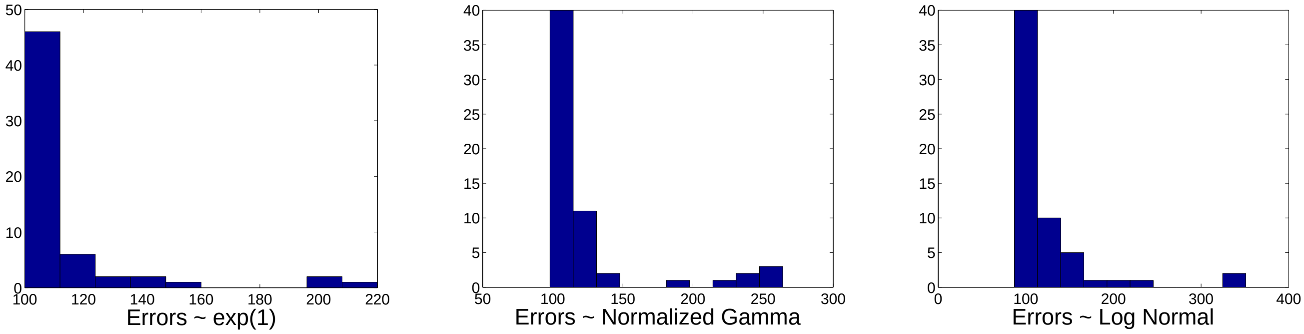

Figure 7: MSE of $\bar{\boldsymbol{\theta}}$ relative to $\hat{\boldsymbol{\theta}}$ for the PACD model.

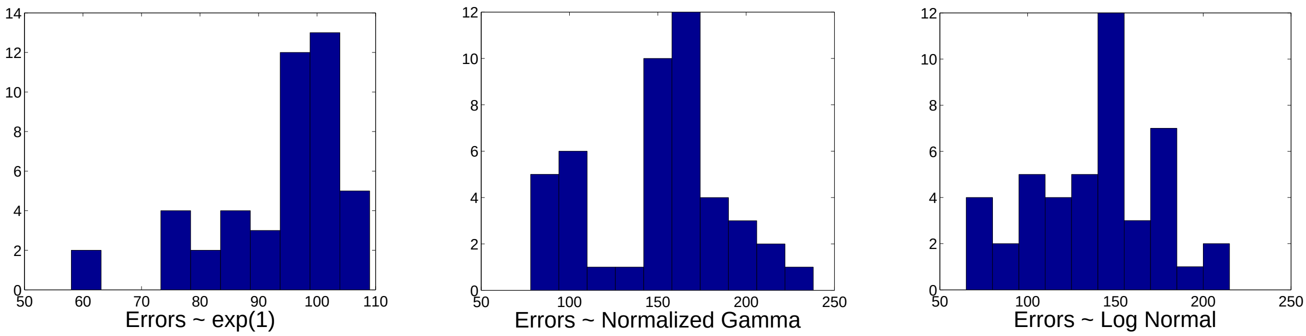

Figure 8: MSE of $\bar{\boldsymbol{\theta}}$ relative to QMLE for the ACD model

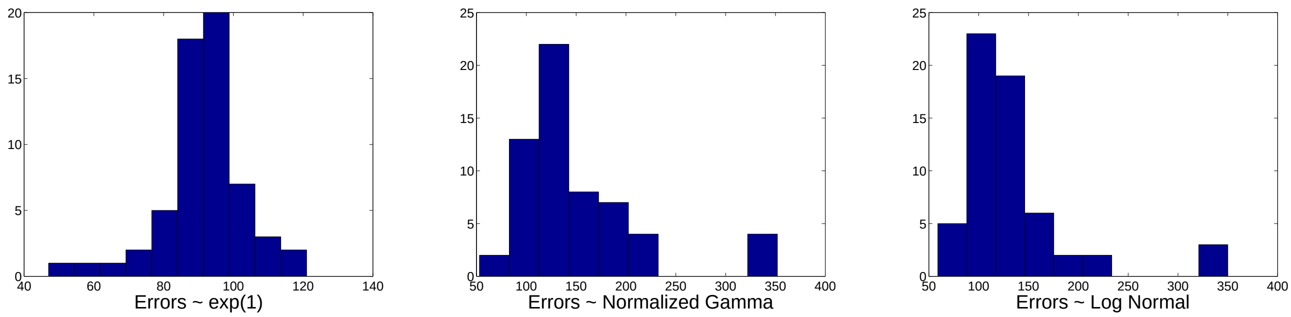

Figure 9: MSE of $\bar{\boldsymbol{\theta}}$ relative to QMLE for the PACD model