ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

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

Kulan Ranasinghe and Mervyn J. Silvapulle

January 2008

Working Paper 1/08

# 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]. This estimator is efficient if the errors are iid and exponentially distributed. Otherwise, it may not be the most efficient. Motivated by this, a class of estimators has been introduced by Drost and Werker (2004). Their estimator is asymptotically most efficient when the error distribution is unknown. However, the practical relevance of their method remains to be evaluated. Further, although some parameters in several common duration models are known to be nonnegative, this estimator may turn out to be negative. This paper addresses these two issues. We propose a new semiparametric estimator when there are inequality constraints on parameters, and a simulation study evaluates the two semiparametric estimators. The results lead us to conclude the following when the error distribution is unknown: (i) If there are no inequality constraints on parameters then the Drost-Werker estimator is better than the QMLE, and (ii) if there are inequality constraints on parameters then the estimator proposed in this paper is better than the Drost-Werker estimator and the QMLE. In conclusion, this paper recommends estimators that are better than the often used QMLE for estimating duration models.

*Key Words:* Adaptive inference; Conditional duration model; Constrained inference; Efficient semiparametric estimation; Order restricted inference; Semiparametric efficiency bound.

*JEL Classification:* C41, C14.

# 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 class of such models forms an essential tool for the study of market microstructure (Bauwens and Giot 2001). The objectives of this paper are to evaluate a recently developed asymptotically efficient method of estimating duration models when the error distribution is unknown, which is almost always the case in practice, and to propose an improvement to the foregoing method when there are inequality constraints on some parameters, for example some 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 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 $\varepsilon_i$ is referred to as the *error term* which is usually standardized so that $E(\varepsilon_i) = 1$. The main objective of duration analysis is to model $\psi_i$ 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, the model may take the form $\psi_i = g(\ldots, X_{i-1}; \ldots, \psi_{i-1}; \boldsymbol{\theta})$ 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)). Since $f$ is usually unknown, the quasi maximum likelihood estimator[QMLE] corresponding to exponential distribution for the error terms, is the standard choice. 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 estimator is *semi-parametrically efficient*, in the sense that it reaches the highest asymptotic efficiency bound. Detailed accounts of this topic are given in Bickel *et al.* (1993) and Tsiatis (2006). While the estimator is asymptotically efficient, a detailed evaluation of the practical relevance of this estimator is not yet available.

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. 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 $\beta \geq 0$ and $\gamma \geq 0$ in (1), 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 $\alpha$, $\beta$ and $\gamma$ in (1) must be nonnegative as well. Further, we also have $\alpha + \beta \leq 1$. However, the DW-estimator does not incorporate such inequality constraints and hence it may turn out to be negative even when the true parameter is known to be nonnegative. If the DW-estimator $\hat{\beta}$ of $\beta$ 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. The literature on statistical inference under inequality constraints, also known as order restrictions, is quite extensive indeed. Some recent relevant references are El Barmi and Mukerjee (2005), El Barmi *et al.* (2006), Peddada *et al.* (2005), Peddada *et al.* (2006), Hwang and Peddada (1994) and Silvapulle and Sen (2005). The latter one provides a comprehensive account on the topic. 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 in section 2.2 provides the asymptotic distribution of our inequality constrained estimator $\bar{\boldsymbol{\theta}}$ and shows that it is 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 summarized as follows:

1. *The errors are iid and their 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] turned out 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:* If the true parameter does not lie in a small particular region of the parameter space, which we shall refer to as $A$, then our proposed estimator is better than the QMLE and the DW-estimator. In several published empirical studies we observed that the estimators were not in the region $A$. Therefore, 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 widely used in practice.

The plan of the paper is as follows. Section 2 discusses the methodological aspects. In subsection 2.1, we recall some known results on efficient semiparametric inference, and in subsection 2.2 we develop the methodological aspects and propose new inequality constrained semiparametric estimators. 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 up to and including the $i^{th}$ observation $X_i$, $\psi_i = E(X_i \mid \mathcal{F}_{i-1})$ 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 in the duration model; for example, $\boldsymbol{\theta} = (\alpha, \beta, \gamma)^\top$ 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 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 $\varepsilon_i$ 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 $\varepsilon_i$ given the past. Now, with $\psi_i = E(X_i \mid \mathcal{F}_{i-1})$, the semiparametric[SP] model is defined formally by

$$X_i = \psi_i \varepsilon_i,\ \psi_i = g(\ldots, X_{i-1}; \ldots, \psi_{i-1}; \boldsymbol{\theta}),\ \text{and } \mathcal{L}(\varepsilon_i \mid \mathcal{F}_{i-1}) = \mathcal{L}(\varepsilon_i \mid \mathcal{H}_{i-1}) \qquad (2)$$

where $g$ is a known function and $\mathcal{L}(\varepsilon_i \mid \mathcal{F}_{i-1})$ refers to the distribution of $\varepsilon_i$ 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 $\{\phi, \Omega\}$.

The next subsection provides the essentials on semiparametric inference, and states the relevant results in a concise form. For convenience, previously known results are discussed in the next subsection and the new methodological developments are given in subsection 2.2

## 2.1 Semi-parametric Estimation

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). The latter book provides the basic technical arguments and concepts in detail. 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 such estimators and their relevance for inference are also given in Newey (1990). In this subsection, we shall state the main relevant results, without the technical details or proofs.

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, $\{\tilde{\boldsymbol{\theta}} + \{n^{-1}\Sigma_{i=1}^{n}\dot{\ell}_i(\tilde{\boldsymbol{\theta}})\dot{\ell}_i(\tilde{\boldsymbol{\theta}})^{\top}\}^{-1}n^{-1}\Sigma_{i=1}^{n}\dot{\ell}_i(\tilde{\boldsymbol{\theta}})\}$, is called the *one-step estimator*. It is asymptotically equivalent to the MLE, and is obtained by applying a Newton-Raphson type 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 obtained by Drost and Werker (2004) 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. When we say that $\tilde{\dot{\ell}}_i^*$ is a "suitable" estimator of $\dot{\ell}_i^*$, it essentially means that the difference between the two converges to zero 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 of the error terms. 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 neighborhood 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 semiparametrically efficient DW-estimator:

$$\hat{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}} + \left(n^{-1}\sum_{i=1}^{n}\tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}})\tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}})^{\top}\right)^{-1} n^{-1}\sum_{i=1}^{n}\tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}}) \tag{3}$$

We close this section with three special cases of the set $\mathcal{H}_i$ in (2) and the corresponding expressions for $\tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}})$. 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:

$$\{\varepsilon_i - 1/var(\varepsilon_i)\}\dot{\psi}_i - \{1 + \varepsilon_i f_i'(\varepsilon_i)/f_i(\varepsilon_i)\}(\partial/\partial\theta)log(\psi_i) - \dot{\psi}_i \tag{4}$$

$$\frac{\varepsilon_i - 1}{var\{\varepsilon_i|\varepsilon_{i-1}\}} E\left[\frac{\partial}{\partial\theta}log(\psi_i)|\varepsilon_{i-1}\right] - \left(1 + \varepsilon_i \frac{f_i'(\varepsilon_i)}{f_i(\varepsilon_i)}\right)\left[\frac{\partial}{\partial\theta}log(\psi_i) - E\left[\frac{\partial}{\partial\theta}log(\psi_i)|\varepsilon_{i-1}\right]\right] \tag{5}$$

$$\{(\varepsilon_i - 1)/\mathrm{var}(\varepsilon_i|\mathcal{H}_i)\}(\partial/\partial\theta)\log(\psi_i) \tag{6}$$

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

The estimator $\hat{\boldsymbol{\theta}}$ in (3) for the foregoing three choices of $\mathcal{H}_{i-1}$, 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 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 $\alpha$, $\beta$ and $\gamma$ are nonnegative because $\psi_i \geq 0$ and $X_i \geq 0$ for every $i$. However, their estimators in (3) 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 results from constrained statistical inference (Silvapulle and Sen (2005)). There is no unique way to define suitable constrained estimators. We propose the following.

Let $\Theta$ denote the parameter space of $\boldsymbol{\theta}$. We shall assume that $\Theta$ is convex. Some of the results presented here would hold even if $\Theta$ is not convex, but is *Chernoff Regular*; discussions on Chernoff Regularity may be found in Geyer (1994), Silvapulle and Sen (2005)) and Shapiro (2000). However, we will not consider such general shapes for $\Theta$ here. For the linear ACD(1,1) model in (1), we have $\Theta = \{\boldsymbol{\theta} : \boldsymbol{\theta} = (\alpha, \beta, \gamma)^\top, \alpha \geq 0, \beta \geq 0, \gamma \geq 0, \beta + \gamma \leq 1\}$, 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 $V$ where $\hat{\boldsymbol{\theta}}$ is the DW-estimator. 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 $(-1/2)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top V^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ and hence the corresponding MLE of $\boldsymbol{\theta}_0$ is

$$\bar{\boldsymbol{\theta}}^* = \arg\min_{\boldsymbol{\theta}\in\Theta}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top V^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}). \tag{7}$$

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

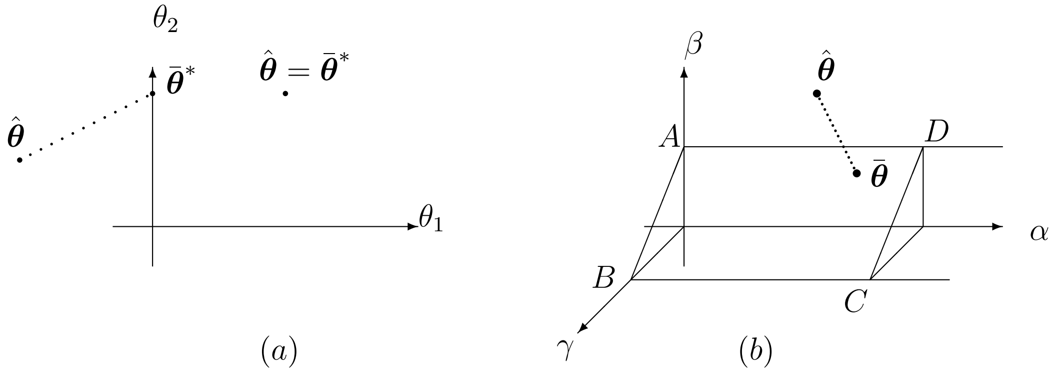


Figure 1: (a) The unconstrained estimator $\hat{\boldsymbol{\theta}}$ and the constrained estimator $\bar{\boldsymbol{\theta}}^*$ of $\boldsymbol{\theta}_0$ subject to $\boldsymbol{\theta} \in \Theta = \{(\theta_1, \theta_2) : \theta_1 \geqslant 0, \theta_2 \geqslant 0\}$, when $V = (1, 0.5 \mid 0.5,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

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 the definition of $\bar{\boldsymbol{\theta}}^*$, a natural constrained semiparametric estimator is

$$\bar{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Theta}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top W_n^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \tag{8}$$

where $W_n$ is positive definite. In general, we would choose $W_n$ to 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. In general, we would choose $W_n = \left(n^{-1}\sum_{i=1}^n \tilde{\tilde{\ell}}_i^*(\tilde{\boldsymbol{\theta}})\tilde{\tilde{\ell}}_i^*(\tilde{\boldsymbol{\theta}})^\top\right)^{-1}$. If the true value $\theta_0$ is a point on the boundary of the parameter space, it will be seen that the forgoing constrained estimator $\bar{\boldsymbol{\theta}}$ in (8) is an

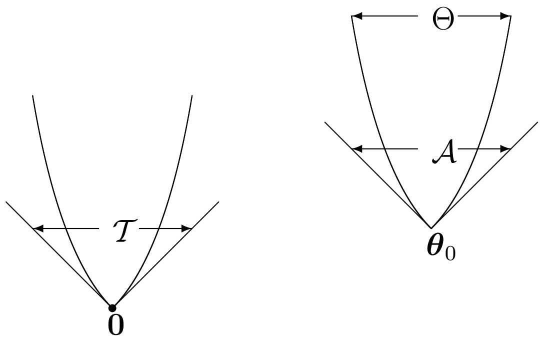


Figure 2: The *tangent cone,* $\mathcal{T}$, and the *approximating cone,* $\mathcal{A}$, of $\Theta$ at $B$.

improvement over the DW-estimator for finite samples and asymptotically. On the other hand, if the true value $\theta_0$ is an interior point, then the constrained estimator is better than the DW-estimator in finite samples, but they are equivalent asymptotically.

Now, to discuss the theoretical results on $\bar{\boldsymbol{\theta}}$, let $\mathcal{T}(\Theta; \boldsymbol{\theta}_0)$ denote the *tangent cone* of $\Theta$ at $\boldsymbol{\theta}_0$. For completeness, its definition is stated here:

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

For more details and references, see Silvapulle and Sen (2005). Intuitively, the tangent cone $\mathcal{T}(\Theta; \boldsymbol{\theta}_0)$ is constructed as follows: First, approximate the boundaries of $\Theta$ at $\boldsymbol{\theta}_0$ by tangents, and then approximate $\Theta$ by the cone formed by these tangents. This is called the *approximating cone* of $\Theta$ 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. 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 $\|\boldsymbol{x}\|_W = \{\boldsymbol{x}^\top W^{-1} \boldsymbol{x}\}^{1/2}$ and $\Pi_W\{\boldsymbol{z} \mid \mathcal{C}\} = \arg\min_{\boldsymbol{\theta} \in \mathcal{C}} \|\boldsymbol{z} - \boldsymbol{\theta}\|_W$. Thus, $\Pi_W\{\boldsymbol{z} \mid \mathcal{C}\}$ denotes the projection of $\boldsymbol{z}$ onto $\mathcal{C}$ with respect to the inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_W = \boldsymbol{x}^\top W^{-1} \boldsymbol{y}$. A simple illustration of $\Pi(\hat{\boldsymbol{\theta}} \mid \mathcal{C})$, which is equal to $\bar{\boldsymbol{\theta}}^*$, 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* $\Theta$ *is convex and that* $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \boldsymbol{Z}$ *where* $\boldsymbol{Z} \sim N(\boldsymbol{0}, V)$ *for some positive definite matrix* $V$ *and* $W_n \xrightarrow{p} W$ *where* $W$ *and* $W_n$ *are positive definite.*

*Then*

$$n^{1/2}(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \Pi_W\{\boldsymbol{Z} \mid \mathcal{T}(\Theta; \boldsymbol{\theta}_0)\} \tag{9}$$

*where* $\bar{\boldsymbol{\theta}}$ *is the constrained estimator defined in (8). Further,* $\bar{\boldsymbol{\theta}}$ *is closer to the true value* $\boldsymbol{\theta}_0$ *than* $\hat{\boldsymbol{\theta}}$ *in the following sense:*

$pr\{\|\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{W_n} \leq \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{W_n}\} = 1$*,* $pr(\bar{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}) \to pr(\boldsymbol{Z} \in \mathcal{T}(\Theta; \boldsymbol{\theta}_0)$*, and*

$$pr\{\|\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{W_n} \lneqq \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{W_n}\} \to pr(\boldsymbol{Z} \notin \mathcal{T}(\Theta; \boldsymbol{\theta}_0). \tag{10}$$

*Proof. Main steps only:* The technical details of the proof of (9) uses the result that the parameter space $\Theta$ 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 $\Theta$ and onto the approximating cone $\mathcal{A}(\boldsymbol{\theta}_0)$ of $\Theta$ at $\theta_0$ 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 $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \boldsymbol{x}^\top W_n^{-1}\boldsymbol{y}$, we have that $(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top W_n^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq 0$. 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 (10) follows. □

In the rest of this subsection, we shall comment on other possible alternatives to the foregoing approach. The general approach to constructing a constrained estimator exploits the fact that one needs to use only the local behavior of the objective function in an $n^{-1/2}$-neighborhood of the true value $\boldsymbol{\theta}_0$. The foregoing $\bar{\boldsymbol{\theta}}$ adopts this approach. It is also possible to construct other similar estimators. For example, another estimator may be defined as $\hat{\boldsymbol{\theta}}(\lambda_0)$ where $\hat{\boldsymbol{\theta}}(\lambda) = \big[\tilde{\boldsymbol{\theta}} + \lambda\big(n^{-1}\sum_{i=1}^n \tilde{\ell}_i^*(\tilde{\boldsymbol{\theta}})\tilde{\ell}_i^*(\tilde{\boldsymbol{\theta}})^\top\big)^{-1} n^{-1}\sum_{i=1}^n \tilde{\ell}_i^*(\tilde{\boldsymbol{\theta}})\big]$ for $0 \leq \lambda \leq 1$ and $\lambda_0$

is the maximum value of $\lambda$ in $[0,1]$ for which $\hat{\boldsymbol{\theta}}(\lambda)$ lies in $\Theta$. This says that the one-step iteration in (3) moves from $\tilde{\boldsymbol{\theta}}$ in the direction suggested by the DW-estimator but stops before crossing the boundary of $\Theta$.

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

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

which may be seen as a pseudo likelihood with score function $n^{-1}\sum_{i=1}^{n} \tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}})$ and information $\big(n^{-1}\sum_{i=1}^{n} \tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}})\tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}})^\top\big)$. 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. This estimator turns out to the same as $\bar{\boldsymbol{\theta}}$ in (8) if the $W_n$ in (8) is equal to $\Big(n^{-1}\sum_{i=1}^{n} \tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}})\tilde{\dot{\ell}}_i^*(\tilde{\boldsymbol{\theta}})^\top\Big)^{-1}$.

Finally, let us note that different choices for $W_n$ in (8) would result in different constrained estimators that may not be asymptotically equivalent. As an example, let $\bar{\boldsymbol{\theta}}^{(a)}$ and $\bar{\boldsymbol{\theta}}^{(b)}$ be the estimator in (8) with $W_n = A_n$ and $W_n = B_n$ respectively, where $A_n \xrightarrow{p} A$ and $B_n \xrightarrow{p} B$. Now, it may be shown that $n^{1/2}(\bar{\boldsymbol{\theta}}^{(a)} - \bar{\boldsymbol{\theta}}^{(b)}) \xrightarrow{d} \boldsymbol{Z}^*$ where $\boldsymbol{Z}^* = [\Pi_A\{\boldsymbol{Z} \mid \mathcal{T}(\Theta;\boldsymbol{\theta}_0)\} - \Pi_B\{\boldsymbol{Z} \mid \mathcal{T}(\Theta;\boldsymbol{\theta}_0)\}] \neq 0$ and $\boldsymbol{Z} \sim N(\mathbf{0}, V)$. Therefore, if $A \neq B$ then $\boldsymbol{Z}^* \neq 0$ and hence $\bar{\boldsymbol{\theta}}^{(a)}$ and $\bar{\boldsymbol{\theta}}^{(b)}$ are not asymptotically equivalent.

# 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 $\lambda_i$ 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.

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. 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 $\hat{\boldsymbol{\theta}}_{Mart}$ defined in (4) - (6). We observed that $\hat{\boldsymbol{\theta}}_{Mart}$ performed better. Further, sconce $\hat{\boldsymbol{\theta}}_{Mart}$ is based on the least amount of

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

| True value | | | $\varepsilon \sim EXP$ | | | $\varepsilon \sim NG$ | | | $\varepsilon \sim LN$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha_0$ | $\beta_0$ | $\gamma_0$ | $\alpha$ | $\beta$ | $\gamma$ | $\alpha$ | $\beta$ | $\gamma$ | $\alpha$ | $\beta$ | $\gamma$ |
| 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 $\theta_i$ is defined as MSE(QMLE)/MSE($\bar{\boldsymbol{\theta}}$).

assumptions, we shall present the results for $\hat{\boldsymbol{\theta}}_{Mart}$ only in the rest of this section, and write $\hat{\boldsymbol{\theta}}$ for $\hat{\boldsymbol{\theta}}_{Mart}$. The results for the other estimators are available in a working paper.

The histograms of relative MSE of $\hat{\boldsymbol{\theta}}$ are shown in Figures 3 - 8. Each figure has three diagrams: the one on left, middle and right correspond to $\varepsilon_i$ 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 3 - 8 show that, as expected, the QMLE performed at least as well as the semiparametric estimator.

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

| | $\varepsilon \sim EXP$ | | | | $\varepsilon \sim NG$ | | | | $\varepsilon \sim LN$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\boldsymbol{\theta}_0 = (\alpha_0, \beta_0, \gamma_0, \lambda_0)$ | $\alpha$ | $\beta$ | $\gamma$ | $\lambda$ | $\alpha$ | $\beta$ | $\gamma$ | $\lambda$ | $\alpha$ | $\beta$ | $\gamma$ | $\lambda$ |
| 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 $\theta_i$ is defined as MSE(QMLE)/MSE($\bar{\theta}_i$).

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 the 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 $\Theta$. 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 deep 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 $A$, (for eg., see the rows marked with '*' in Tables 1 and 2) 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 $A$ QMLE does not dominate $\bar{\boldsymbol{\theta}}$; Tables 1 and 2 show that, in region A, QMLE is better than $\bar{\boldsymbol{\theta}}$ for $(\alpha, \gamma)$ but not for $\beta$.

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.

*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 $\alpha$, $\beta$ and $\gamma$ must be nonnegative and $\beta + \gamma \leq 1$, the constrained estimator $\bar{\boldsymbol{\theta}}$ performed better than the unconstrained DW-estimator $\hat{\boldsymbol{\theta}}$ and also better than the 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 importance of 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, and 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

$$\psi_i = \alpha + \beta_1 X_{i-1} + \beta_2 X_{i-2} + \gamma_1 \psi_{i-1} + \gamma_2 \psi_{i-2}, \tag{11}$$

by QMLE and the semiparametric methods. The parameter space $\Theta$ is given by

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

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 (4)- (6). 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 $\Theta$, it is also equal to the QMLE under the constraint $\boldsymbol{\theta} \in \Theta$.

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

| | $\alpha$ | $\beta_1$ | $\beta_2$ | $\gamma_1$ | $\gamma_2$ |
|---|---|---|---|---|---|
| | 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 |

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 the 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 computational task, but possible to do. In any case, it is important to note that the constrained estimator is closer to the true value than the unconstrained estimator $\hat{\boldsymbol{\theta}}$.

Note that, based on the semiparametric estimators corresponding to Martingale errors, the estimate of $\beta_2$ has now moved from $-0.041$ to its boundary $\beta_2 = 0$, the estimate of $\gamma_2$ moved from $-.082$ to 0.27 a value that is interior to its allowed range, and the estimate of $\gamma_1$ moved from 1.005 to 0.616, a value that is also interior to its allowed range. This example illustrates that when some or all components of $\hat{\boldsymbol{\theta}}$ lie 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. Since such situations are expected to be common in practice, this semiparametric method of estimation is of significant practical importance.

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. An empirical application involving the ACD(2,2) model illustrates the relevance, importance and the ease with which $\bar{\boldsymbol{\theta}}_{Mart}$ can be used.

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.

## 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 helpful comments and suggestions and for providing a copy of their computer program, 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.

El Barmi, H. and Mukerjee, H. (2005). Inference under a stochastic ordering constraint: the k-sample case. *Journal of the American Statistical Association*, **100**, 252–261.

El Barmi, H., Kochar, S., and Tsimikas, J. (2006). Likelihood ratio tests for and against ordering of the cumulative incidence functions in multiple competing risks and discrete mark variable models. *Journal of Statistical Planning and Inference*, **136**(5), 1588–1607.

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.

Geyer, C. J. (1994). On the asymptotics of constrained $M$-estimation. *Annals of Statistics*, **22**, 1993–2010.

Hwang, J. T. G. and Peddada, S. D. (1994). Confidence interval estimation subject to order restrictions. *Annals of Statistics*, **22**, 67–93.

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

Peddada, S. D., Dunson, D. B., and Tan, X. (2005). Estimation of order-restricted means from correlated data. *Biometrika*, **92**, 703–715.

Peddada, S. D., Haseman, J., Tan, X., and Travlos, G. (2006). Tests for simple tree order restriction with application to dose-response studies. *Jour. Royal Statist. Soc., Ser - C*, **55**, 493–506.

Shapiro, A. (2000). On the asymptotics of constrained local $M$-estimators. *Annals of Statistics*, **28**(3), 948–960.

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.

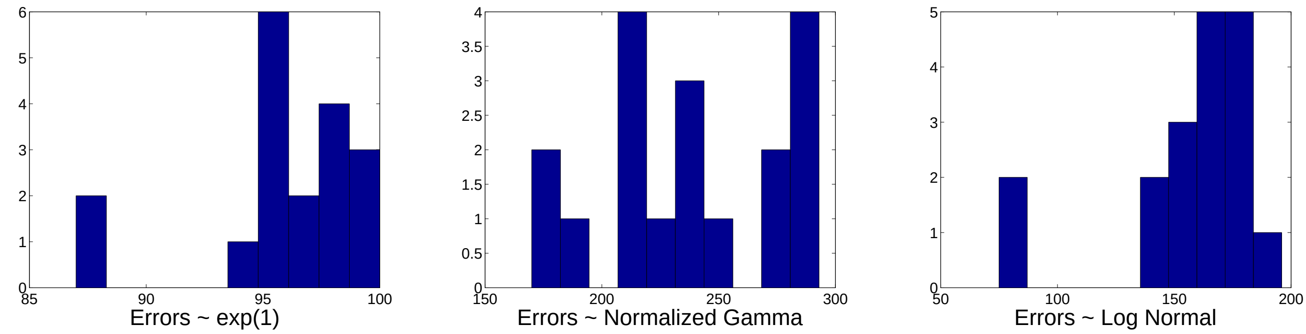


Figure 3: MSE of $\hat{\boldsymbol{\theta}}$ relative to QMLE for the LACD-1 model.

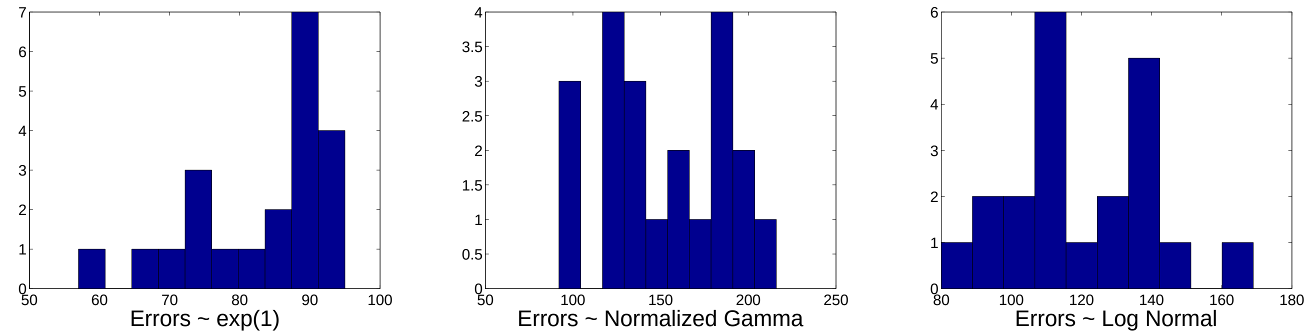


Figure 4: MSE of $\hat{\boldsymbol{\theta}}$ relative to QMLE for the LACD-2 model.

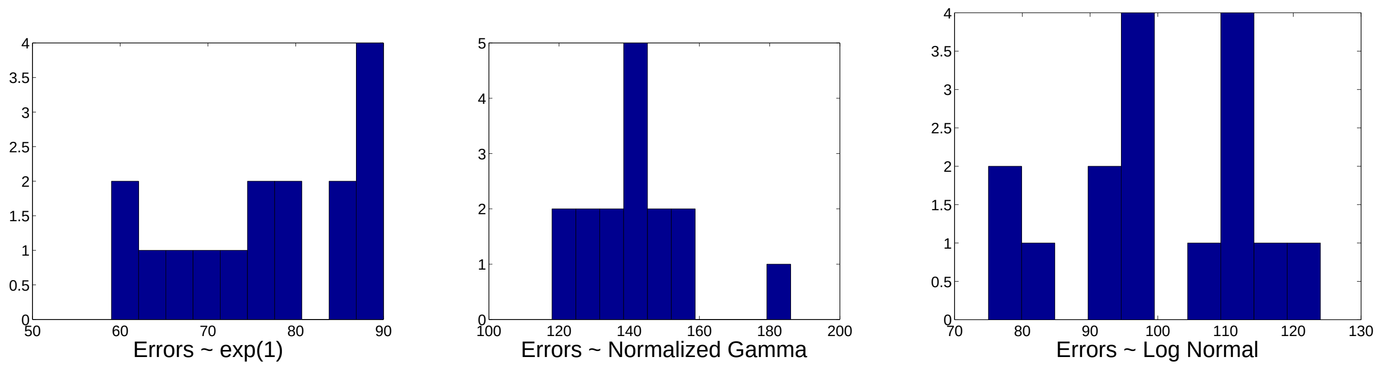


Figure 5: MSE of $\hat{\boldsymbol{\theta}}$ relative to QMLE for the BCACD model

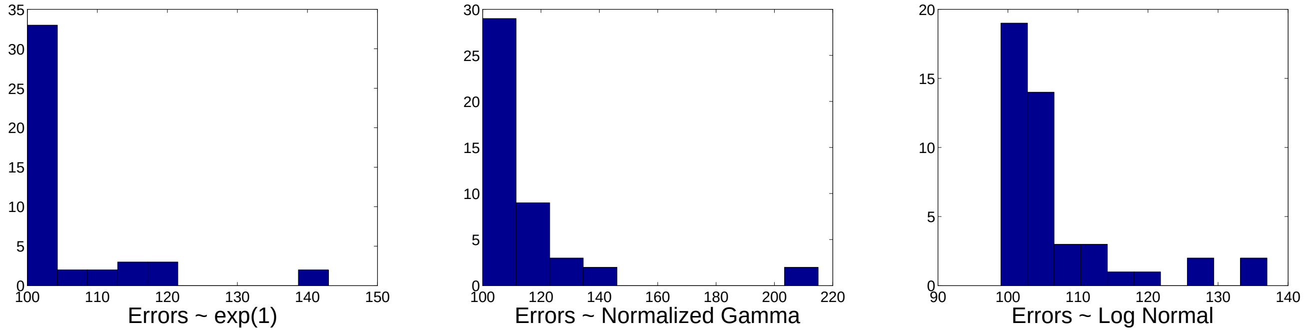


Figure 6: MSE of $\bar{\boldsymbol{\theta}}$ relative to $\hat{\boldsymbol{\theta}}$ for the ACD model.

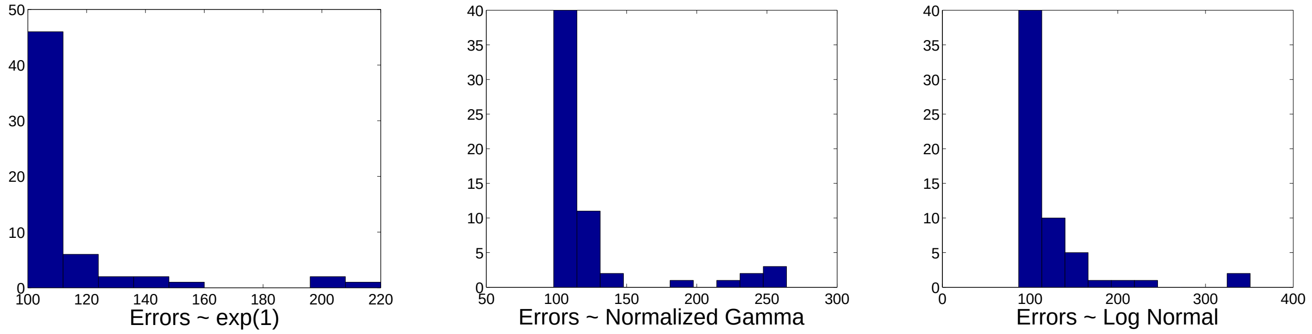


Figure 7: MSE of $\bar{\boldsymbol{\theta}}$ relative to $\hat{\boldsymbol{\theta}}$ for the PACD model.

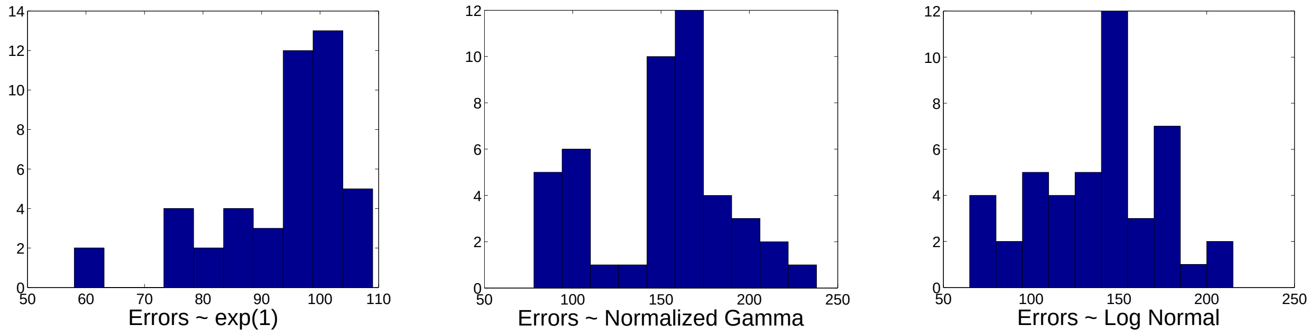


Figure 8: MSE of $\bar{\boldsymbol{\theta}}$ relative to QMLE for the ACD model

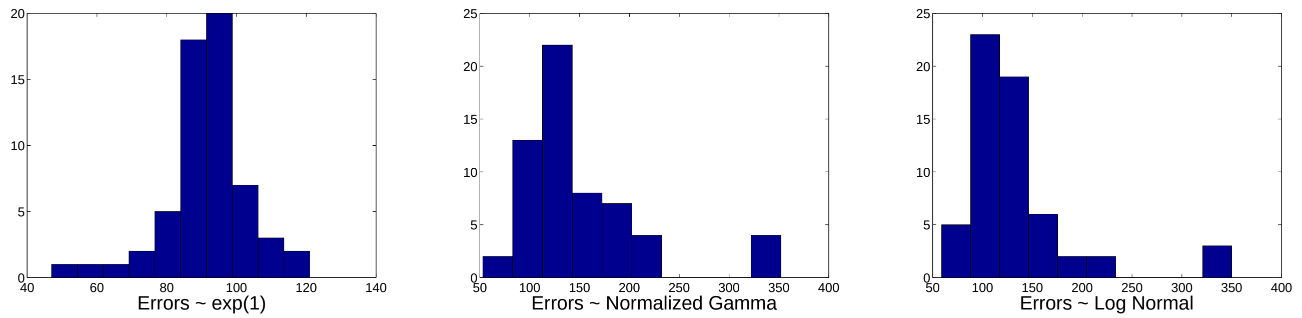


Figure 9: MSE of $\bar{\boldsymbol{\theta}}$ relative to QMLE for the PACD model