UNIVERSITE CATHOLIQUE DE LOUVAIN
DEPARTEMENT DE SCIENCES ECONOMIQUES

Essays in the econometrics of dynamic duration models with application to tick by tick financial data

Thèse presenté par:
Fausto Galli

Jury de thèse:

Prof. Luc Bauwens	promoteur
Prof. Joachim Grammig	
Prof. Olivier Scaillet	
Prof. Fatemeh Shadman-Mehta	
Prof. David Veredas	

Septembre 2009

Contents

To the memory of my father

Acknowledgements

This thesis would not have been possible without the contribution of two persons that with trust and patience supported, guided and advised me in an great number of occasions and of ways. Therefore, my special gratitude goes to Luc Bauwens and Antonio Cosma.

Moreover, many people contributed to this thesis in various forms. Some helped me with technical problems I could not find a way out from, some helped me on a more personal basis, encouraging and often tolerating me with patience. So here is a list, in strict alphabetic order, of the elements of the intersection between the set of these helpful people to whom my gratitude goes and the feasibility set of my memory: Alba Cappelletti, Ana, Andrea, Andreas, Anna, Arie, Athena, Augustin, Bernadette, Carmen, Caroline, Catarina, Catherine, Chiara, Christian, Christophe, Claudio, Daniela, David, Delia, Devi, Dimitri, Eloisa, Fabio, Filippo, Francisco, Fulvio, Genaro, Gilda, Giovanni Barone-Adesi, Giuseppe, Guy, Ilaria, Irene Gijbels, Julian, Ke-Li, Luca, Luciano, Mady, Magda, Malika, Manu, Maurice Mouchart, Mustafa, Oana, Osvaldo, Patrick, Pascale, Pierre, Sarah, Sebastien, Sergio, Sheila, Simona, Susana, Valeria, Veronique.

Chapter 1

Tick-by-tick data and ACD models

1.1 Introduction

The introduction of automated electronic systems of trading allows the keeping of precise records of a vector of characteristics of all the quotes posted and all the trades executed in a regulated market[1]. The availability of such an exhaustive set of data is fairly recent, dating back to the early nineties. Its introduction has stimulated the development of a lively field of econometric analysis, focusing on the trading mechanisms, the intraday characteristics of the markets and the price formation process. The data used in this work are extracted from the Trade And Quote (TAQ) database, which provides intraday information on prices and quotes for stocks traded on the NYSE and on the NASDAQ regulated markets. It is worth remarking that these intraday databases are now available for a large number of exchanges and that in the recent literature, the analytical framework employed here has been extended to tick-by-tick data from regulated markets of various countries.

The availability of a complete set of information about the exact (down to the second or even less) timing of trades and quotes allows the consideration of the duration between two financial events as a random variable, that can be studied by employing the tools of duration analysis. Different kinds of financial events can be defined, such as trade and quote formation, price evolution and accumulation of traded volume, in order to study the process followed by their durations.

Once these financial events have been defined and their durations have been computed from the tick-by-tick database, it is possible to conduct a statistical analysis of their characteristics. The main tools that have been used in the literature so far have been the one of

[1]In the literature, different terms have been used to indicate this kind of data sets. Some authors call them "high frequency data", others "ultra high frequency data", others "tick-by-tick data". Throughout this work, the last denomination will be mostly used.

transition and of time series analysis. Most of the works on the duration of tick-by-tick data are in fact characterized by the direct specification of the density of the duration between events augmented by the introduction of an autoregressive component. In principle, other frameworks of analysis are possible (which can employ, for instance, a perspective based on counting or on intensity) and are being explored by the literature. In this work, though, these last methods will not be considered, even if their development is very interesting. Instead, the work concentrates on the duration density specification.

The statistical models used to describe the point process followed by events defined on the basis of tick-by-tick data can also be seen as a part of a larger analytical framework, whose aim is to jointly describe various dynamics, such as, for instance, price evolution, liquidity or strategic interaction of different agents.

1.2 Tick-by-tick financial data

The data we employ in the essays in the following chapters have a great deal of common features. They consist in tick-by-tick observations from the Trade and Quote (TAQ) database produced by the New York Stock Exchange (NYSE). The data are first filtered to eliminate possible errors in recording or minor discretization problems (durations of zero seconds) and then aggregated to compute trade, price and volume durations, after controlling for daily and weekly seasonality.

A precise account of how these data are treated is provided by Bauwens and Giot (2001), in the next subsections we only provide a quick introduction of how these steps are performed.

1.2.1 Common features of Tick-by-Tick data

A first common characteristic of tick-by-tick data is that they consist in *all* trades executed and quotes posted in a specific market, that is for a specific traded asset in a specific exchange. Tick-by-tick data therefore represent the richest possible set of information about the trading activity of an asset, with the possible exception of the complete list of all book entries and their variations (which would contain more information about the behavior of offer and demand).

Given that trade completions and quote postings happen irregularly, tick-by-tick are not uniformly spaced in time. This is probably the most significant difference with the data that are most commonly used in financial econometrics, where equally spaced observations (daily, weekly, monthly...) are considered. The effect of irregular timing of the data on the scope and methodology of statistical analysis are of course numerous. In this thesis we will concentrate on the consideration that the timing itself of trades and quote postings can now be object of statistical analysis using as a starting point the tools provided by the theory of point processes. This approach has a long tradition in statistics, and we suggest as an introduction to the subject the manuals of Lancaster (1990) and Cox and Isham (1980) A further step could be to consider the new information provided by timing in an analysis of other processes, like return volatility or trading volume dynamics, but we will not explore these problems further here, leaving these subjects to further research.

It is important to notice that in the dataset used in the next chapters of this thesis (composed of observations on stocks traded on NYSE and AMEX), the timing of trades and quote postings is recorded with a rounding of a second. This degree of precision is usually enough to observe and detail the behavior of the point process under analysis, but can sometimes lead to attributing the same time of several trades/quotes to the same second, even if they are separated by some fractions. This can especially happen in periods of fast market for heavily traded stocks and unfortunately in these cases some aggregation is necessary to avoid durations of zero.

Before moving to the statistical analysis of the data, it is important to introduce some relevant features of durations between trades and/or quote postings in tick-by-tick data. These feature appear to be common to durations of various stocks and exchanges and are denoted therefore as "stylized facts" (in pretty much the same way common features of return volatility have been singled out and studied in ARCH literature).

The main stylized facts of durations in financial tick-by-tick data can be considered to be the following:

- **Autocorrelation**. There is an extremely clear evidence of the presence of significant positive autocorrelation in the process of durations. This rules out the possibility of limiting ourselves to the basic Poisson or any iid process (even if the density of the durations is richer than the negative exponential of the Poisson case).
- **Long memory**. The pattern of autocorrelations seems to have a slow rate of decay. The persistence of (relatively low but significant) autocorrelations even at high orders is denoted as "long memory".
- **Overdispersion**. This stylized fact can be somehow considered as the counterpart in the realm of durations of the presence of fat tails in the distribution of returns. The observed ratio between standard deviation to the mean (dispersion index) is larger than the benchmark value of one implied by an exponential distribution.
- **Nonlinearities**. There is some evidence of switching regimes, characterized by transitions in the mean and variance of durations.
- **Seasonality**. Durations appear to be strongly dependent on seasonal factors such as the time of the day and in less marked way, also the day of the week. A typical inverted "U" shape is observed in the means of durations, with more frequent trading

happening after the opening and before the closing of trading sessions, while times between trades tend to widen in the middle of the trading day. Informational issues could be suggested as an explanation of this phenomenon (new information needs to be incorporated at opening and discounted before closing), but we will not deal with a thorough analysis of this issue and simply will consider data deseasonalized by a nonparametric regression of durations on time.

1.2.2 The TAQ database

The data we used for the empirical parts of the following chapters are drawn from the Trade and Quote (TAQ) database, which provides the full record of transactions for stocks on NYSE and NASDAQ markets.

The database has information both about actual trades (informations about prices and size) and about best bid-ask posted quoted, with details on price, size and of course the time of the trade or the bid-ask. Unfortunately, no further data on the book is available at least in the version we used. Another piece of information that is not offered is the one about the direction of the trading and the regime of a trade (large trades have a special procedure, called "upstairs trading").

A further characteristic of these data that must be remarked is that a degree of discretization in the recorded traded time is present. The precision is in fact down to a second which in the case of heavily traded stocks, could lead to several trades sharing the same timestamp. In treating the data we chose to merge these trades, therefore eliminating zero durations.

1.2.3 Durations

Once the raw data are treated in order to deal with the discretization effect and with other technical issues like recording errors or trades registered out of the official opening hours of the market (9:30-16:30 for our data), we can aggregate the quotes and the trades information on the basis of the time of their execution or posting and define other kind of durations. In the next chapters we will in particular use three types of durations: trade, price and volume durations.

- **Trade durations** are defined as the time between two consecutive trades. The presence of discretization could lead to durations of zero, which are eliminated.
- **Price durations** are defined as the time necessary to the price to abandon a predefined range. To compute them we select a price threshold, say c_p and we record that it took to the price to get out of the interval $[X_p \pm c_p]$ where X_p is the price

corresponding to the end of the previous price duration. In order to avoid the effect of bid-ask bounce, we usually use the average between the best bid and ask quotes rather than the realized price.

- An analogous approach is used to define and compute **volume durations**. This time we consider the volume of trades and register the time needed to accumulate in excess of a predefined volume range. Volume durations are less sensitive to bid-ask spread, so they are computed on the basis of realized trades.

1.2.4 The issue of seasonality

An important feature of the trade and quote data is a marked seasonal pattern observed both on a daily basis and inside the trading day.

Intraday seasonality tends to manifest itself with a stronger trading activity after the opening and before the closing of the market. This results in an "inverse-U" shape in both trade, price and volume durations, which are longer in the central hours of the day. Explanations of this phenomenon vary, here we simply mention the possibility that lunch affects the presence of operators and the idea that in the morning the information that reached the market overnight is factored in and that traders often balance or close their positions towards the end of the day.

A similar, though definitely less marked, effect is observable on Mondays and Fridays, which usually register a highly trading activity than days in the middle of the week.

Bid-ask spreads too appear to display a similar seasonal behavior. They in fact result larger in times of fast trading while they restrict in the middle of the day and of the week.

In order to take into account the effect of seasonality, we compute adjusted durations x_i by dividing the raw durations X_i by a seasonality index $\phi(t_i j)$ which represents the time-of-day or the time-of-week effect at time t_i:

$$x_i = X_i/\phi(t_i j). \tag{1.1}$$

The seasonality index $\phi(t_i j)$ is computed in two slightly different ways in this thesis.

- In chapters 2 and 4, we used the average-and-spline method, which consists in calculating the average duration over 30 minutes intervals for each day of the week and joining their midpoints with cubic splines.

- In chapter 3, data are deseasonalized by performing a Nadaraya-Watson estimation with a quartic kernel and a bandwidth computed with normal reference.

An example comparison of the values of the seasonality indices calculated by kernel regression and average-and-spline method for the same data is presented in figure 1.1. The kernel based methods seems to provide smoother estimates. However, the general shape and the magnitude of the seasonal effect are strictly comparable between the two approaches.

1.3 ACD family models

The family of parametric extensions of the original ACD model has been steadily growing since its introduction. In this subsection, some of the models that are more relevant or more useful in order to read the rest of the work will be introduced.

1.3.1 Autocorrelated Conditional Duration

The ACD model, introduced by Engle and Russell (1998a) belongs to the class of accelerated time (also called time deformation) models of the form

$$x = x_0 \, g(z, \beta), \tag{1.2}$$

where durations x are specified as the product of a baseline duration x_0 and a function $g(z, \beta)$ of exogenous variables z and a vector of parameters β. If the expectation $E(x_0) = 1$ and x is independent on $g(z, \beta)$, the second term can be seen as the conditional expectation $E(x|z)$ of x given z.

In the ACD framework, the i-th duration between two financial events, can be modeled as the product

$$x_i = \Psi_i \epsilon_i, \tag{1.3}$$

of a baseline duration ϵ_i, positive and independent and identically distributed[2] and the time deformation factor Ψ_t. The baseline duration ϵ_t is usually parametrized in order to have an unconditional mean of 1, so that the conditional duration is $E(x|I_i) = \Psi_i$.

In the original ACD specification by Engle and Russell (1998a) the conditional duration Ψ_i is specified as an autoregressive model

$$\Psi_i = \omega + \alpha x_{i-1} + \beta \Psi_{i-1} \tag{1.4}$$

with the constraints $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ (with $\beta = 0$ if $\alpha = 0$), to ensure the positivity of the conditional durations and $\alpha + \beta < 1$ to ensure stationarity.[3]

[2] Several different distribution are commonly employed in the ACD literature: exponential, Weibull, Gamma, Burr, Generalized Gamma, for a detailed explanation of their characteristics see Bauwens and Giot(2000).

[3] Formula 1.4 corresponds to an ACD(1,1) specification. The presence of more lags in x and Ψ are defined as higher order ACD(p, q) models.

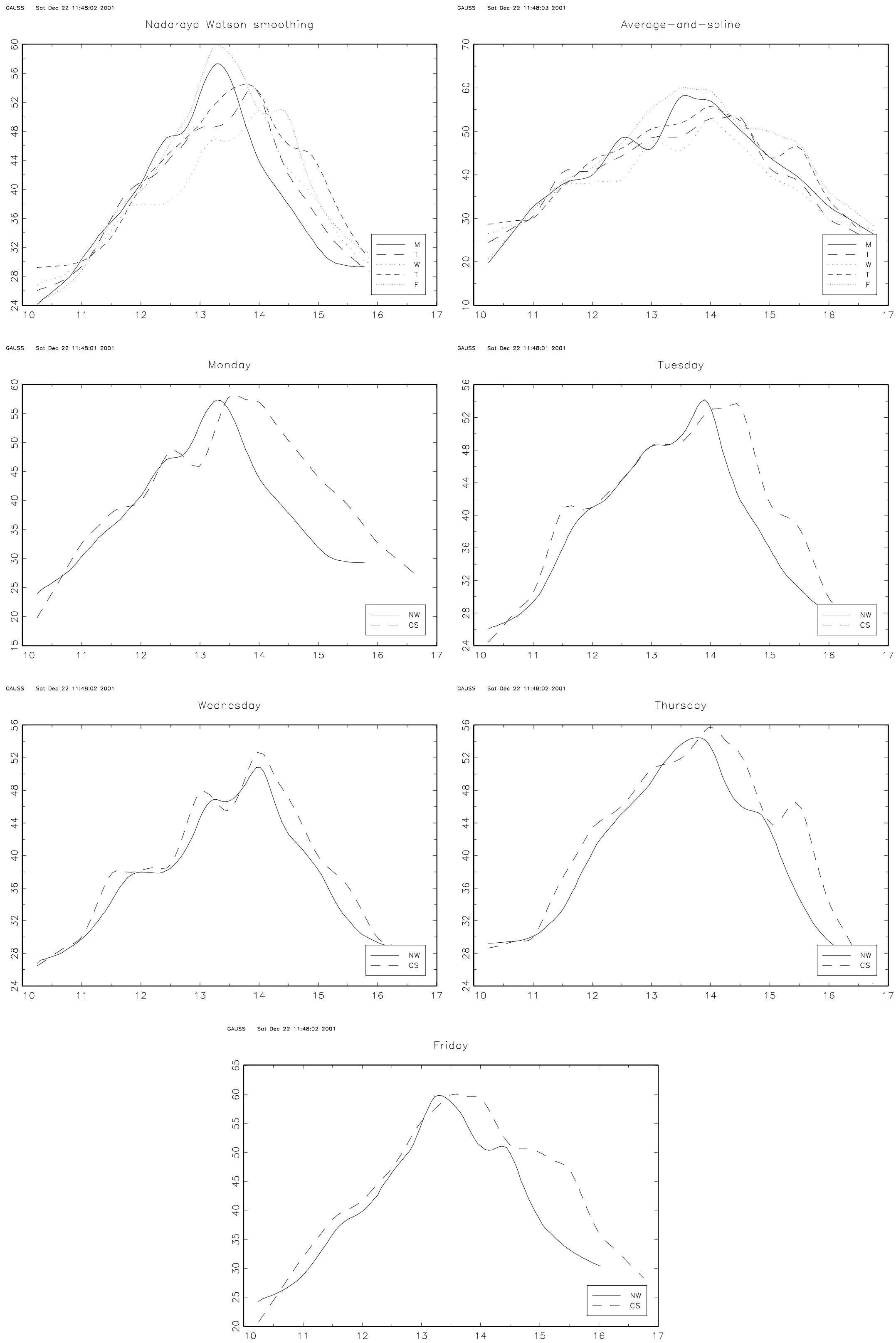

Figure 1.1: Nonparametrically and average-and-spline computed tod-tow for trade durations of Disney stock in the period January-May 1997.

The ACD specification is able to render a number of relevant stylized facts displayed by durations in tick-by-tick financial data, such as autocorrelation, overdispersion and, to some degree, long memory.

Estimation of the ACD model can easily be performed by maximum likelihood. If $f_\epsilon(\epsilon_i;\theta_2)$ stands for the density function of ϵ_i, which can depend on some parameters θ_2, the density function of x_i given $\mathcal{H}_i$ is

$$f_x(x_i|\mathcal{H}_i;\theta) = f_\epsilon(\frac{x_i}{\Psi_i};\theta_2)\,\Psi_i^{-1}, \tag{1.5}$$

so that the log-likelihood function for the parameter $\theta = (\theta_1, \theta_2)$ where θ_1 corresponds to the parameters of the autoregressive equation (typically ω, α, and β) is

$$l(\theta) = \sum_{i=1}^{n} \ln f_x(x_i|\mathcal{H}_i;\theta) = \sum_{i=1}^{n}\left[\ln f_\epsilon(\frac{x_i}{\Psi_i};\theta_2) - \ln \Psi_i\right]. \tag{1.6}$$

In this expression, Ψ_i can be also be any of the alternative specifications proposed in the literature, some of which are introduced below.

The log-likelihood function can be maximized by a standard numerical algorithm, which at convergence delivers the maximum likelihood estimate (MLE) $\hat{\theta}$.

When using a precise hypothesis about the distribution of ϵ_i, one runs the risk of making a mistake, in which case the MLE is not even consistent. A simple alternative is to rely on the quasi-maximum likelihood estimate (QML) $\tilde{\theta}_1$. The appropriate quasi-log-likelihood function is obtained by proceeding as if the distribution of ϵ_i were exponential. Of course, the QML approach is robust to misspecification of the baseline distribution, but it does not provide an estimate of the parameter set θ_2.

1.3.2 Logarithmic ACD

The ACD formulation for the conditional duration requires positivity constraints on the parameters in (1.4) to avoid negative durations. The consequences of imposing these restrictions are particularly troublesome when we introduce additional explanatory variables. For this reason, Bauwens and Giot (2000) introduced a logarithmic version of the ACD model.

In Log-ACD models, equation (1.3) is written as

$$x_i = \exp(\psi_i)\epsilon_i, \tag{1.7}$$

such that ψ_i is the logarithm of the conditional duration $\Psi_i = \exp(\psi_i)$. The difference with ACD models is that the autoregressive equation bears on the logarithm of the conditional duration, rather than on the conditional duration itself. In the simple Log-ACD$(1,1)$ specification, two possible specifications of this equations are

$$\psi_i = \omega + \alpha \log \epsilon_{i-1} + \beta\psi_{i-1} \quad \text{(Type I)} \tag{1.8}$$

and

$$\psi_i = \omega + \alpha\epsilon_{i-1} + \beta\psi_{i-1} \quad \text{(Type II)}. \tag{1.9}$$

Here, no sign restrictions are needed on the parameters to ensure the positivity of the conditional duration. The hypotheses about ϵ_i are the same as in ACD models, as well as the possible probability distributions. Like in the ACD case, the estimation for the Log-ACD can easily be performed by quasi-maximum likelihood.

In the following chapter, we will further detail the properties of Log-ACD models is be provided in the following chapter. The chapter will introduce the more general Log-ACD(p, p) and derive analytical expressions for moments and autocorrelation function, performing an empirical comparison with other specification.

1.3.3 Stochastic Conditional Duration

Like in the case of GARCH literature, upon which the research on ACD heavily relies, the introduction of a second element of randomness appears to be a valid tool for the treatment of the dynamics of the data.

Bauwens and Veredas (2004) suggest that durations are driven by a stochastic dynamic latent factor, which can be interpreted as inversely related to the information dynamically and randomly reaching the market. The authors assume that the latent variable follows the first order autoregressive process

$$\psi_i = \omega + \beta\psi_{i-1} + u_i, \tag{1.10}$$

where u_i is independently normally distributed with zero mean and variance σ^2. Combining (1.10) with (1.7) yields a state space model where (1.7) is transformed in logarithm. Given this structure, the latent factor $\Psi_i = \exp(\psi_i)$ is lognormally distributed, both conditionally on the past durations and unconditionally. Notice that even if ϵ_i is assumed to follow a standard distributions for durations (like those used for ACD models), the conditional distribution of x_i is the mixture of the assumed distribution of ϵ_i and the conditional lognormal of Ψ_i. The authors compute these mixture distributions by unidimensional numerical integration and illustrate the wide range of shapes of the hazard function that can be obtained depending on the distribution specified for ϵ_i. For example, if ϵ_i is distributed as a Weibull, the shape of the hazard function varies form that of the Weibull distribution for very small values of σ^2 to humped shapes for large values of σ^2.

Estimation of the parameters of the SCD model can be done in several ways. Since the model is a nonlinear space state model, due to the non-normality of the distribution of $\ln\epsilon_i$ (unless ϵ_i is itself assumed to be lognormal), simulation-based methods are required for ML estimation, such as it is done in Chapter 3. The approach followed by Bauwens and Veredas (2004) consists in approximating the distribution of $\ln\epsilon_i$ by a normal distribution with mean and variance that depend on the parameters of the assumed true distribution.

A consistent (albeit inefficient) estimator of all the parameters can then be obtained by quasi-maximum likelihood using the Kalman filter and the prediction error decomposition.

In the third chapter of this thesis we sill instead propose to use the an importance sampling algorithm to numerically compute the multidimensional integral that is necessary to calculate the likelihood without approximations.

1.3.4 A short review of other ACD family specifications

This subsection presents a very short review of the alternative specifications of the ACD models. Some of these models represent a parametric extension of the original framework, while others try to analyze the dynamics of financial durations with a slightly different perspective.

Threshold ACD

In the standard ACD framework, the conditional mean dynamics are determined by the simple linear specifications (1.3)-(1.4). argue that financial duration processes require a more flexible specification. In the threshold ACD (TACD) model, proposed by Zhang et al. (1999), conditional duration Psi_i is modeled as a three regime model where the regimes are allowed to have different duration persistence and error distribution:

$$\Psi_i = \omega_i^{(j)} + \alpha_i^{(j)} x_{i-1} + \beta_i^{(j)} \Psi_{i-1} \qquad if \quad x_{i-1} \in R_j, \tag{1.11}$$

with $R_j = [r_{j-1}, r_j)$, $j = 1, 2, \ldots, J$, where the positive integer J is the number of regimes and $0 = r_0 < r_1 < \ldots < r_J = \infty$ are the threshold values. The TACD model can be estimated by QML.

Meitz and Terasvirta (2006) propose a generalization of this specification, characterized by the presence of a smooth transition between regimes.

A generalization based on Box-Cox transformation

Fernandes and Grammig (2006) develop a family of ACD models encompassing many existing formulations existing in the literature. The nesting relies on a Box and Cox (see Box and Cox (1964)) transformation with shape parameter $\lambda \geq 0$ to the conditional duration process and on an asymmetric response to shocks, giving way to:

$$\frac{\Psi_i^\lambda - 1}{\lambda} = \omega_* + \alpha_* \Psi_{i-1}^\lambda [|\epsilon_{i-1} - b| - c(\epsilon i - 1 - b)]^v + \beta \frac{\Psi_{i-1}^\lambda - 1}{\lambda}. \tag{1.12}$$

The shape parameter λ determines whether the Box-Cox transformation is concave ($\lambda \leq 1$) or convex ($\lambda \geq 1$). The augmented autoregressive conditional duration (AACD) then ensues by rewriting (1.4) as

$$\Psi_i^\lambda = \omega + \alpha\Psi_{i-1}^\lambda \left[|\epsilon_{i-1} - b| - c(\epsilon_{i-1} - b)\right]^v + \beta\Psi_{i-1}^\lambda, \tag{1.13}$$

where $\omega = \lambda\omega_* - \beta + 1$ and $\alpha = \lambda\alpha_*$. The AACD model provides a flexible functional form that permits the conditional duration process $\{\Psi_i\}$ to respond in distinct manners to small and large shocks. The shocks impact curve $g(\epsilon_i) = [|\epsilon_i - b| + c(\epsilon_i - b)]^v$ incorporates such asymmetric responses through the shift and rotation parameters b and c, respectively.

The original ACD model is recovered by imposing $\lambda = v = 1$ and $b = c = 0$, whereas letting $\lambda \to 0$ and $b = c = 0$ renders another simpler Box-Cox specification put forward by Dufour and Engle (2000). Further (1.12) reduces to the Log-ACD models either if $\lambda \to 0$, $v = 1$ and $b = c = 0$ (Type I) or if $\lambda, v \to 0$ and $b = c = 0$ (Type II). Other conditional duration models may be built by imposing restrictions on (1.12). Some examples considered by Fernandes and Grammig include the asymmetric Log-ACD ($\lambda \to 0$ and $v = 1$), asymmetric power ACD ($\lambda = v$), asymmetric ACD ($\lambda = v = 1$), and power ACD ($\lambda = v$ and $b = c = 0$). The specifications here found can be estimated by maximum likelihood and are compared by the authors, who argue that both concavity in the shocks impact curve and asymmetric response are useful in tracking the behavior of the data.

Fractionally integrated ACD

Empirically, the evidence for a long range of time dependence intertrade duration is revealed by a highly persistent pattern of the autocorrelations, displaying a slow, hyperbolic rate of decay. The ACD model and most of its derivations account for short serial dependence in expected durations and thus impose an exponential decline pattern on the autocorrelation function. In empirical application of ACD models, the estimated coefficients on lagged variables often sum up nearly to one. Such evidence may indicate a potential misspecification that arises when an exponential decay pattern is fitted to a process showing an hyperbolic rate of decay. This could suggest that a more flexible structure allowing for longer term dependencies might improve the fit. This is the motivation of the introduction by Jasiak (1998) of a class of fractionally integrated ACD models (FIACD).

Considering lag operators, equation (1.4) for a generic ACD(p, q) model can be rewritten as:

$$\Psi_i = \omega + \alpha(L)x_i + \beta(L)\Psi_i, \tag{1.14}$$

where $\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \ldots + \alpha_p L^p$, and $\beta(L) = \beta_1 L + \beta_2 L^2 + \ldots + \beta_q L^q$. This specification implies that the effect of past durations on the current conditional expected value decays exponentially with the lag length. Indeed, the ACD(p, q) can be rewritten as an ARMA (m, m) process in x_i, where $m = \max(p, q)$, and

$$[1 - \phi(L)]x_i = \omega + [1 - \beta(L)]v_i, \tag{1.15}$$

where $\phi(L) = \alpha(L) + \beta(L) = \phi_1 L + \phi_2 L^2 + \ldots + \phi_m L^m$, and $v_i = x_i - \Psi_i$ is the linear innovation of the duration process. The stationarity and invertibility conditions require that the roots of $[1 - \phi(L)]$ and $1 - \beta(L)$ lie outside the unit circle in the complex space. The corresponding fractionally integrated process is obtained by introducing the fractional differencing operator

$$[1 - \phi(L)](1 - L)_i^x = \omega + [1 - \beta(L)]v_i. \tag{1.16}$$

with $0 \leq d \leq 1$. By substituting $x_i - \Psi_i$ for v_i, one can obtain the FIACD(p, d, q) model:

$$\begin{aligned} [1 - \beta(L)]\Psi_i &= \omega + [1 - \beta(L) - [1 - \phi(L)](1 - L)^d]x_i \\ &= \omega + \lambda(L)x_i \end{aligned} \tag{1.17}$$

where $\lambda(L) = \lambda_1 L + \lambda_2 L^2 + \ldots$, a polynomial of infinite order (in estimation it is approximated by truncating after 1000 lags). In order to guarantee the positive sign of the expected duration, all coefficients in the last equation have to be nonnegative. It is remarkable that, when $d = 0$, the FIACD is led back to an ACD, while for $d = 1$ an integrated ACD is obtained.

If durations follow a FIACD process, the first unconditional moment of x_i is infinite and the process is not weakly stationary. However, it can be shown that under some conditions this process can be strictly stationary and ergodic. The parameters of the FIACD specification can be consistently estimated by QML once a viable probability density function is specified for the baseline duration ϵ_i. A small set of estimations on empirical samples appear to deliver significant estimates for the fractional parameter of the FIACD model, which seems to reveal a persistence phenomenon and to suggest a misspecification of the ACD model.

Models for the distribution for the baseline duration

The alternative specifications described so far were related to the functional form of the conditional duration. The ACD model can be further articulated by considering various forms for the distribution of the baseline duration. A series of "standard" distributional models are employed in the literature on ACD models. These distributions (Exponential, which is nested by all others, Weibull, Burr, Gamma, Generalized Gamma, Generalized F...) are usually derived from the usual toolbox of duration analysis and can allow for different dynamics of the hazard function (see Pacurar (2008) for an account of these distributions).

An interesting generalization of the distributional law of the conditional duration is provided by DeLuca and Gallo (2009) and DeLuca and Gallo (2004). In the second and more general work, the authors adopt a mixture of exponentials with time-varying mixing weights, maintaining the hypothesis that the innovation process has unit expected value. The fact that mixing weights depend on time is suggested to be the effect of different regimes in intensity in trading. These weights can be also be directly modeled as depending on some indicators of market activity, allowing for a market microstructure interpretation of their evolution in time.

Multivariate duration models

While the literature on volatility processes (GARCH, SV,...), abounds with multivariate specifications, in the ACD analysis of point processes the extensions in this direction are still a limited number and this seems to be the most promising field of research. The difficulties inherent to this analysis are mainly due to the fact that the components of the vectors of processes that are to be jointly analyzed (trades of different stocks, quote adjustments, financial events defined on the basis of price or volume...) do not necessarily occur at regularly spaced time intervals, which makes traditional time series analysis impossible. As spells in different processes overlap and have different mean durations, it is difficult to retrieve a good set of variables to condition the distribution of the durations between events. In general, it is difficult to model the distribution of a duration when new information can arrive within the duration itself. Furthermore, in an univariate point process durations can be easily ordered, so that only the duration of the spells and their relative positions need to be considered, while in a multivariate setup a clear ordering of the durations of different processes is difficult to obtain and the study of the density of durations needs to be performed conditionally to the clock time too. As it is pointed out in Cox and Isham (1980), a point process can be analyzed under three main different perspectives: distribution of the duration between events, instantaneous probability of an occurrence (termed as intensity of the process) and distribution of the number of occurrence in an arbitrarily fixed span of time. The attempts to obtain a joint model of financial point processes have followed these three lines of research.

For what concerns the direct study of the distribution of the durations between events (the very same perspective employed in the most of the literature on univariate processes), Engle and Lunde (2003) propose a censored bivariate ACD model. The goal of their work us to assess how quickly information in the transaction process impacts the prices via quote adjustments. Their primary interest lies therefore in the time between transactions, considered a the "driving process", and subsequent quote revisions. The two processes are not treated symmetrically and some information is lost if multiple quote revisions occur without intervening transactions. This asymmetry is overcome by Mosconi and Olivetti (2005) who propose a bivariate competing risk model, where the transaction and the quote revision processes can be considered to be collapsed into a single marked point process. Contrarily to the model of Engle and Lunde, the latter could be in principle extended to a multivariate case.

The intensity of the multivariate process is the object of the analysis of Russell (1999), who proposes to model the instantaneous arrival rate conditionally of the multivariate filtration of arrival times and associated marks. Though the ensuing Autocorrelated Conditional

Intensity (ACI) model is conceptually simple and appealing, it is analytically difficult and its estimation requires a great amount of computations. An interesting extension of this approach is provided by Bauwens and Hautsch (2002) who, in the Stochastic Conditional Intensity (SCI) model, add a latent factor to the specification of the conditional intensity function. This stochastic parametrization seems to improve the descriptive power of the model, though it obliges to recur to numerical methods in order to study the likelihood.

the Multivariate Autoregressive Conditional Poisson (MACP) model, a multivariate counting specification is proposed by Heinen and Rengifo (2003), who employ a Poisson and a double Poisson (with an additional parameter) distributions to model the number of financial events that take place during a fixed span of time. Conditionally on past observations, the vector of the means of the occurrences follow a VAR process. The interdependence between the different univariate processes is further modeled via a multivariate normal copula, which introduces contemporaneous correlation between the series. The model can be estimated by maximum likelihood. This framework of analysis seems to be very flexible, as it can be applied to a series of point processes representing in principle any financial event.

1.4 Empirical applications of the ACD models

The models of the ACD family have found their place in the literature also as a component of a set of more general models, that jointly consider price dynamics and duration between financial events. This subsection will hint at some of the most relevant contributions, without going too deeply in the usually analytically involved details of these interesting models.

The field where this class of models has been used the earliest and most frequently is the one of microstructure of financial markets. The hypotheses on the presence of different types of agents and their impact on the trading activity on the market (see for instance Admati and Pfeiderer (1988), Diamond and Verrecchia (1987), Easley and O'Hara (1992) or O'Hara (2000)) require to analyze irregularly spaced data whose frequency of arrival matters, a task the ACD model is very well suited for.

Models for the study of market microstructure usually rely on the mix between a model for the evolution of returns (usually a GARCH or a VAR) and an ACD which function is to account for the changes in frequency of transactions (and therefore, in the context of tick-by-tick data, the frequency of arrival of observations).

Engle and Russell (1998b) employ in transaction arrival times and accompanying measures of price, volume and bid-ask spread to develop a marked point process, i.e. a point process where each occurrence is associated with a vector of marks. The authors propose to decompose the joint distributions of arrival times and price changes into the products of the conditional distribution of price changes and the marginal distribution of arrival time. The latter is modeled an ACD. The distribution of changes in price is instead described with of a multi-state transition model, by specifying the vector of probabilities of being in each state. These probabilities obviously need to sum up to one and are considered to be conditional to the durations of past transactions and to a set of other covariates, such as volume and spread of past transactions. The process is defined Autoregressive Conditional Multinomial (ACM) and can be estimated by maximizing the joint likelihood.

Engle (2000) applies the ACD model to IBM transaction arrival times to develop semi-parametric hazard estimates and conditional intensities. Combining these intensities with a GARCH model of prices gives rise to the UHF-GARCH model, where the dynamics of volatility are conditioned on transaction times. Evidence is found for both short and long run components of volatility and that longer duration and longer expected durations are associated with lower volatilities, coherently with the main results in market microstructure theory.

Ghysels and Jasiak (1998) develop a class of ARCH models for series sampled at unequal time intervals set by trade or quote arrivals. The class of models they introduce is called ACD-GARCH and can be described as a random coefficient GARCH, or doubly stochastic GARCH, where the durations between transaction determine the parameter dynamics. The ACD-GARCH model becomes genuinely bivariate when past asset return volatilities are allowed to affect transaction durations and vice versa. In a different perspective, the spacings between trades can be considered exogenous to the volatility dynamics. This assumption is required in a two-step estimation procedure. The bivariate setup enables to test for Granger causality between volatility and intra-trade durations. Under general conditions, the authors propose several GMM estimation procedures, some having a QML interpretation. As illustration, an empirical study of the IBM tick-by-tick data is presented, where it is argued that the volatility of IBM stock prices Granger causes intra-trade durations and that the persistence in GARCH drops dramatically once intra-trade durations are taken into account.

Dufour and Engle (2000) use an ACD model to study the effect of trading intensity in the price formation process. The authors augment by an ACD Hasbrouck's (see Hasbrouck (1991)) vector autoregressive model for the dynamics of price variations measured on quotes and trades. The coefficients of this model pertaining to trades are allowed to vary with time, with the objective to investigate if the time between trades (modeled by an ACD process) affects the price adjustment to trades and the correlation between current and past trades. As a result of the empirical analysis, the authors find that as time duration between transactions decreases, the price impact of trades, the speed of price adjustment to trade-related information and the positive autocorrelation of trades all increase. This paper was successfully extended both by Spierdijk (2004) and Manganelli (2005) who further augment the original model to include the effect of volume, volatility and classes of frequently and infrequently traded stocks.

Wang (2004) develops an asymmetric variant of the ACD-GARCH model that attempts to capture the asymmetric effect of good news and bad news on intertrade durations. From the analysis it emerges that good-news-based trading will generally lead to increased trading intensity, while bad-news-based trading will generally contribute to longer durations and that that long durations tend to lead to declining prices and low volatility. Noticing that most financial markets allow investors to submit both limit and market orders but it is not always clear why agents choose one over the other, Lo and Sapp (2008) empirically investigate how several microstructure factors influence the choice and timing of submissions and specifically the expected time (duration) between successive orders using an ACD model. The authors find that the order submission process is not symmetric for market and limit orders. For example, only the lagged average volume seems to affects the expected duration of market orders while the lagged market imbalance, quote intensity, average volume and bid-ask spread all shorten the expected duration of limit orders. ACD is finally used as a component of the interdependent model for volatility and inter-transaction duration processes proposed by Grammig and Wellner (2002). The authors overcome the estimation problems (the parameters for the equations of volatility and trade frequency are to be estimated simultaneously) by using a GMM procedure. This model allows to indirectly test the common microstructure hypothesis that volatility is caused by private information affecting prices when informed investors are active.

Not limited to market microstructure, ACD models are a useful tool in other fields of financial economics where the object of analysis is the frequency of a recurring phenomenon. Ivanov and Lewis (2008) for instance examine the determinants of issue cycles for initial public offerings by modeling the time between IPO's with an ACD model. By including proxies for different explanations of issue activity, they evaluate whether the data support the business conditions, time varying adverse selection costs, or sentiment hypotheses. Focardi and Fabozzi (2005) discuss a ACD based theoretical model for explaining credit-risk contagion in credit portfolios. Another example of the flexibility of the ACD framework is provided by Christoffersen and Pelletier (2004) who employ an ACD model to build tests for the evaluation of VaR models based on the dynamics of the violations. Finally, Fischer and Zurlinden (2004), noticing that the timing of interventions offers important information for central bank watchers, consider whether the duration intervals of past interventions matter for future interventions by modeling them as an ACD and applying it to the persistence of interventions by the Federal Reserve, the Bundesbank and the Swiss National Bank.

The Conditional Autoregressive Range (CARR) model, developed by Chou (2005) can be considered an application of the analytics of the ACD model, in that it adapts the ACD equations to describe the evolution of he high/low range of asset prices within fixed time intervals, providing an efficient estimator of local volatility. Though the two models (ACD and CARR) basically share the same functional and statistical forms, there is an essential distinctions between them in that while duration is measured at some random intervals, range is measured at fixed intervals; hence, the natures of the variables of interest are different although they share the common property that all observations are positively valued. Extensions of the CARR model are proposed by Chou and Wang (2007), who include exogenous explanatory variable such as lagged return and lagged trading volume, and by Brunetti and Lidholdt (2002) who develop a fractionally integrated version of the model (FICARR) to accommodate for long memory in the range of log-prices and test it to a database of exchange rates.

1.5 Overview of the following chapters

In chapter 2, we provide existence conditions and analytical expressions of the moments of logarithmic autoregressive conditional duration (Log-ACD) models. We focus on the dispersion index and the autocorrelation function and compare them with those of ACD and SCD models. Using duration data for several stocks traded on the New York Stock Exchange, we compare the models in terms of their ability at fitting some stylized facts.

Moving from the fact that the evaluation of the likelihood function of the SCD model requires to compute an integral that has the dimension of the sample size. In chapter 3, we apply the efficient importance sampling method for computing this integral. We compare EIS-based ML estimation with QML estimation based on the Kalman filter. We find that EIS-ML estimation is more precise statistically, at a cost of an acceptable loss of quickness of computations. We illustrate this with simulated and real data. We show also that the EIS-ML method is easy to apply to extensions of the SCD model.

Finally, the aim of chapter 4 is to carry out a nonparametric analysis of financial durations. We make use of an existing algorithm to describe nonparametrically the dynamics of the process in terms of its lagged realizations and of a latent variable, its conditional mean. The devices needed to effectively apply the algorithm to our dataset are presented. On simulated data, the nonparametric procedure yields better estimates than the ones delivered by an incorrectly specified parametric method. On a real dataset, the nonparametric analysis can convey information on the nature of the data generating process that may not be captured by the parametric specication. In this view, the nonparametric method proposed can be a valuable preliminary analysis able to suggest the choice of a good parametric specication, or a complement of a parametric estimation.

Chapter 2

The moments of Log-ACD models

2.1 Introduction

Until the present contribution, one drawback of the Log-ACD model, was that its unconditional moments were not available analytically. Bauwens and Giot (2000) relied therefore on numerical simulations to compute the moments of several Log-ACD models, in particular their autocorrelation function (ACF) and dispersion index (i.e. the ratio of standard deviation to mean). This led them to conclude that Log-ACD models were able to fit the stylized facts of stock market durations 'as well' as ACD models. These facts are a rather slowly decreasing ACF that starts from a relatively low positive value, a consequence of the clustering of activity, and overdispersion. The latter implies that very small and very large durations occur in higher proportions than is compatible with an exponential distribution.

In this chapter[1] we thus provide analytical expressions for the unconditional moments and ACF for the models belonging to the Log-ACD class as defined in Bauwens and Giot (2000), focusing on its most general parametrization. The results of this work are proved using the method that has been proposed by He et al. (2002) and He (2000) for the moments of exponential GARCH models. We also provide an empirical application in which we compute the unconditional moments and ACF for the ACD and Log-ACD models estimated on financial durations for several stocks traded on the New York Stock Exchange.

The chapter is organized as follows. In Section 2, we detail the class of Log-ACD models, introduced in the previous chapter, and provide the conditions of existence and the general formulae of the moments. In Section 3, we look at the properties of the dispersion index and

[1]This chapter is the result of a joint work with Luc Bauwens (Université Catholique de Louvain, Louvain-la-Neuve) and Pierre Giot (Facultés Universitaires Notre-Dame de la Paix, Namur), published on *Quantitative and Qualitative Analysis in Social Sciences*, 2008-1.

the ACF. In Section 4, a comparison between the conditions for the existence of moments and autocorrelations is carried out between Log-ACD, ACD and SCD models. Section 5 presents the comparison using real data. Section 6 concludes. Proofs are relegated in an appendix.

2.2 Log-ACD(r, q) models: definition and moments

An outline of the main characteristics of the Log-ACD model is provided in the introduction. Here we introduce its more general (r, q) form, which will be the object of the analysis of this chapter.

We denote by x_i the duration between two events that happened at times t_{i-1} and t_i, i.e. $x_i = t_i - t_{i-1}$. We assume that the stochastic process $\{x_i\}$ generating the durations is doubly infinite (i goes from $-\infty$ to $+\infty$).

A Log-ACD model specifies the observed duration as the mixing process

$$x_i = e^{\psi_i}\epsilon_i, \tag{2.1}$$

where the ϵ_i are independent and identically distributed, with

$$\mathrm{E}\,\epsilon_i = \mu \tag{2.2}$$

$$\mathrm{Var}\,\epsilon_i = \sigma^2, \tag{2.3}$$

so that $\mathrm{E}(x_i|\mathcal{H}_i) = \mu \exp(\psi_i)$, where $\mathcal{H}_i$ denotes the information set available at time t_{i-1} (the beginning of the duration x_i), which includes the past durations.

The important assumption, which is the same as for ACD models (see Engle and Russell 1998), is that the dependence in the duration process can be subsumed in the conditional expectation $\mathrm{E}(x_i|\mathcal{H}_i)$, in such a way that $x_i/\mathrm{E}(x_i|\mathcal{H}_i)$ is IID. For further reference, we define

$$\Psi_i = \exp(\psi_i). \tag{2.4}$$

To introduce dependence in the process, which can produce a clustering of durations, ψ_i is specified as an autoregressive equation,[2] which in its most general form (in this chapter) is written as

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j g(\epsilon_{i-j}) + \sum_{j=1}^{p} \beta_j \psi_{i-j}, \tag{2.5}$$

[2] The results derived for the Log-ACD(p, p) can be directly applied to any Log-ACD(r, q) model with $r \neq q$, as the latter specification can always be nested in the former by simply choosing $p = \max\{r, q\}$.

which is equivalent to

$$\Psi_i = e^{\omega} \prod_{j=1}^{p} e^{\alpha_j g(\epsilon_{i-j})} \prod_{j=1}^{p} \Psi_{i-j}^{\beta_j}. \tag{2.6}$$

Two choices of the function $g(\epsilon_{i-j})$ are $\ln \epsilon_{i-j}$ or ϵ_{i-j}. The first one corresponds to the Log-ACD$_1$ model, in which (2.5) becomes

$$\begin{aligned} \psi_i &= \omega + \textstyle\sum_{j=1}^{p} \alpha_j \ln \epsilon_{i-j} + \sum_{j=1}^{p} \beta_j \psi_{i-j} \\ &= \omega + \textstyle\sum_{j=1}^{p} \alpha_j \ln x_{i-j} + \sum_{j=1}^{p} (\beta_j - \alpha_j)\, \psi_{i-j}, \end{aligned} \tag{2.7}$$

and the second one to the Log-ACD$_2$ model, for which

$$\begin{aligned} \psi_i &= \omega + \textstyle\sum_{j=1}^{p} \alpha_j \epsilon_{i-j} + \sum_{j=1}^{p} \beta_j \psi_{i-j} \\ &= \omega + \textstyle\sum_{j=1}^{p} \alpha_j (x_{i-j}/ \exp \psi_{i-j}) + \sum_{j=1}^{p} \beta_j \psi_{i-j}. \end{aligned} \tag{2.8}$$

As it was already remarked in the introduction, several choices are available for the distribution of ϵ_i: exponential, gamma, generalized gamma, Weibull, Burr, lognormal, Pareto..., in principle any distribution with positive support. The choice of a particular distribution should be guided by the desire of having a 'correct' specification, and perhaps by its convenience for estimation. Among the distributions cited above, the Burr and the Pareto do not necessarily have finite moments, so that restrictions on their parameters must be imposed to ensure that the variance and the mean exist. The Burr family includes the Weibull (and the exponential) as a particular case, while the generalized gamma includes the gamma and the Weibull (hence the exponential). All these distributions depend on a scale parameter that we normalize at 1. For distributions that are indexed by a single shape parameter (gamma, Weibull), μ and σ^2 are linked through that parameter. For the exponential distribution, the parameter is fixed to 1 so that $\mu = \sigma^2 = 1$. The Burr and generalized gamma depend on two shape parameters, and are therefore more flexible, in particular they can have a non-monotonous hazard function. The moments of a Log-ACD model depend of course on the moments of ϵ_i.
In order to proceed, let us introduce the matrix

$$\mathbf{\Omega} = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \dots & \beta_{p-1} & \beta_p \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}, \tag{2.9}$$

and the coefficients

$$\phi_k = \beta' \mathbf{\Omega}^{k-p-1} \phi \qquad k > p \tag{2.10}$$

where $\beta = (\beta_1, \ldots, \beta_p)'$, and $\phi = (\phi_1, \ldots, \phi_p)$ such that

$$\begin{aligned}
\phi_0 &= 1 \\
\phi_1 &= \beta_1 \\
\phi_s &= \textstyle\sum_{j=1}^{s} \beta_j \phi_{s-j} \qquad s = 2, \ldots, p \\
\phi_s &= \textstyle\sum_{j=1}^{p} \beta_j \phi_{s-j} \qquad s > p.
\end{aligned} \tag{2.11}$$

Let $\lambda(\mathbf{\Omega})$ be the absolute value of the maximum eigenvalue of the matrix $\mathbf{\Omega}$. The unconditional moments of x_i exist and are independent of i as $k \to \infty$ if and only if $\lambda(\mathbf{\Omega}) < 1$. In this case, $\mathbf{\Omega}^k \to 0$ and $\sum_{j=0}^{k} \mathbf{\Omega}^j \to (I - \mathbf{\Omega})^{-1}$ as $k \to \infty$, which is necessary for the sequence $\{\phi_i\}$ to converge to a finite value (see for example Hamilton (1994) page 20).

Theorem 1 *Assume that* $\mathrm{E}\exp[m\theta_j g(\epsilon_i)]$ *and* $\mu_m = \mathrm{E}|\epsilon_i|^m$ *exist for an arbitrary* $m \in \mathbb{R}_+$. *For the Log-ACD process defined by (2.1)-(2.5), the condition* $\lambda(\Omega) < 1$ *is necessary and sufficient for the existence of the m-th moment* $\mathrm{E}x_i^m$. *Under this condition,*

$$\mathrm{E}x_i^m = \mu_m \exp\left[m\omega(1 - \sum_{j=1}^{p} \beta_j)^{-1}\right] \prod_{j=1}^{\infty} \mathrm{E}\exp\left[m\theta_j g(\epsilon_i)\right]), \tag{2.12}$$

where

$$\begin{aligned}
\theta_1 &= \alpha_1 \\
\theta_s &= \begin{cases} \sum_{j=1}^{s} \alpha_j \phi_{s-j}, & s = 2, ..., p \\ \sum_{j=1}^{p} \alpha_j \phi_{s-j}, & s > p \end{cases}
\end{aligned} \tag{2.13}$$

In the following corollary, we adapt this result to the Log-ACD (1,1) case.

Corollary 1 *For the Log-ACD (1,1), the hypotheses of Theorem 1 reduce to be the following:* $\mathrm{E}\exp[\{m\alpha\beta^{j-1}g(\epsilon_i)\}] < \infty$, $\mu_m < \infty$ *for an arbitrary positive integer* m *and* $|\beta| < 1$. *Under these conditions,*

$$\mathrm{E}x_i^m = \mu_m \exp(\frac{m\omega}{1-\beta}) \prod_{j=1}^{\infty} \mathrm{E}\exp[m\alpha\beta^{j-1}g(\epsilon_i)]. \tag{2.14}$$

For the practical computation of (2.12), the infinite product that appears in the moment expression can be truncated after a sufficiently large number of terms since β^j tends to 0.[3]

[3] In practice, we found that for first and second-order moments, truncation after 1000 terms was more than sufficient to get a high accuracy.

For example, if we use an exponential distribution, $\mathrm{E}(\epsilon^{\alpha\beta^j}) = \Gamma(1+\alpha\beta^j)$ and $\mathrm{E}\exp(\epsilon\alpha\beta^j) = 1/(1-\alpha\beta^j)$, so that both expectations tend to 1 when j tends to infinity.

If α and β are both positive (as is practically always the case), computing the moment given in the previous theorem requires to know $\mathrm{E}(\epsilon^p)$ for any positive p (not necessarily integer) in the Log-ACD$_1$ case, and $\mathrm{E}\exp(p\epsilon)$ in the Log-ACD$_2$ case. The (non-integer) moments $\mathrm{E}(\epsilon^p)$ are available for the generalized gamma and Burr distributions, and all their particular cases. The moment generating function which provides $\mathrm{E}\exp(p\epsilon)$ is only available analytically for the gamma distribution (including the exponential).

To be able to obtain an approximation of the moment generating function for the other distributions considered, namely the Weibull, the Burr and the generalized gamma, one can notice that the following Taylor expansion can be used:

$$\mathrm{E}\exp(p\epsilon) = \sum_{k=0}^{\infty} \frac{p^k}{k!}\mathrm{E}\epsilon^k. \tag{2.15}$$

For any of the p-th order moments $\mathrm{E}\exp(p\epsilon)$ to exist, the infinite series of integer moments $\mathrm{E}\epsilon^k$ must converge to a finite value. In the Burr case, this condition is never satisfied, as the maximum fractional finite moment is determined by the ratio of its two shape parameters. For the Weibull and the generalized gamma instead, the infinite moment series converges only if the shape parameter common to the two distributions is larger than one. In this case, it is possible to truncate the infinite sum and obtain an approximation of the p-th moment $\mathrm{E}\exp(p\epsilon)$.

2.3 Dispersion and autocorrelation function

Durations between stock market events are often characterized by overdispersion, meaning that the standard deviation of the data is larger than their mean (see Section 5). Another important stylized fact is the shape of the ACF, which usually decreases slowly from a relatively low positive first-order autocorrelation. It is therefore essential that Log-ACD models be able to fit such stylized facts, for some parameter values.

Let us measure the degree of dispersion of the random variable x by the variation coefficient, or its square root (= standard deviation/mean) that we call the dispersion index and we denote by δ_x. This ratio is larger than 1 in the case of overdispersion. This measure is a direct by-product of Theorem 1, and we have the following result:

Corollary 2 *For the Log-ACD process defined by (2.1)-(2.5), assume that the hypotheses*

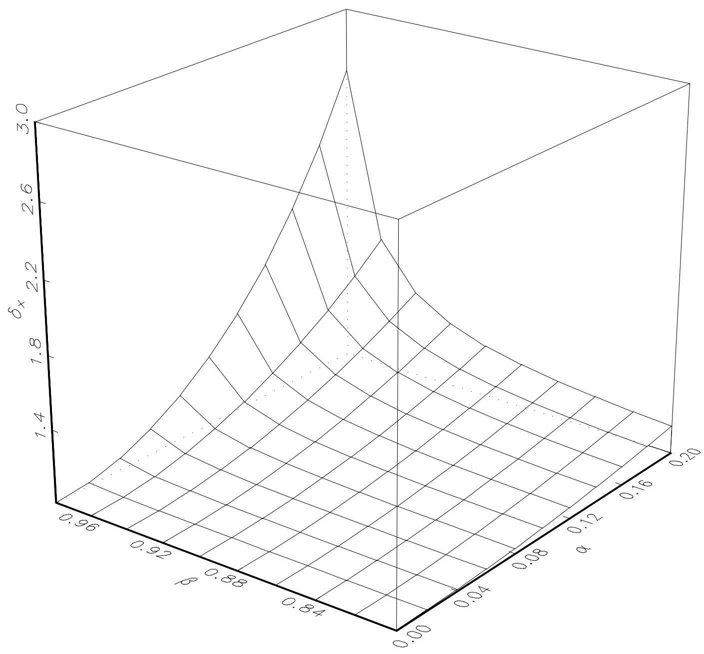

Figure 2.1: Dispersion Index of Log-ACD$_1$ Model (Exponential Distribution)

of Theorem 1 hold for $m = 1, 2$. *Then*

$$1 + \delta_x^2 = \left(1 + \delta^2\right) \frac{\prod_{j=1}^{\infty} \mathrm{E} e^{2\theta_j g(\epsilon_i)}}{\left[\prod_{j=1}^{\infty} \mathrm{E} e^{\theta_j g(\epsilon_i)}\right]^2} \geq 1 + \delta^2 \tag{2.16}$$

where $\delta = \sigma/\mu$ *is the dispersion index of* ϵ_i..

The dispersion index of x_i cannot be smaller than that of ϵ_i. Thus, it suffices that ϵ_i be equidispersed ($\delta = 1$) for x_i to be overdispersed, as long as $\alpha \neq 0$. Figure 2.1 illustrates the variation of δ_x as a function of α (from 0 to 0.2) and β (from 0.8 to 0.98) when ϵ_i is exponential (so that $\delta = 1$) and the model is a Log-ACD$_1$(1,1). For the Log-ACD$_2$(1,1) model, the figure is almost identical, the difference being that the values of δ_x are slighlty smaller (except for the combinations $\alpha = 0.2$ and $0.8 < \beta < 0.94$).

The next theorem provides the autocorrelation function.

Theorem 2 *For the Log-ACD process defined by (2.1)-(2.5), assume that* $\mu < \infty$, $\lambda(\Omega) < 1$, $\mathrm{E}e^{\delta g(\epsilon_i)} < \infty$ *for any* $\delta \in \mathbb{R}$, $\mathrm{E}\epsilon_{i-n} \exp[\theta_n g(\epsilon_{i-n})] < \infty$ *for any* $n \in \mathbb{N}_+$, *and* $\mathrm{E}\exp[(\phi_{n-j}\alpha_{j+h}+$

$\theta^*_{hn})g(\epsilon_{i-n-h})] < \infty$ *for* j *and* h *such that* $n \geq 1$. *Then, for* $n \geq 1$, *the* n*-th order autocorrelation of* $\{x_i\}$ *has the form*

$$\rho_n = \frac{\mu \mathrm{E}\epsilon_i e^{\theta_n g(\epsilon_i)} \prod_{j=1}^{n-1} \mathrm{E}e^{\theta_j g(\epsilon_i)} \prod_{j=p}^{\infty} \mathrm{E}(e^{\theta^*_{jn} g(\epsilon_i)}) M_{n,p} - \mu^2 [\prod_{j=1}^{\infty} \mathrm{E}e^{\theta_j g(\epsilon_i)}]^2}{\mu_2 \prod_{j=1}^{\infty} \mathrm{E}e^{2\theta_j g(\epsilon_i)} - \mu^2 [\prod_{j=1}^{\infty} \mathrm{E}e^{\theta_j g(\epsilon_i)}]^2}, \tag{2.17}$$

where

$$M_{n,p} = \begin{cases} = \prod_{h=1}^{p-n} \mathrm{E}e^{(\sum_{j=1}^{n} \phi_{n-j}\alpha_{h+j} + \theta^*_{hn}) g(\epsilon_{i-n-h})} & \\ \quad \cdot \prod_{h=1}^{n-1} \mathrm{E}e^{(\sum_{j=1}^{h} \phi_{n-j}\alpha_{p-h+j} + \theta^*_{p-h,n}) g(\epsilon_{i-n-h})} & \textit{for } 1 \leq n \leq p \\ = \prod_{h=1}^{p-1} \mathrm{E}e^{(\sum_{j=1}^{p-h} \phi_{n-j}\alpha_{h+j} + \theta^*_{hn}) g(\epsilon_{i-n-h})} & \textit{for } n > p, \end{cases} \tag{2.18}$$

θ^*_{jn} *is defined in (2.54), and* $\mu_2 = \sigma^2 + \mu^2$.

The following corollary specializes the previous to the Log-ACD (1,1) case.

Corollary 3 *For the Log-ACD (1,1) process, the hypotheses of Theorem 2 reduce to be the following:* $|\beta| < 1$ *and* $\mathrm{E}\exp[2\alpha g(\epsilon_i)] < \infty$. *Under these conditions*

$$\rho_n = \frac{\mu \mathrm{E}\left[\epsilon_i e^{\alpha\beta^{n-1} g(\epsilon_i)}\right] \prod_{j=1}^{n-1} \mathrm{E}e^{\alpha\beta^{j-1} g(\epsilon_i)} \prod_{j=1}^{\infty} \mathrm{E}e^{\alpha(1+\beta^n)\beta^{j-1} g(\epsilon_i)} - \mu^2 \left[\prod_{j=1}^{\infty} \mathrm{E}e^{\alpha\beta^{j-1} g(\epsilon_i)}\right]^2}{\mu_2 \prod_{j=1}^{\infty} \mathrm{E}e^{2\alpha\beta^{j-1} g(\epsilon_i)} - \mu^2 \left[\prod_{j=1}^{\infty} \mathrm{E}e^{\alpha\beta^{j-1} g(\epsilon_i)}\right]^2}. \tag{2.19}$$

Some remarks can be made on the features of the autocorrelation function provided by Theorem 2.

First, it is worthwhile to notice that $\lim_{n\to\infty} \rho_n = 0$. This can be easily seen, in the Log-ACD(p,p) instance, by considering that, as $n \to \infty$,

$$\begin{aligned} &\mathrm{E}\left[\epsilon_{i-n} e^{\theta_n g(\epsilon_{i-n})}\right] \to \mu, \\ &\prod_{j=1}^{n-1} \mathrm{E}e^{\theta_j g(\epsilon_i)} \prod_{j=p}^{\infty} \mathrm{E}e^{\theta^*_{jn} g(\epsilon_i)} \to \left[\prod_{j=1}^{n-1} \mathrm{E}e^{\theta_j g(\epsilon_i)}\right]^2, \text{ and} \\ &M_{n,p} \to 1. \end{aligned} \tag{2.20}$$

Hence, the numerator of (2.17) tends to zero.

Another remark is that the shape of ρ_n as a function of n in Theorem 2 is determined by the absolute value of the maximum eigenvalue of the $\mathbf{\Omega}$ matrix. The closer $\lambda(\mathbf{\Omega})$ to 1, the more persistent the autocorrelation. Notice that $\lambda(\Omega) = \beta$ in the Log-ACD$_1$ case.

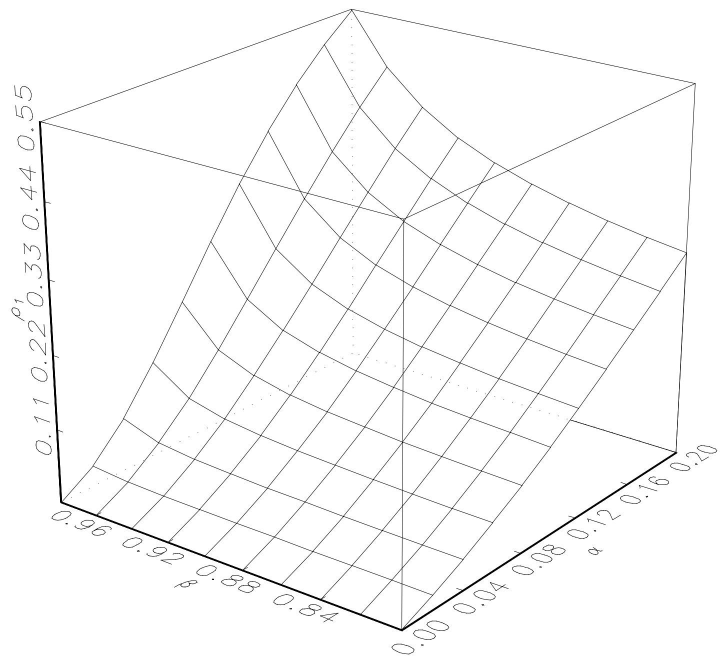

Figure 2.2: First Autocorrelation of Log-ACD$_2$ Model (Exponential Distribution)

Figure 2.2 illustrates the variation of ρ_1 in the same setup as in Figure 2.1 (again with ϵ_i exponential, so that $\mu = \sigma = 1$). For the Log-ACD$_1$(1,1) model, the figure is almost the same, but the value od ρ_1 in the Log-ACD$_1$(1,1) case is larger than in the Log-ACD$_2$(1,1) whenever $\alpha < 0.08$ and smaller whenever $\alpha > 0.14$, while in the intermediate cases it is larger when $\beta > 0.9$ (approximately). However, the differences are never larger than 0.04. These features are not necessarily the same for other distributions of ϵ_i. From this Figure, we see that for $\alpha < 0.10$, ρ_1 does not exceed 0.20 (roughly) when β is smaller than 0.96.

Another feature of interest is the rate of decrease of the ACF. We assume that $0 < \beta < 1$ to avoid oscillation of the signs of the autocorrelations. If we consider for example the Log-ACD$_1$(1,1) model, it can be written as the ARMA(1,1) process

$$\ln x_i = \omega + \beta \ln x_{i-1} + u_i - (\beta - \alpha) u_{i-1} \tag{2.21}$$

where $u_i = \ln x_i - \psi_i$ is a martingale difference. The autocorrelations of the logarithm of the duration therefore decrease geometrically at the rate β. However, by computing (2.19) for many parameter configurations, we found that the autocorrelations of the duration decrease at the above rate only after a 'large' lag. For small lags, the rate of decrease is less than β, although not much. Table 2.1 provides, for several parameter values, the value of ρ_1, the ratio ρ_2/ρ_1, and the value of n from which the rate of decrease is equal to β (for a precision of 4 decimal digits). The results in the table show that i) for fixed β, the larger

Table 2.1: Properties of the ACF of Log-ACD$_2$ Model (Exponential Distribution)

(Exponential Distribution)

	β					
	0.800	0.840	0.880	0.920	0.960	0.980
α						
0.04	*0.045*	*0.046*	*0.048*	*0.051*	*0.061*	*0.079*
	0.793	*0.834*	*0.875*	*0.917*	*0.958*	*0.979*
	24	*30*	*38*	*53*	*93*	*162*
0.08	*0.100*	*0.104*	*0.111*	*0.123*	*0.157*	0.213
	0.785	*0.827*	*0.869*	*0.912*	*0.955*	0.976
	28	*34*	*44*	*63*	*115*	212
0.12	*0.164*	*0.172*	*0.185*	*0.209*	0.270	0.353
	0.775	*0.818*	*0.861*	*0.905*	0.950	0.972
	30	37	48	69	129	244
0.16	0.234	0.247	0.267	0.302	0.380	0.467
	0.763	0.807	0.851	0.896	0.942	0.965
	31	39	51	74	140	268
0.20	0.306	0.324	0.350	0.392	0.474	0.548
	0.749	0.793	0.838	0.885	0.932	0.955
	33	41	53	78	149	288

In each cell, from top to bottom, one finds the value of ρ_1, the ratio ρ_2/ρ_1, and the value of n from which $\rho_{n+1}/\rho_n = \beta$ to four decimal places.

α, the larger the difference $\beta - \rho_2/\rho_1$ and the value of n, and ii) for fixed α, the larger β, the smaller the difference $\beta - \rho_2/\rho_1$ but the larger the value of n.

From Figure 2.2 and Table 2.1, we see that there is a region of parameter values for which the autocorrelation function starts at a low positive value (say less than about 0.2) and decreases "slowly" (see the italicized entries of Table 2.1)

2.4 Comparison with ACD and SCD models

The ACD model, introduced by Engle and Russell (1998a), is defined by the following equations:

$$
\begin{aligned}
&x_i = \Psi_i \epsilon_i \\
&\Psi_i = \omega + \alpha x_{i-1} + \beta \Psi_{i-1} \\
&\omega > 0, \quad \alpha \geq 0, \quad \beta \geq 0, \quad \beta = 0 \text{ if } \alpha = 0,
\end{aligned}
\tag{2.22}
$$

where the baseline duration ϵ_i follows the same assumptions as in the Log-ACD case and $(\alpha + \beta)$ in the ACD conditional duration Ψ_i is analogous to the β term in the logarithmic specification.

For this class of models, computing moments and autocorrelation functions is easy and one can obtain the following simple expression in the ACD (1,1) instance:

$$
\begin{aligned}
&\mu_x = \mathrm{E}x = \frac{\mu\omega}{1 - \mu\alpha - \beta} \qquad \text{if } 0 \leq \mu \text{ and } (\alpha + \beta) < 1, \\
&\delta_x^2 = \frac{\sigma_x^2}{\mu_x^2} = \frac{\sigma^2}{\mu^2} \frac{1 - \beta^2 - 2\mu\alpha\beta}{1 - (\mu\alpha + \beta)^2 - (\alpha\sigma)^2} \geq \delta^2, \\
&\rho_1 = \frac{\alpha(1 - \beta^2 - \alpha\beta)}{1 - \beta^2 - 2\alpha\beta}, \\
&\rho_n = (\alpha + \beta)\rho_{n-1} \quad (n > 1).
\end{aligned}
\tag{2.23}
$$

It must be however noticed that the conditions for the existence of the moments of higher order becomes involve the parameters α and β in the formula for the conditional duration, which is not the case for the Log-ACD model, where conditions on β do not change. Furthermore, the ACF of the durations decreases geometrically at the rate $\alpha + \beta$, since the ACD can be rewritten as an ARMA model with AR parameter $\alpha + \beta$.

Like the Log-ACD model, the SCD model (SCD), Bauwens and Veredas (2004), has a non linear expression for the conditional duration Ψ_i. The model has the following specification:

$$
\begin{aligned}
&x_i = \Psi_i \, \epsilon_i = e^{\psi_i} \, \epsilon_i \\
&\psi_i = \omega + \beta\psi_{i-1} + \eta_i, \quad |\beta| < 1,
\end{aligned}
\tag{2.24}
$$

where again the baseline duration term follows the same assumptions as in the Log-ACD case, but is independent of η_i, the other random term present in the model, characterized by an IID normal distribution with mean 0 and variance σ^2.

Table 2.2: Point Estimates

	expon.	Weibull	gamma	Burr	gen. gamma
ACD					
ω	0.063	0.062	0.063	0.089	0.112
α	0.098	0.097	0.098	0.114	0.128
β	0.840	0.842	0.840	0.803	0.760
Log-ACD$_1$					
ω	0.042	0.042	0.042	0.057	0.050
α	0.090	0.089	0.090	0.109	0.108
β	0.928	0.929	0.928	0.900	0.900
Log-ACD$_2$					
ω	-0.084	-0.083	-0.084	-0.087	-0.089
α	0.082	0.082	0.082	0.089	0.087
β	0.938	0.939	0.938	0.920	0.919

Maximum likelihood estimates of the parameters of the ACD, Log-ACD$_1$ and Log-ACD$_2$ models, assuming various distributions for ϵ_i. Data: price durations (at \$1/8) for IBM, january-april 1997, 18878 observations.

The SCD model allows for a simple structure for moments and ACF, which is

$$\begin{aligned}
\mu_x &= \mu e^{\frac{\omega}{1-\beta}+\frac{1}{2}\frac{\sigma^2}{1-\beta^2}} \\
1+\delta_x^2 &= (1+\delta^2)\, e^{\frac{\sigma^2}{1-\beta^2}} \geq 1+\delta^2 \\
\rho_k &= \frac{e^{\frac{\sigma^2\beta^k}{1-\beta^2}}-1}{(1+\delta^2)\,(e^{\frac{\sigma^2}{1-\beta^2}}-1)} \approx \frac{\sigma^2\beta^k/(1-\beta^2)}{(1+\delta^2)\,(e^{\frac{\sigma^2}{1-\beta^2}}-1)} \approx \beta\rho_{k-1}.
\end{aligned} \tag{2.25}$$

A relevant remark is that, like in the Log-ACD case, the autocorrelation function ρ_k geometrically decreases at rate β only asymptotically, while for small k the decrease rate is smaller.

2.5 Fitting the stylized facts

In this section, we consider an application to financial durations for stocks traded on the NYSE. The objective of this empirical application is to provide an illustrative example of the use of the formulae derived in the previous section. The possibility of calculating the moments that are implied by the estimated parameters allows us also to compare various specifications (ACD, Log-ACD$_1$ and Log-ACD$_2$) and distributions for the baseline durations in their ability to "fit" the sample moments of the data.

As reviewed by Giot (2000), while durations can simply be defined as the time elapsed between two market events, by judiciously defining the notion of market event one can highlight several important features of intraday market activity. For example, a duration between two quotes is a quote duration and the modeling of these using ACD or Log-ACD type models can quantify the notion of quoting activity, i.e. the rate at which the specialists post quotes.

Important extensions related to the quote process are the notions of price and volume durations. Price durations are defined as the minimum time for the stock price to escape from a given price interval. In our application, we focus on the mid-price of the specialist quote, i.e., the average of the bid and ask prices, and the price interval considered is \$0.125. It can be shown (see Giot 2000) that there is a relationship between the volatility of the price process and the conditional hazard of the ACD or Log-ACD model. Thus this provides a strong motivation for the use of such high frequency duration models in the modeling of intraday volatility. A volume duration is defined as the time required for total traded volume to cumulate until a given amount (25000 shares in our application). This duration can be considered as a partial measure of market liquidity, as it indicates the time needed to trade a given amount of shares.

The data set considered in the empirical evaluation consists of series of price and volume durations of five stocks (Boeing, Coca Cola, Disney, Exxon and Ibm) taken from the Trade and Quote (TAQ) database of the New York Stock Exchange. For each stock, we have considered two periods. The first period ranges from september to november 1996, while the second goes from january to april 1997.

To take into account the known seasonal effects, we followed Engle and Russell (1998a) in computing adjusted durations as

$$x_i = X_i/\phi(t_i, j) \tag{2.26}$$

where X_i is the original duration (extracted from the data base) and $\phi(t_i, j)$ is the seasonal effect, considered as the function of the time (t_i) and the day of the week (j) of the transaction. The function $\phi(t_i, j)$ is estimated by averaging over thirty minute intervals for each day of the week and smoothing with a cubic spline. The resulting time-of-day and time-of-week adjusted duration is denoted by x_i.

Each deseasonalized sequence of data has been estimated by ACD(1,1), Log-ACD$_1$(1,1) and Log-ACD$_2$(1,1), and for each one of these models we have considered a series of distributions for the conditional durations, namely: exponential (0 shape parameters), Weibull and gamma (1 shape parameter), and Burr and generalized gamma (2 shape parameters). In all these distributions, a further parameter, the scale one, is present. We have chosen to constrain this parameter to the value such that the expectation of the baseline duration

ϵ_i equals 1 in order to avoid an identification problem with the parameters of the autoregressive factor (another possible choice could have been to fix it to 1). The number of observations is different in each sequence of data, ranging from a minimum of 1609 (for the Coca Cola price durations of 1996) to a maximum of 19680 (for the Ibm price durations of 1996). Table 2.2 reports the ML estimates for the case of Ibm price durations in the 1997 data set.[4] The ML estimates for each model, distribution and data sequence were then used to compute the analytical expressions for the unconditional moments and autocorrelation functions. The results based on the analytical expressions were then compared with the empirical (unconditional) moments and ACF.

Tables 2.4 and 2.5 (at the end of the chapter) report the first two empirical moments and the dispersion indices resulting from the analytical expressions for the three models. Broadly speaking, the unconditional moments computed from the analytical formulae. As one can see from the first moment, the second and the dispersion index, the models are quite capable of reproducing the empirical moments in the fitted distribution of the unconditional durations. The first moment and the dispersion ratio, in particular, seem to be the ones that can be better matched by the analytical values. Of course, some extreme cases arise, in which the estimation can not really catch much of the features of the data or the estimated parameters are very close to some conditions for the existence of moments in the conditional distribution (as it can be the case for the Burr). The analytical (estimated) moments for the Log-ACD$_2$ model are not reported for the Burr and generalized gamma distribution. The reason is that the conditions on the convergence of the series in (2.15) to a finite value are never satisfied in the Burr case and were not satisfied by the parameters resulting from the estimations with the generalized gamma Log-ACD$_2$ model. Figure 2.3 reports as a graphical example the empirical ACF of a series of data (Ibm price durations for the january-april 1997 period) and the ACF computed from the estimated parameters of various models.

In order to summarize the large amount of empirical results obtained, we make a ranking of models. The results of this ranking may serve as a guide for the interpretation of the results. The steps followed have been kept as simple as possible. First, for each stock, period and distribution we computed the percentage difference between the empirical and the theoretical (i.e. resulting from the estimated parameters) first moment and dispersion index (which is also a function of the second moment). We also computed a weighted sum of the absolute difference between the values taken by the empirical and theoretical autocorrelations. Only the first 50 first values were considered and we assigned decreasing weights (0.975^n to the n-th autocorrelation, which assignes a weight of 0.28 to the 50-th lag). Second, the stocks, for each period, were then ranked for each one of the three considered criteria (deviation of the first moment, dispersion and autocorrelations) and the numbers denoting their positions in the rankings were added to provide a global ranking. Third, the resulting ranks were finally added for all the stocks and periods, keeping the distinction between price and volume durations. In the resulting ranks, the models and

[4]We do not report standard errors since they are not needed in the following discussion.

distributions with the lowest values are the ones that better perform globally on the three criteria together.

Table 2.3: Ranking results

Price		**Volume**	
Model	*sum of ranks*	*Model*	*sum of ranks*
eLACD2	24	ggLACD1	33
ggLACD1	27	wLACD1	36
ggACD	38	bLACD1	38
gLACD2	46	gLACD1	46
gLACD1	57	ggACD	49
wLACD2	57	gACD	58
wACD	58	wACD	63
eACD	66	bACD	79
wLACD2	67	eLACD2	79
eLACD1	76	gLACD2	80
gACD	86	wLACD2	82
bLACD1	105	eLACD1	83
bACD	112	eACD	93

Sum of rank points for first moment, dispersion and autocorrelation for all the stocks and periods. A lower value of the sum indicates a better performance. The capital letters denote the model (ACD, Log-ACD$_1$ or Log-ACD$_2$) while the small ones denote the conditional distribution (e for exponential, w for Weibull, g for gamma, b for Burr and gg for generalized gamma).

Table 2.3 displays the results of the rank computation. It is quite evident that the performance of the models and distributions considered varies with the kind of duration, price or volume, that we fitted. For price durations, the generalized gamma seems to be the best distribution, followed by the Weibull. The Burr is strongly penalized by its constraint on the number of existing moments, often failing to correctly model the second moment, which reflects in a poorly fitted dispersion index and ACF. The ranking does not seem to give many hints about what specification (ACD, Log-ACD$_1$ or Log-ACD$_2$) may be preferable, though the exponential Log-ACD$_2$ is the model that performs the best. The results on volume durations lead instead to markedly prefer the Log-ACD$_1$ specification, followed by the ACD one. Here again, one can see that the generalized gamma seems to grant a significant gain over other distributions. This should not come as a surprise, as its parametrization is richer than the one of Weibull and gamma and it does not suffer from the constraints for the existence of moments that characterize the Burr.

2.6 Conclusion

We provide analytical formulae for the moments of Log-ACD(p,p) models. The formulae are more complex than for the ACD model, since the ACD model is actually a linear process (ARMA) whereas the Log-ACD is non-linear. We have shown that the shape of the autocorrelation function of Log-ACD models is different from the shape of the ACF of the ACD model. The formulae can be used to check implied moments from parameter estimates, as in the illustration of this chapter. They could also be used to select parameter values in order to match desired moments (e.g. for designing a Monte Carlo experiment). In an empirical analysis, we tried to illustrate the different aptitudes of various models and distributions in estimating the empirical moments.

Appendix

Proof: [Theorem1]

(i)
For simplicity in the notation, let us define the vectors $\alpha = (\alpha_1, ..., \alpha_p)'$, $\mathbf{g}_i = (g(\epsilon_{i-1}), ..., g(\epsilon_{i-p}))'$. Suppose that $1 \leq k \leq p$. If we apply the definition of Ψ_i in (2.6) to Ψ_{i-1} in (2.6), after rearranging and substituting with ϕ_1 we can write

$$\Psi_i = \exp\{\omega(1+\phi_1)\} \cdot \exp\{\alpha'\mathbf{g}_i + \phi_1\alpha'\mathbf{g}_{i-1}\} \cdot \prod_{j=1}^{p-1} \Psi_{i-j-1}^{\phi_1\beta_j+\beta_{j+1}} \cdot \Psi_{i-p-1}^{\phi_1\beta_p} . \tag{2.27}$$

If we apply it again to Ψ_{i-2} in (2.27) and substitute with ϕ_2, we get

$$\begin{aligned} \Psi_i = \ & \exp\{\omega(1+\phi_1+\phi_2)\} \cdot \exp\{\alpha'\mathbf{g}_i + \phi_1\alpha'\mathbf{g}_{i-1} + \phi_2\alpha'\mathbf{g}_{i-2}\} \\ & \cdot \textstyle\prod_{j=1}^{p-2} \Psi_{i-j-2}^{\beta_j\phi_2+(\beta_{j+1}\phi_1+\beta_{j+2})} \cdot \ \Psi_{i-p-1}^{\beta_{p-1}\phi_2+\beta_p\phi_1} \cdot \ \Psi_{i-p-2}^{\beta_p\phi_2}. \end{aligned} \tag{2.28}$$

Continuing applying the definition given in (2.6) and substituting with ϕ_k until $k = p$, yields

$$\Psi_i = \prod_{j=0}^{p} e^{\omega\phi_j} \prod_{j=0}^{p} e^{\phi_j\alpha'\mathbf{g}_{i-j}} \cdot \Psi_{i-p-1}^{\sum_{j=1}^{p}\beta_j\phi_{p-j+1}} \cdot \Psi_{i-p-2}^{\sum_{j=2}^{p}\beta_j\phi_{p-j+2}} \cdot ... \cdot \Psi_{i-p-p-1}^{\beta_{p-1}\phi_p+\beta_p\phi_{p-1}} \cdot \Psi_{i-p-p}^{\beta_p\phi_p}. \tag{2.29}$$

In order to be able to iterate further, we need to derive an expression of ϕ_k when $k > p$, given (2.9), (2.10) and (2.11). This can be done by noticing that the following equalities

$$\begin{aligned}\mathbf{\Omega}^{p}\,[\phi_0,0,0,...,0]' &= \mathbf{\Omega}^{p-1}\,[\phi_1,\phi_0,0,\ldots,0]' \\ &= \ldots \\ &= \mathbf{\Omega}\,[\phi_{p-1},\phi_{p-2},\ldots,\phi_1,\phi_0]' \\ &= [\phi_p,\phi_{p-1},\ldots,\phi_2,\phi_1]'\end{aligned} \tag{2.30}$$

hold and by applying them to (2.10), to show that

$$\begin{aligned}\phi_k &= \beta'\mathbf{\Omega}^{k-p-2}\mathbf{\Omega}\,[\phi_p,\phi_{p-1},\ldots,\phi_2,\phi_1]' \\ &= \beta_1\beta'\mathbf{\Omega}^{k-p-2}\,[\phi_p,\phi_{p-1},\ldots,\phi_2,\phi_1]+ \\ &\quad\ \beta_2\beta'\mathbf{\Omega}^{k-p-2}\,[\phi_{p-1},\phi_{p-2},\ldots,\phi_1,\phi_0]+ \\ &\quad\ \ldots+ \\ &\quad\ \beta_p\beta'\mathbf{\Omega}^{k-p-2}\,[\phi_1,\phi_0,0,\ldots,0,] \\ &= \beta_1\phi_{k-1}+\beta_2\phi_{k-2}+...+\beta_p\phi_{k-p} \\ &= \textstyle\sum_{j=1}^{p}\beta_j\phi_{k-j}.\end{aligned} \tag{2.31}$$

Let us consider the case $k = p+1$. Applying the definition of Ψ_i in (2.6) to Ψ_{i-p-1} in (2.29), we get

$$\begin{aligned}\Psi_i = \ & \textstyle\prod_{j=0}^{p+1} e^{\omega\phi_j}\prod_{j=0}^{p+1} e^{\phi_j\alpha'\mathbf{g}_{i-j}}\cdot \\ & \Psi_{i-p-2}^{\sum_{j=1}^{p}\beta_j\phi_{p-j+2}}\cdot\Psi_{i-p-3}^{\sum_{j=2}^{p}\beta_j\phi_{p-j+3}}\cdot\ldots\cdot\Psi_{i-p-p}^{\beta_{p-1}\phi_{p+1}+\beta_p\phi_p}\cdot\Psi_{i-p-p-1}^{\beta_p\phi_{p+1}}.\end{aligned} \tag{2.32}$$

For notational simplicity again, let us define the the parameters

$$\begin{aligned}&\xi_{k+1}=\phi_{k+1} \\ &\xi_{k+2}=\beta_2\phi_k+...+\beta_p\phi_{k-p+2} \\ &\xi_{k+3}=\beta_3\phi_k+...+\beta_p\phi_{k-p+3} \\ &\ldots \\ &\xi_{k+p}=\beta_p\phi_k\end{aligned} \tag{2.33}$$

which enable us to write (2.32) as

$$\Psi_i=\prod_{j=0}^{p+1}\exp\{\omega\phi_j\}\cdot\prod_{j=0}^{p+1}\exp\{\phi_j\alpha'\mathbf{g}_{i-j}\}\cdot\prod_{j=1}^{p}\Psi_{i-j-p-1}^{\xi_{p+1+j}}. \tag{2.34}$$

Let us consider now the case $k = m > p + 1$. By recursively applying the definition of Ψ_i in (2.6) to $\Psi_{i-p-1}, \cdots, \Psi_{i-m+1}$, Ψ_{i-m}, and substituting with the ξ_j's we can write

$$\Psi_i = \prod_{j=0}^{m} \exp\{\omega\phi_j\} \cdot \prod_{j=0}^{m} \exp\{\phi_j \alpha' \mathbf{g}_{i-j}\} \cdot \prod_{j=1}^{p} \Psi_{i-j-m}^{\xi_{m+j}}. \tag{2.35}$$

So, if $k > p$, we can use the following general form to express Ψ_i:

$$\Psi_i = \prod_{j=0}^{k} \exp\{\omega\phi_j\} \cdot \prod_{j=0}^{k} \exp\{\phi_j \alpha' \mathbf{g}_{i-j}\} \cdot \prod_{j=1}^{p} \Psi_{i-j-k}^{\xi_{k+j}}. \tag{2.36}$$

(ii) In order to compute the first unconditional moment of x_i, we can multiply (2.36) by ϵ_i and take expectations on both sides, which yields:

$$\mathrm{E}(x_i) = \mu_1 \exp\{\omega \sum_{j=0}^{k} \phi_j\} \cdot \mathrm{E}\left[\prod_{j=0}^{k} \exp\{\phi_j \alpha' \mathbf{g}_{i-j}\} \cdot \prod_{j=1}^{p} \Psi_{i-j-k}^{\xi_{k+j}}\right]. \tag{2.37}$$

As $k \to \infty$, noting that $\lim_{k\to\infty} \xi_{k+i} = 0$, if and only if $\lambda(\mathbf{\Omega}) < 1$, and that the ϵ_i's are iid, we obtain

$$\begin{aligned} \mathrm{E}(x_i) &= \mu_1 \exp\{\omega \textstyle\sum_{j=0}^{\infty} \phi_j\} \cdot \mathrm{E}\left[\textstyle\prod_{j=0}^{\infty} \exp\{\phi_j \alpha' \mathbf{g}_{i-j}\}\right] = \\ &= \mu_1 \exp\{\omega \textstyle\sum_{j=0}^{\infty} \phi_j\} \cdot [\mathrm{E}(\exp\{\alpha_1 \phi_0 g(\epsilon_{i-1})\} \cdot \\ &\mathrm{E}(\exp\{(\alpha_1\phi_1 + \alpha_2\phi_0) g(\epsilon_{i-2})\}) \cdot ... \cdot \mathrm{E}(\exp\{(\textstyle\sum_{j=1}^{p} \alpha_j \phi_{s-j}) g(\epsilon_{i-j})\}) \cdot ...]. \end{aligned} \tag{2.38}$$

If we define θ_j, $j \geq 1$ as the coefficients of $g(\epsilon_{i-j})$ in (2.38), we can see that (2.13) holds and that the first moment of x_i can be written as

$$\mathrm{E}(x_i) = \mu_1 \exp\{\omega \sum_{j=0}^{\infty} \phi_j\} \cdot \prod_{j=0}^{\infty} \mathrm{E} \exp\{\theta_j g(\epsilon_{i-j})\} \tag{2.39}$$

In order to complete the proof, we must show that, from (2.9) and (2.10) $\sum_{j=0}^{\infty} \phi_j = (1 - \sum_{j=1}^{p} \beta_j)^{-1}$ if and only if $\lambda(\mathbf{\Omega}) < 1$. In fact, from (2.10) it follows that

$$\begin{aligned} \textstyle\sum_{j=0}^{\infty} \phi_j &= \textstyle\sum_{j=0}^{p} \phi_j + \sum_{j=p+1}^{\infty} \phi_j = \sum_{j=0}^{p} \phi_j + \beta' \sum_{j=p+1}^{\infty} \mathbf{\Omega}^{j-p-1} \phi = \\ &= \textstyle\sum_{j=0}^{p} \phi_j + \beta'(I - \mathbf{\Omega})^{-1} \phi = (1 - \sum_{j=1}^{p} \beta_j)^{-1}, \end{aligned} \tag{2.40}$$

if and only if $\lambda(\mathbf{\Omega}) < 1$, since

$$(I-\mathbf{\Omega})^{-1} = (1-\sum_{j=1}^{p}\beta_j)^{-1}\begin{bmatrix} 1 & \sum_{j=2}^{p}\beta_j & \sum_{j=3}^{p}\beta_j & \dots & \beta_{p-1} & \beta_p \\ 1 & 1-\beta_1 & \sum_{j=3}^{p}\beta_j & \dots & \dots & \dots \\ 1 & 1-\beta_1 & 1-\beta_1-\beta_2 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \beta_{p-1}+\beta_p & \beta_p \\ 1 & 1-\beta_1 & 1-\beta_1-\beta_2 & \dots & 1-\sum_{j=1}^{p-2}\beta_j & \beta_p \\ 1 & 1-\beta_1 & 1-\beta_1-\beta_2 & \dots & 1-\sum_{j=1}^{p-2}\beta_j & 1-\sum_{j=1}^{p-1}\beta_j \end{bmatrix}. \tag{2.41}$$

As the proof was given for $m = 1$, it must be noted that the same results for $m > 1$ can be derived by raising both sides of (2.37) to the power m.

$\diamond$

Proof: [Corollary 1]
In the Log-ACD (1,1), $\beta_s = \begin{cases} \beta & s=1 \\ 0 & 1<s\le p \end{cases}$ and $\alpha_s = \begin{cases} \alpha & s=1 \\ 0 & 1<s\le p \end{cases}$.
Then $\lambda(\mathbf{\Omega}) = |\beta|$.
Furthermore, (2.14) implies that in (2.31)

$$\phi_k = \beta\phi_{k-1} = \beta^2\phi_{k-2} = \cdots = \beta^{k-1}\phi_1 = \beta^k, \tag{2.42}$$

therefore, in (2.13), θ_s reduces to

$$\theta_s = \alpha\phi_{s-1} = \alpha\beta^{s-1}. \tag{2.43}$$

$\diamond$

Proof: [Corollary 2]
(2.16) follows directly from (2.12). Since $Ey^2 \ge (\mathrm{E}y)^2$, defining y as $\exp[\alpha\beta^{j-1}g(\epsilon_i)]$, we see that each term of the infinite product in (2.16) is not smaller than 1, and equal to 1 if $\alpha = 0$. This implies that $\delta_x \ge \delta$. $\diamond$

Proof: [Theorem 2]

(i)
For notational simplicity, let us define the following parameters

$$\begin{aligned}
&\beta^*_{1n} = \phi_n + 1 \quad \text{for } n \geq 1,\\
&\beta^*_{jn} = \textstyle\sum_{h=1}^{n} \beta_{h+j-1}\phi_{n-h} \quad \text{for } 1 \leq n \leq p-j+1 \text{and } 2 \leq j \leq p-1,\\
&\beta^*_{jn} = \textstyle\sum_{h=j}^{p} \beta_{h}\phi_{n+j-1-h} \quad \text{for } p-j+2 \leq n \leq p \text{and } 2 \leq j \leq p-1,\\
&\beta^*_{pn} = \beta_p\phi_{n+1} \quad \text{for } 1 \leq n \leq p,\\
&\beta^*_{jn} = \textstyle\sum_{h=1}^{p+1-j} \beta_{j+h-1}\phi_{n-h} \quad \text{for} n \geq p \text{and } 2 \leq j \leq p.
\end{aligned} \tag{2.44}$$

and show how they are determined.
We can start by considering the product $(\Psi_i\Psi_{i-n})$ for $1 \leq n \leq p$. If we apply (2.6) to Ψ_i in the product and make use of the results of the first part of the proof of Theorem 1, we obtain

$$\Psi_i\Psi_{i-n} = \prod_{j=0}^{n-1} \exp\{\omega\phi_j\alpha'\mathbf{g}_{i-j}\} \cdot \prod_{h=1}^{p-n+1} \Psi_{i-j-n+1}^{\sum_{j=1}^{n}\beta_{h+j-1}\phi_{n-j}} \cdot \prod_{h=1}^{n-1} \Psi_{i-p-h}^{\sum_{j=1}^{h}\beta_{p-h+j}\phi_{n-j}}\Psi_{i-n}. \tag{2.45}$$

If we suppose that $h = 1$ in the second product term of (2.45) If we multiply by Ψ_{i-n}, it takes the form

$$\Psi_{i-n}^{\sum_{j=1}^{n}\beta_j\phi_{n-j}+1} \tag{2.46}$$

which implies

$$\beta^*_{1n} = \sum_{j=1}^{n} \beta_j\phi_{n-j} + 1 = \phi_n + 1 \quad \text{for } 1 \leq n \leq p, \tag{2.47}$$

which shows how the first expression of (2.44) is determined.
If we suppose that $h = 1$ in the third product term of (2.45), which yields $\Psi_{i-p-1}^{\beta_p\phi_{n-1}}$. So,

$$\beta^*_{pn} = \beta_p\phi_{n-1} \quad \text{for } 1 \leq n \leq p. \tag{2.48}$$

This shows how the fourth expression of (2.44) is determined.
If we consider the remaining cases, defined by $h = 2, ..., p-n+1$ in the second product term and $h = 2, ..., n-1$ in the third. Thus

$$\begin{aligned}
&\Psi_{i-h-n+1}^{\sum_{j=1}^{n}\beta_{h+j}\phi_{n-j}} \qquad \text{for } 2 \leq h \leq p-n+1\\
&\Psi_{i-p-h}^{\sum_{j=1}^{h}\beta_{p-h}\phi_{n-j}} \qquad \text{for } 2 \leq h \leq n-1.
\end{aligned} \tag{2.49}$$

(2.49) indicates that β^*_{jn}, $j = 2, ..., p-n+1$ can be defined by setting $h = 2, ..., p-n+1$ in the first expression of (2.49) and β^*_{jn}, $p-n-2 \leq j \leq p-1$ can be defined by setting

$h = n-1, ..., 2$ in the second. Analogously, for $2 \leq j \leq p-1$

$$\beta^*_{jn} = \begin{cases} \sum_{h=1}^{n} \beta_{h+j-1}\phi_{n-h} & 1 \leq n \leq p-j+1 \\ \sum_{h=1}^{p} \beta_h \phi_{n+j-h-1} & p-j+2 \leq n \leq p. \end{cases} \quad (2.50)$$

Thus the second and the third expressions of (2.44) are derived.
If we finally consider the case of $n > p$. If we set $k = n-1$ in (2.27) and (2.28), we obtain a corresponding representation of $(\Psi_i \Psi_{i-n})$ which reads:

$$\Psi_i \Psi_{i-n} = (\prod_{j=0}^{n-1} \exp\{\omega \phi_j \alpha' \mathbf{g}_{i-j}\}) \cdot (\prod_{j=1}^{p} \Psi_{i-j-n+1}^{\xi_{n+j-1}} \Psi_{i-n}) \quad (2.51)$$

where

$$\begin{aligned} &\xi_n = \phi_n = \beta^*_{1n} \\ &\xi_{n+1} = \beta_2 \phi_{n-1} + ... + \beta_p \phi_{n-p+1} = \beta^*_{2n} \\ &\xi_{n+2} = \beta_3 \phi_{n-1} + ... + \beta_p \phi_{n-p+2} = \beta^*_{2n} \\ &... \\ &\xi_{n+p-1} = \beta_p \phi_{n-1} = \beta^*_{pn}. \end{aligned} \quad (2.52)$$

This shows how the first expression for $n \geq p+1$ and the fifth for $n > p$ of (2.44) are determined.

If now we substitute β_j with β^*_{jn} in (2.31), and suppose

$$\begin{aligned} &\phi^*_{0n} = 1, \\ &\phi^*_{1n} = \beta^*_1, \\ &\phi^*_{kn} = \textstyle\sum_{j=1}^{j-1} \beta_j \phi^*_{k-1,n} + \beta^*_{kn} \quad j = 2, \cdots, p, \text{ and} \\ &\phi^*_k = \beta' \mathbf{\Gamma^{k-p-1}} \phi^* \quad j > p, \end{aligned} \quad (2.53)$$

we obtain an analogous expression for the parameter ϕ^*_{jn}.
Let us then define the following parameters, which will be useful in the remainder of the proof:

$$\theta^*_{hn} = \begin{cases} \sum_{j=1}^{h} \alpha_j \phi^*_{h+1-j,n} & h = 1, ..., p \\ \sum_{j=1}^{p} \alpha_j \phi^*_{h+1-j,n} & h > p \end{cases} . \quad (2.54)$$

(ii)
We now take the expected value of $x_i x_{i-n}$, and we obtain the following expression:

$$\begin{aligned} E(x_i x_{i-n}) &= E(\epsilon_i \epsilon_{i-n} \Psi_i \Psi_{i-n}) = \\ &= E(\epsilon_i \epsilon_{i-n} \textstyle\prod_{j=0}^{n-1} \exp\{\omega \phi_j\} \cdot \prod_{j=0}^{n-1} \exp\{\phi_j \alpha' \mathbf{g}_{i-j}\} \cdot \prod_{j=1}^{p} \Psi_{i-j-n+1}^{\beta^*_{jn}}). \end{aligned} \quad (2.55)$$

If $n \geq p+1$ we can write (2.55) as

$$\begin{aligned} E(x_i x_{i-n}) &= E(\epsilon_i \epsilon_{i-n} \Psi_i \Psi_{i-n}) \\ &= \mu \prod_{j=0}^{n-1} \exp\{\omega\phi_j\} E(\epsilon_{i-n} \prod_{j=1}^{n} \exp\{\phi_j g(\epsilon_{i-j})\} \cdot \\ &\quad \cdot \prod_{j=1}^{p-1} \exp\{(\sum_{j=1}^{p-h} \phi_{n-j}\alpha_{h+j}) g(\epsilon_{i-n-h})\} \cdot \prod_{j=1}^{p} \Psi_{i-j-n+1}^{\beta^*_{jn}}). \end{aligned} \tag{2.56}$$

If we apply the result in (2.36) to the last two products of the right hand side of (2.56) and let $k \to \infty$, we obtain that

$$\begin{aligned} & E\left[\prod_{h=1}^{p-1} \exp\{(\sum_{j=1}^{p-h} \phi_{n-j}\alpha_{h+j}) g(\epsilon_{i-n-j})\}\right] \cdot (\prod_{j=1}^{p} \Psi_{i-j-n+1}^{\beta^*_{jn}}) = \\ &\quad = (\prod_{j=1}^{\infty} \exp\{\omega\phi^*_{jn}\}) \cdot \\ &\quad\quad \cdot E\left[\left(\prod_{h=1}^{P-1} \exp\{(\sum_{j=1}^{p-h} \phi_{n-j}\alpha_{h+j}) g(\epsilon_{i-n-j})\}\right) \cdot (\prod_{j=1}^{p} \exp\{\phi^*_{jn} g(\epsilon_{i-n-j})\})\right] \cdot \\ &\quad\quad \cdot E\left[(\prod_{j=p}^{\infty} \exp\{\phi^*_{jn} g(\epsilon_{i-n-j})\})\right] \\ &\quad = (\prod_{j=1}^{\infty} \exp\{\omega\phi^*_{jn}\}) \cdot E\left[\left(\prod_{h=1}^{P-1} \exp\{(\sum_{j=1}^{p-h} \phi_{n-j}\alpha_{h+j} + \theta^*_{hn}) g(\epsilon_{i-n-j})\}\right)\right] \cdot \\ &\quad\quad \cdot E\left[(\prod_{j=p}^{\infty} \exp\{\phi^*_{jn} g(\epsilon_{i-n-j})\})\right] \\ &\quad = (\prod_{j=1}^{\infty} \exp\{\omega\phi^*_{jn}\}) \cdot E(\prod_{j=p}^{\infty} \exp\{\theta^*_{jn} g(\epsilon_i)\}) \cdot \\ &\quad\quad \cdot E[\prod_{h=1}^{p-1} \exp\{(\sum_{j=1}^{p-h} \phi_{n-j}\alpha_{h+j} + \theta^*_{hn}) g(\epsilon_{i-n-h})\}]. \end{aligned} \tag{2.57}$$

Hence, we can rewrite (2.56) in the following form:

$$\begin{aligned} E(x_i x_{i-n}) = \; & \mu E[\epsilon_{i-n} \exp\{\theta_n g(\epsilon_{i-n})\}] \cdot (\prod_{j=1}^{\infty} \exp\{\omega\phi^*_{jn}\}) \cdot (\prod_{j=1}^{n-1} \exp\{\omega\phi_j\}) \cdot \\ & \cdot E(\prod_{j=1}^{n-1} \exp\{\theta_j g(\epsilon_i)\}) \cdot E(\prod_{j=p}^{\infty} \exp\{\theta^*_{jn} g(\epsilon_i)\}) \cdot \\ & \cdot E[\prod_{h=1}^{p-1} \exp\{(\sum_{j=1}^{p-h} \phi_{n-j}\alpha_{h+j} + \theta^*_{hn}) g(\epsilon_{i-n-h})\}]. \end{aligned} \tag{2.58}$$

If $1 \leq n \leq p$ (2.55) reads

$$\begin{aligned} E(x_i x_{i-n}) &= E(\epsilon_i \epsilon_{i-n} \Psi_i \Psi_{i-n}) = \\ &= \mu \prod_{j=0}^{n-1} \exp\{\omega\phi_j\} E[\epsilon_{i-n} \prod_{j=1}^{n} \exp\{\phi_j g(\epsilon_{i-j})\} \cdot \\ &\quad \cdot \prod_{h=1}^{n} \exp\{(\phi_{n-h} \sum_{j=1}^{p-h} \alpha_{h+j}) g(\epsilon_{i-n-j})\} \cdot \prod_{j=1}^{p} \Psi_{i-j-n+1}^{\beta^*_{jn}}]. \end{aligned} \tag{2.59}$$

If again we apply the result in (2.36) to the last two products of the left hand side of (2.59) and let $k \to \infty$, we obtain

$$\begin{aligned} E\left(\prod_{h=1}^{n} \exp\{(\phi_{n-h} \sum_{j=1}^{p-h} \alpha_{h+j}) g(\epsilon_{i-n-j})\}\right) \cdot (\prod_{j=1}^{p} \Psi_{i-j-n+1}^{\beta_{jn}^*}) = \\ = (\prod_{j=1}^{\infty} \exp\{\omega \phi_{jn}^*\}) \cdot E(\prod_{h=1}^{p-n} \exp\{(\sum_{j=1}^{n} \phi_{n-j}\alpha_{h+j} + \theta_{hn}^*) g(\epsilon_{i-n-j})\}) \cdot \\ \cdot E(\prod_{h=1}^{n-1} \exp\{(e^g, \sum_{j=1}^{h} \phi_{n-j}\alpha_{p-h+j} + \theta_{p-h,n}^*) g(\epsilon_{i-n-j})\}) \cdot \\ \cdot E(\prod_{j=p}^{\infty} \exp(\theta_{jn}^* g(\epsilon_i)\}). \end{aligned} \tag{2.60}$$

Hence, we get the following expression for (2.59)

$$\begin{aligned} E(x_i x_{i-n}) = \; & \mu E[\epsilon_{i-n} \exp\{\theta_n g(\epsilon_{i-n})\}]) \cdot (\prod_{j=1}^{\infty} \exp\{\omega \phi_{jn}^*\}) \cdot (\prod_{j=1}^{n-1} \exp\{\omega \phi_j\}) \cdot \\ & \cdot E(\prod_{j=1}^{n-1} \exp\{\theta_j g(\epsilon_i)\}) \cdot E(\prod_{j=p}^{\infty} \exp(\theta_{jn}^* g(\epsilon_i)\}) \cdot \\ & \cdot E(\prod_{h=1}^{p-n} \exp\{(\sum_{j=1}^{n} \phi_{n-j}\alpha_{h+j} + \theta_{hn}^*) g(\epsilon_{i-n-j})\}) \cdot \\ & \cdot E(\prod_{h=1}^{n-1} \exp\{(\sum_{j=1}^{h} \phi_{n-j}\alpha_{p-h+j} + \theta_{p-h,n}^*) g(\epsilon_{i-n-j})\}). \end{aligned} \tag{2.61}$$

(iii)
Finally, to be able to simplify and derive expressions (2.17)-(2.18), we need to show that, for any $n \geq 1$

$$\prod_{j=0}^{n-1} \exp\{\omega \phi_j\} \prod_{j=1}^{\infty} \exp\{\omega \phi_{jn}^*\} = \exp\{2\omega(1 - \sum_{j=1}^{p} \beta_j)^{-1}\} \tag{2.62}$$

holds. To do so, we can first show that

$$\sum_{j=1}^{\infty} \phi_{jn}^* = (\sum_{j=1}^{p} \beta_{jn}^*)(1 - \sum_{j=1}^{p} \beta_j)^{-1}. \tag{2.63}$$

Let us consider

$$\begin{aligned} \sum_{j=1}^{\infty} \phi_{jn}^* & = \sum_{j=1}^{p} \phi_{jn}^* + \sum_{j=p+1}^{\infty} \phi_{jn}^* \\ & = \sum_{j=1}^{p} \phi_{jn}^* + \beta' \sum_{j=p+1}^{\infty} \Omega^{j-p-1} \phi_n^* \\ & = \sum_{j=1}^{p} \phi_{jn}^* + \beta' (I - \Omega)^{-1} \phi_n^*. \end{aligned} \tag{2.64}$$

Since $(I - \Omega)^{-1}$ is known from (2.41), it is sufficient to consider the case $p = 2$. Then (2.64) becomes,

$$\begin{aligned} \sum_{j=1}^{\infty} \phi_{jn}^* & = \phi_{1n}^* + \phi_{2n}^* + \tfrac{1}{1-\beta_1-\beta_2} \begin{pmatrix} \beta_1 & \beta_2 \end{pmatrix} \begin{pmatrix} 1 & \beta_2 \\ 1 & 1-\beta_1 \end{pmatrix} \begin{pmatrix} \phi_{1n}^* \\ \phi_{2n}^* \end{pmatrix} = \\ & = \tfrac{\phi_{1n}^* + \phi_{2n}^* - \beta_1 \phi_{1n}^*}{1-\beta_1-\beta_2} = \\ & = \tfrac{\beta_{1n}^* + \beta_{2n}^*}{1-\beta_1-\beta_2}. \end{aligned} \tag{2.65}$$

Next, we can show that (2.62) holds for any $n \geq 1$.
Let $n = 1$, then (2.62) has the form

$$\exp\{\omega\} \prod_{j=1}^{\infty} \exp\{\omega\phi^*_{j1}\} = \exp\{\omega\} \exp\{\omega \frac{\beta^*_{11} + \beta^*_{21}}{1 - \beta_1 - \beta_2}\} = \exp\{\frac{2\omega}{1 - \beta_1 - \beta_2}\}. \tag{2.66}$$

Similarly, we can check for $n = 2$.
Assume now that (2.62) holds for $n = m > 2$, that is,

$$\prod_{j=0}^{m-1} \exp\{\omega\phi_j\} \prod_{j=1}^{\infty} \exp\{\omega\phi^*_{jm}\} = \exp\{2\omega(1 - \sum_{j=1}^{2} \beta_j)^{-1}\}, \tag{2.67}$$

we can show that it holds for $n = m + 1$. From (2.67) we have

$$\begin{aligned} \textstyle\sum_{j=0}^{m} \phi_j \quad & + \textstyle\sum_{j=1}^{\infty} \phi^*_{j,m+1} \\ & = \textstyle\sum_{j=0}^{m-1} \phi_j + \phi_m + \frac{\beta^*_{1,m+1} + \beta^*_{2,m+1}}{1-\beta_1-\beta_2} \\ & = \textstyle\sum_{j=0}^{m-1} \phi_j + \frac{\beta^*_{1m} + \beta^*_{2m}}{1-\beta_1-\beta_2} + \left(\phi_m + \frac{\beta^*_{1,m+1} + \beta^*_{2,m+1}}{1-\beta_1-\beta_2} - \frac{\beta^*_{1m} + \beta^*_{2m}}{1-\beta_1-\beta_2}\right) \\ & = 2\omega(1 - \textstyle\sum_{j=1}^{p} \beta_j)^{-1} + \left(\phi_m + \frac{\beta^*_{1,m+1} + \beta^*_{2,m+1}}{1-\beta_1-\beta_2} - \frac{\beta^*_{1m} + \beta^*_{2m}}{1-\beta_1-\beta_2}\right). \end{aligned} \tag{2.68}$$

Now, the second term on the right-hand of (2.68) equals zero, because $\beta^*_{1,m+1} = \phi_{m+1} + 1$, $\beta^*_{2,m+1} = \beta_2\phi_m$ and $\phi_{m+1} = \beta_1\phi_m + \beta_2\phi_{m+1}$. Thus, (2.62) holds for any $n \geq 1$.

◇

Proof: [Corollary 3]
As in Corollary 1, if $p = 1$, then $\lambda(\mathbf{\Omega}) = |\beta|$ and $\theta_s = \alpha\beta^{s-1}$. Hence $E(\epsilon_i e^{\theta_n g(\epsilon_i)})$ in (2.17) reduces to $E(\epsilon_i e^{\alpha\beta^{n-1} g(\epsilon_i)})$, which is finite if $E(\epsilon_i e^{\alpha g(\epsilon_i)}) < \infty$. For the same reason $E(e^{\theta_j g(\epsilon_i)})$ reduces to $E(e^{\alpha\beta_{j-1} g(\epsilon_i)})$, which is finite if $E(e^{2\alpha g(\epsilon_i)}) < \infty$. This last condition also ensures the existence of the second moment of x_i.

Then, as

$$\theta^*_{jn} = \begin{array}{ll} \alpha\phi^*_{1n} = \alpha(\phi_n + 1) = \alpha\beta^{j-1}(\beta^n + 1) & j = 1 \\ \alpha\phi^*_j = \alpha\beta^{j-1}\phi^*_{1n} = \alpha\beta^{j-1}(\beta^n + 1) & j > 1, \end{array} \tag{2.69}$$

the factor $\prod_{j=p}^{\infty} E(e^{\theta^*_{jn} g(\epsilon_i)})$ reduces to $\prod_{j=1}^{\infty} E(e^{\alpha\beta^{j-1}(\beta^n+1) g(\epsilon_i)})$, which is finite if $E(e^{2\alpha g(\epsilon_i)}) < \infty$, as $\lim_{j\to\infty} \beta^{j-1} = 0$.

Noticing that the products of $M_{n,p}$ reduce to 1 if $p = 1$ completes the proof. ◇

Table 2.4: Volume durations - Moments Implied by Point Estimates

	data	eACD	wACD	gACD	bACD	ggACD	eLACD$_1$	wLACD$_1$	gLACD$_1$	bLACD$_1$	ggLACD$_1$	eLACD$_2$	wLACD$_2$	gLACD$_2$
BOEING v96														
Mean	*1.006*	1.098	1.096	1.094	1.111	1.088	0.303	1.001	1.033	1.039	1.034	1.164	1.775	1.082
Variance	*1.539*	10.09	2.242	2.469	2.504	2.306	0.273	1.582	1.723	1.738	1.702	4.627	5.257	3.992
Dispersion	*0.731*	2.714	0.932	1.031	1.015	0.973	1.406	0.761	0.785	0.781	0.768	1.553	0.818	1.553
COCA COLA v96														
Mean	*0.999*	1.017	1.026	1.015	1.023	1.016	0.7954	0.9976	1.006	1.003	1.006	1.019	1.157	1.013
Variance	*1.766*	2.32	1.894	1.894	1.898	1.889	1.383	1.758	1.816	1.801	1.826	2.287	2.392	2.258
Dispersion	*0.876*	1.115	0.894	0.916	0.902	0.911	1.089	0.875	0.891	0.889	0.897	1.097	0.887	1.097
DISNEY v96														
Mean	*0.999*	1.003	1.003	1.002	1.001	0.999	0.538	0.994	1.002	0.999	1.002	1.046	1.279	1.042
Variance	*1.414*	2.353	1.425	1.491	1.432	1.426	0.653	1.388	1.455	1.414	1.420	2.555	2.341	2.540
Dispersion	*0.644*	1.157	0.646	0.697	0.656	0.653	1.120	0.637	0.671	0.645	0.643	1.156	0.655	1.156
EXXON v96														
Mean	*1.000*	1.010	1.016	1.009	1.012	1.013	0.694	1.000	1.005	1.006	1.006	1.030	1.193	1.029
Variance	*1.433*	2.246	1.484	1.524	1.479	1.480	1.049	1.430	1.484	1.451	1.451	2.319	2.060	2.318
Dispersion	*0.658*	1.096	0.662	0.705	0.666	0.665	1.084	0.656	0.685	0.659	0.656	1.089	0.669	1.089
IBM v96														
Mean	*1.006*	1.279	1.638	1.256	1.379	1.305	0.318	4.906	2.855	3.423	1.006	1.251	1.056	0.259
Variance	*1.656*	na	94.74	5.983	11.04	7.115	0.441	46.30	15.07	22.42	1.826	6.572	2.067	0.274
Dispersion	*0.798*	10.00	5.857	1.670	2.193	1.783	1.834	0.962	0.921	0.955	0.896	1.788	0.924	1.763
BOEING v97														
Mean	*1.001*	1.081	1.079	1.077	1.089	1.072	0.308	0.999	1.030	1.033	1.031	1.156	1.736	1.084
Variance	*1.532*	7.579	2.096	2.290	2.279	2.144	0.278	1.564	1.702	1.703	1.677	4.474	4.979	3.931
Dispersion	*0.727*	2.342	0.895	0.986	0.958	0.929	1.392	0.753	0.777	0.771	0.759	1.530	0.807	1.530
DISNEY v97														
Mean	*0.999*	1.001	0.999	0.999	0.999	0.997	0.531	0.995	1.003	1.001	1.003	1.042	1.268	1.038
Variance	*1.399*	2.318	1.405	1.473	1.410	1.405	0.631	1.382	1.449	1.405	1.410	2.518	2.279	2.499
Dispersion	*0.635*	1.147	0.636	0.689	0.644	0.642	1.114	0.628	0.664	0.635	0.634	1.148	0.646	1.148
EXXON v97														
Mean	*1.000*	1.002	1.008	1.000	1.004	0.999	0.401	1.000	1.062	1.038	1.053	1.056	1.515	0.849
Variance	*1.486*	2.594	1.520	1.559	1.543	1.521	0.386	1.469	1.691	1.606	1.646	2.769	3.449	1.789
Dispersion	*0.697*	1.258	0.704	0.748	0.729	0.724	1.183	0.685	0.707	0.700	0.696	1.218	0.709	1.217
IBM v97														
Mean	*1.003*	1.025	1.030	1.024	1.039	1.024	0.288	0.959	1.015	1.013	1.014	1.118	1.643	1.074
Variance	*1.495*	3.759	1.631	1.666	1.762	1.658	0.243	1.381	1.538	1.562	1.540	3.689	4.084	3.399
Dispersion	*0.698*	1.606	0.732	0.768	0.794	0.763	1.392	0.706	0.703	0.723	0.705	1.397	0.717	1.396

Unconditional moments for the ACD, Log-ACD$_1$ and Log-ACD$_2$ models computed by applying the analytical expressions with the estimated parameters. The first column (in italics) gives the empirical moments computed from the data. The capital letters denote the model (ACD, Log-ACD$_1$ or Log-ACD$_2$) while the small ones denote the conditional distribution (e for exponential, w for Weibull, g for gamma, b for Burr and gg for generalized gamma).

Table 2.5: Price durations - Moments Implied by Point Estimates

	data	eACD	wACD	gACD	bACD	ggACD	eLACD$_1$	wLACD$_1$	gLACD$_1$	bLACD$_1$	ggLACD$_1$	eLACD$_2$	wLACD$_2$	gLACD$_2$
BOEING p96														
Mean	*1.006*	1.070	1.055	1.070	1.200	1.021	1.096	0.938	1.060	1.038	1.017	0.988	0.928	0.972
Variance	*2.938*	2.894	3.072	2.947	na	3.488	2.915	2.325	2.776	5.836	3.345	2.203	2.094	2.138
Dispersion	*1.380*	1.236	1.327	1.254	4.008	1.532	1.195	1.283	1.213	2.102	1.494	1.124	1.197	1.124
COCA COLA p96														
Mean	*1.001*	1.008	1.005	1.008	1.034	0.995	1.012	0.986	1.006	1.019	1.011	0.997	0.983	0.997
Variance	*2.377*	2.116	2.216	2.139	3.212	2.435	2.136	2.137	2.134	3.163	2.714	2.048	2.098	2.047
Dispersion	*1.171*	1.040	1.092	1.051	1.415	1.207	1.041	1.094	1.052	1.430	1.286	1.029	1.080	1.029
DISNEY p96														
Mean	*1.001*	1.017	1.016	1.017	1.055	0.990	0.962	0.952	1.013	1.017	1.003	0.995	0.991	0.991
Variance	*2.517*	2.260	2.269	2.176	4.084	2.391	2.036	2.012	2.169	3.707	2.619	2.112	2.109	2.096
Dispersion	*1.229*	1.088	1.093	1.049	1.631	1.199	1.094	1.103	1.054	1.606	1.265	1.064	1.070	1.064
EXXON p96														
Mean	*1.000*	1.004	1.002	1.004	1.036	0.988	0.999	0.974	1.014	1.023	1.007	0.998	0.989	0.998
Variance	*2.432*	2.056	2.106	2.033	3.708	2.277	2.077	2.030	2.105	3.694	2.614	2.030	2.056	2.029
Dispersion	*1.196*	1.019	1.047	1.008	1.566	1.153	1.038	1.066	1.023	1.591	1.256	1.018	1.049	1.018
IBM p96														
Mean	*1.002*	1.029	1.035	1.029	1.126	1.004	0.888	0.945	1.037	1.031	1.016	0.997	1.032	0.914
Variance	*2.429*	2.421	2.351	2.224	5.876	2.512	1.780	1.943	2.225	4.312	2.632	2.173	2.244	1.826
Dispersion	*1.192*	1.134	1.091	1.049	1.908	1.222	1.121	1.083	1.034	1.748	1.244	1.089	1.052	1.089
BOEING p97														
Mean	*1.007*	1.062	1.049	1.062	1.186	1.018	1.087	0.940	1.058	1.038	1.018	0.988	0.934	0.977
Variance	*2.910*	2.795	2.960	2.837	na	3.374	2.850	2.312	2.736	5.647	3.314	2.199	2.105	2.149
Dispersion	*1.367*	1.216	1.301	1.231	3.441	1.502	1.188	1.270	1.203	2.059	1.482	1.118	1.188	1.118
DISNEY p97														
Mean	*1.000*	1.011	1.012	1.011	1.045	0.991	0.955	0.959	1.008	1.014	1.002	0.995	0.998	0.991
Variance	*2.415*	2.201	2.191	2.106	3.574	2.327	1.976	1.983	2.101	3.348	2.533	2.095	2.097	2.076
Dispersion	*1.190*	1.073	1.067	1.029	1.507	1.170	1.079	1.076	1.033	1.501	1.233	1.056	1.052	1.056
EXXON p97														
Mean	*1.000*	1.014	0.923	1.016	na	0.746	1.468	1.204	1.230	1.014	na	0.584	0.148	0.032
Variance	*3.257*	2.326	2.421	2.595	na	2.406	4.522	3.706	3.507	3.348	na	0.736	0.058	0.002
Dispersion	*1.502*	1.124	1.357	1.229	na	1.822	1.054	1.247	1.147	1.501	1.705	1.078	1.287	1.049
IBM p97														
Mean	*1.002*	1.012	1.017	1.012	1.077	0.994	0.910	0.952	1.008	1.029	1.004	0.998	1.023	0.995
Variance	*2.376*	2.222	2.143	2.041	4.904	2.326	1.811	1.899	2.032	4.427	2.496	2.119	2.136	2.110
Dispersion	*1.169*	1.082	1.035	0.997	1.795	1.163	1.088	1.046	1.000	1.781	1.214	1.063	1.019	1.063

Unconditional moments for the ACD, Log-ACD$_1$ and Log-ACD$_2$ models computed by applying the analytical expressions with the estimated parameters. The first column (in italics) gives the empirical moments computed from the data. The capital letters denote the model (ACD, Log-ACD$_1$ or Log-ACD$_2$) while the small ones denote the conditional distribution (e for exponential, w for Weibull, g for gamma, b for Burr and gg for generalized gamma).

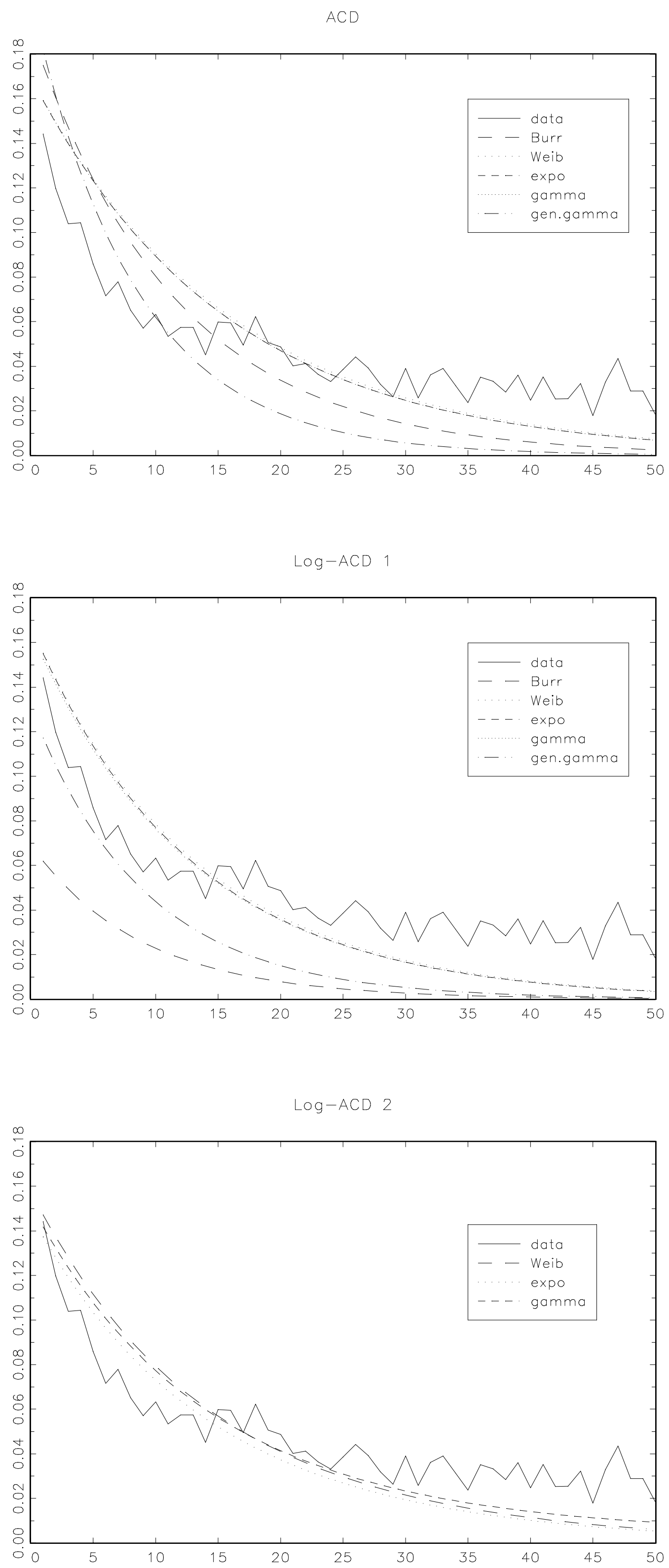

Figure 2.3: ACF for the ACD (top), Log-ACD$_1$ and Log-ACD$_2$ (bottom) models with various conditional distributions (using the analytical expressions computed for the estimated parameters) and empirical data (price duration at \$1/8 for IBM, data 1997).

Chapter 3

EIS for the estimation of SCD models

3.1 Introduction

As it was described in the introduction, the SCD model, proposed by Bauwens and Veredas (2004), is based on the assumption that a second stochastic component is present and it takes the form of a latent variable driving the dynamics of the process. The gain in flexibility granted by the presence of a latent variable in the conditional duration the SCD specification, however, comes at a cost. The fact that the latent conditional duration enters the model nonlinearly leads to the necessity to perform a high-dimensional integral if we want to base our estimation on the exact likelihood.

The solution proposed by Bauwens and Veredas in their original paper was a direct estimation by means of quasi-maximum likelihood (QML) based on the approximation of model with a linear space state representation and the application of the Kalman filter. This method has the advantage of being simple in terms of numerical computation (therefore, fast) and of providing consistent and asymptotically normal estimators, but it is suboptimal in finite samples.

To avoid approximations, in the literature on Stochastic Volatility models a series of estimation procedures have been proposed, for instance GMM, EMM, bayesian estimation based on MCMC sampling. For a survey of these procedures see Ghysels et al. (1996). Some of these solutions have been extended to the estimation of SCD models. Knight and Ning (2008) proposes to estimate SCD models via the Empirical Characteristic Function (ECF) and the Generalized Method of Moments (GMM). Maximum likelihood estimation based on Monte Carlo Markov chain (MCMC) integration of the latent variable is instead adopted by Feng et al. (2004) to estimate both an SCD model in the form proposed by Bauwens and Veredas (2004) and an extended version with a leverage effect determined by the presence of past durations in the mean of the latent variable. MCMC is used also by Strickland, Forbes and Martin (2006), in the context of a Bayesian analysis of the SCD model.

A relatively new method for the estimation of models with latent variables consists in the efficient importance sampling (EIS) procedure, recently developed by Richard and Zhang (2007). This method consists in an extension of the well known importance sampling technique and seems to be particularly well suited for the evaluation of the multidimensional and relatively well behaved integrals of the SCD likelihood. An application to an extended family of SV models is provided in Liesenfeld and Richard (2003) and it well illustrates the flexibility of this algorithm, as well as its speed (particularly if compared to Markov chain methods).

The purpose of this chapter[1] is to apply to an SCD framework the EIS method of numerical sampling, and to use it in order to analyze some extensions of the original specification proposed by Bauwens and Veredas (2004).

The chapter is organized in the following way. In section 2 the main features of the SCD model which are functional to the subsequent analysis will be presented. Section 3 will detail the ML-EIS numeric integration method employed. In section 4 an example with a simulated series of duration data is provided. In section 5 the ML-EIS technique of estimation is applied to the same dataset used in sBauwensVeredas2004 and a comparison is drawn. Section 6 concludes.

3.2 SCD models: main features

In this section, we present briefly the SCD model. A more detailed description can be found in Bauwens and Veredas (2004). If we denote by x_i the duration between two events that happened at times t_{i-1} and t_i, and assume that the stochastic process $\{x_i\}$ generating the durations is doubly infinite (i goes from $-\infty$ to $+\infty$), the stochastic conditional duration model can be written as

$$x_i = e^{\psi_i}\epsilon_i, \tag{3.1}$$

with

$$\psi_i = \omega + \beta\psi_{i-1} + u_i, \tag{3.2}$$

where $|\beta| < 1$ to ensure the stationarity of the process, and

$$u_i \sim iid\ \mathcal{N}(0, \sigma^2), \tag{3.3}$$

and

$$\epsilon_i \sim iid\ p(\epsilon_i), \tag{3.4}$$

[1]This chapter is the result of a joint work with Luc Bauwens (Université Catholique de Louvain, published as CORE DP 2007/53, et accepted for publication in Computational Statistics and Data Analysis

with $p(.)$ a distribution with positive support, and u_j independent of ϵ_i, $\forall i, j$.

The moments and autocorrelation function of this process are

$$
\begin{aligned}
&\mathrm{E}(x_i) = \mu e^{\frac{\omega}{1-\beta}+\frac{1}{2}\frac{\sigma^2}{1-\beta^2}},\\
&1+\delta_x^2 = (1+\delta^2)\, e^{\frac{\sigma^2}{1-\beta^2}}\\
\text{and}\qquad &\\
&\rho_k = \frac{e^{\frac{\sigma^2\beta^k}{1-\beta^2}}-1}{(1+\delta^2)\,(e^{\frac{\sigma^2}{1-\beta^2}}-1)} \approx \frac{\sigma^2\beta^k/(1-\beta^2)}{(1+\delta^2)\,(e^{\frac{\sigma^2}{1-\beta^2}}-1)} \approx \beta\rho_{k-1},
\end{aligned}
\tag{3.5}
$$

where μ stands for $\mathrm{E}(\epsilon_i)$, δ_x for the dispersion index (i.e. the standard deviation to mean ratio) of x_i, and δ for the dispersion index of ϵ_i. The autocorrelation function ρ_k geometrically decreases at rate β only asymptotically with respect to k, while for small k the decrease rate is smaller.

Given a sequence x of n realizations of the process, with density $g(x|\psi, \theta_1)$ indexed by the parameter vector θ_1, conditional on a vector ψ of latent variables of the same dimension as x, and given the density $h(\psi|\theta_2)$ indexed by the parameter θ_2, the likelihood function of x can be written as:

$$L(\theta; x) = L(\theta_1, \theta_2; x) = \int g(x|\psi, \theta_1)h(\psi|\theta_2)d\psi. \tag{3.6}$$

Actually, the integrand in the previous equation is the joint density $f(x, \psi|\theta_1, \theta_2)$. Given the assumptions we made, it can be sequentially decomposed as

$$f(x, \psi|\theta_1, \theta_2) = \prod_{i=1}^{n} p(x_i|\psi_i, \theta_1)q(\psi_i|\psi_{i-1}, \theta_2), \tag{3.7}$$

where $p(x_i|\psi_i, \theta_1)$ is obtained from $p(\epsilon_i)$ using the change of variable in (3.1) (so that θ_1 corresponds to the parameters of $p(.)$), and $q(\psi_i|\psi_{i-1}, \theta_2)$ is the Gaussian density $\mathcal{N}(\omega + \beta\psi_{i-1}, \sigma^2)$ (so that θ_2 includes ω, β and σ^2). Given the functional form usually adopted for $p(\epsilon_i)$ (Weibull, gamma, generalized gamma...), the multidimensional integral in (3.6) cannot be solved analytically and must be computed numerically by simulation.

To perform a QML estimation, one can use the following transformation of the model:

$$
\begin{aligned}
&\ln x_i = \eta + \psi_i + \xi_i\\
\text{and}\qquad &\\
&\psi_i = \omega + \beta\psi_{i-1} + u_i,
\end{aligned}
$$

where $\xi_i = \ln\epsilon_i - \eta$ and $\eta = \mathrm{E}(\ln\epsilon_i)$. This puts the model in state space form with zero mean errors. The ensuing distribution of ξ_i can be approximated by a Gaussian one, and the Kalman filter applied to compute the approximate likelihood.

3.3 EIS based ML estimation

Alternative approaches to QML inference for the SCD model can be based on Monte Carlo methods. These methods are widely used in Bayesian inference to evaluate moments of high-dimensional posterior densities when they are not known analytically. One reason for their success is the relative ease with which they can be applied. EIS, in particular, allows a very accurate evaluation of the likelihood function and has been shown to be quite reliable for the estimation of SV models and of latent factor intensity models, see Bauwens and Hautsch (2006). Another advantage of this algorithm is that its basic structure does not depend on a specific model. This renders changes in the distributional assumptions for the underlying random variables rather simple.

For a detailed presentation of the algorithm, we refer the reader to Richard and Zhang (2007). In this section, we present the basics of its motivation and functioning and we detail its implementation in the context of the ML estimation of SCD models.

Assume one has to evaluate a unidimensional functional integral of the form

$$G(y) = \int_\Lambda g(y, \lambda) p(\lambda) d\lambda, \tag{3.8}$$

where g is an integrable function with respect to a density $p(\lambda)$ with support Λ. The vector y denotes an observed data vector, which in our context corresponds to the observed durations.

A Monte Carlo (MC) estimate of (3.8) is

$$\bar{G}_S(y) = \frac{1}{S} \sum_{i=1}^{S} g(y, \tilde{\lambda}_i), \tag{3.9}$$

where the $\tilde{\lambda}_i$'s are draws from p and S is the number of draws. In cases where one cannot generate directly draws from $p(\lambda)$, one can resort to importance sampling (IS). The IS principle consists of replacing the initial sampler $p(\lambda)$ with an auxiliary parametric importance sampler $m(\lambda, a)$, which is an easy-to-simulate density for λ, indexed by the parameter vector a. To apply importance sampling, equation (3.8) is transformed into

$$G(y) = \int_\Lambda g(y, \lambda) \frac{p(\lambda)}{m(\lambda, a)} m(\lambda, a) d\lambda, \tag{3.10}$$

and the corresponding IS-MC estimate of $G(y)$ is

$$\bar{G}_{S,m}(y, a) = \frac{1}{S} \sum_{i=1}^{S} g(y, \tilde{\lambda}_i) \frac{p(\tilde{\lambda}_i)}{m(\tilde{\lambda}_i, a)}, \tag{3.11}$$

where the $\tilde{\lambda}_i$'s now denote draws from the IS density m.

The aim of *efficient* importance sampling (EIS) is to minimize the MC variance of the estimator in (3.11) by selecting optimally the parameters a of the importance function density m given a functional form for m (for instance the Gaussian density).

Given independent draws $\tilde{\lambda}_i$'s, the sampling variance of $\bar{G}_{S,m}(y,a)$ is given by

$$\mathrm{Var}(\bar{G}_{S,m}(y,a)) = G(y)V(a,y), \tag{3.12}$$

where

$$V(a,y) = \frac{1}{G(y)} \int_{\Lambda} \left[\frac{g(y,\lambda)p(\lambda)}{m(\lambda,a)} - G(y) \right]^2 m(\lambda,a)d\lambda. \tag{3.13}$$

If we denote by $k(\lambda,a)$ the density kernel of the IS sampler $m(\lambda,a)$ and by $\chi(a)$ its integral, such that $m(y,a) = \frac{k(\lambda,a)}{\chi(a)}$, $V(a,y)$ in (3.13) can be rewritten as

$$V(a,y) = \int_{\Lambda} h(d^2(y,a,\lambda))g(y,\lambda)p(\lambda)d\lambda, \tag{3.14}$$

where

$$d(y,a,\lambda) = \ln g(y,\lambda) - \ln p(\lambda) - \ln k(\lambda,a) - \ln G(y) - \ln \chi(a), \tag{3.15}$$

and

$$h(c) = e^{\sqrt{c}} + e^{-\sqrt{c}} - 2 = 2\sum_{i=1}^{\infty} \frac{c^i}{(2i)!}. \tag{3.16}$$

Noting that the term $d(y,a,\lambda)$ is supposed to be small if an efficient sampler is used, the function $h(c)$ can be approximated by its leading term c to get

$$V(a,y) \approx \tilde{V}(a,y) = \int_{\Lambda} d^2(y,a,\lambda)g(y,\lambda)p(\lambda)d\lambda. \tag{3.17}$$

Minimizing the MC variance amounts then to minimizing the quadratic term $d^2(y,a,\lambda)$. It can be shown that if the importance sampler m belongs to the exponential family, the problem remarkably simplifies to a least squares minimization problem for the components of the vector of auxiliary parameters a.

Extending the EIS approach to the case where λ is of high dimension, like in the SCD model (where λ corresponds to the ψ vector), requires that we can decompose the minimization problem in a series of unidimensional subproblems.

A natural MC estimate of the likelihood function in (3.7), which is of the type of (3.9), is given by

$$\tilde{L}(\theta;x) = \frac{1}{S}\sum_{j=1}^{S} \left[\prod_{i=1}^{n} p(x_i|\tilde{\psi}_i^{(j)}, \theta_1) \right], \tag{3.18}$$

where $\tilde{\psi}_i^{(j)}$ denotes a draw from the density $q(\psi_i|\psi_{i-1}^{(j)}, \theta_2)$. This approach bases itself only on the information provided by the distributional assumptions of the model and does not consider the information that comes from the observed sample. It turns out that this estimator is highly inefficient since its sampling variance rapidly increases with the sample size. In any practical case of a duration data set, where the sample size n lies between 500 and 50000 observations, the Monte Carlo sampling size S required to give precise enough estimates of $L(\theta; x)$ would be too high to be affordable and it turns out that this estimator cannot be relied on practically.

EIS tries to make use of the information provided by the observed data in order to come to a reasonably fast and reliable numerical approximation. Let $\{m(\psi_i|\psi_{i-1}, a_i)\}_{i=1}^n$ be a sequence of auxiliary samplers indexed by the set of auxiliary parameter vectors $\{a_i\}_{i=1}^n$. These densities can be defined as a parametric extension of the natural samplers $\{q(\psi_i|\psi_{i-1}, \theta_2)\}_{i=1}^n$. We rewrite the likelihood function as

$$L(\theta; x) = \int \left[\prod_{i=1}^n \frac{f(x_i, \psi_i|x_{i-1}, \psi_{i-1}, \theta)}{m(\psi_i|\psi_{i-1}, a_i)} \prod_{i=1}^n m(\psi_i|\psi_{i-1}, a_i) \right] d\psi. \tag{3.19}$$

Then, its corresponding IS-MC estimator (the equivalent of (3.11)) is given by

$$\tilde{L}(\theta; x, a) = \frac{1}{S} \sum_{j=1}^S \left[\prod_{i=1}^n \frac{f(x_i, \tilde{\psi}_i^{(j)}(a_i)|x_{i-1}, \tilde{\psi}_{i-1}^{(j)}(a_{i-1}), \theta)}{m(\tilde{\psi}_i^{(j)}(a_i)|\tilde{\psi}_{i-1}^{(j)}(a_{i-1}), a_i)} \right]. \tag{3.20}$$

where $\{(\tilde{\psi}_i^{(j)}(a_i)\}_{i=1}^n$ are trajectories drawn from the auxiliary samplers.

Relying on the factorized expression of the likelihood, the MC variance minimization problem can be decomposed in a sequence of subproblems for each element i of the sequence of observations, provided that the elements depending on the lagged values ψ_{i-1} are transferred back to the $(i-1)$-th minimization subproblem. More precisely, if we decompose m in the product of a function of ψ_i and ψ_{i-1} and one of ψ_{i-1} only, such that

$$m(\psi_i|\psi_{i-1}, a_i) = \frac{k(\psi_i, a_i)}{\chi(\psi_{i-1}, a_i)} = \frac{k(\psi_i, a_i)}{\int k(\psi_i, a_i) d\psi_i}, \tag{3.21}$$

we can set up the following minimization problem:

$$\hat{a}_i(\theta) = \arg\min_{a_i} \sum_{j=1}^S \{\ln \left[f(x_i, \tilde{\psi}_i^{(j)}|\tilde{\psi}_{i-1}^{(j)}, x_{i-1}, \theta) \chi(\tilde{\psi}_i^{(j)}, \hat{a}_{i+1}) \right] - c_i - \ln(k(\tilde{\psi}_i^{(j)}, a_i))\}^2, \tag{3.22}$$

where c_i is constant that must be estimated along with a_i. If the density kernel $k(\psi_i, a_i)$ belongs to the exponential family of distributions, the problem becomes linear in a_i, and this greatly improves the speed of the algorithm, as a least squares formula can be employed instead of an iterative routine.

The estimated $\hat{a}_i$ are then substituted in (3.20) to obtain the EIS estimate of the likelihood. The EIS algorithm can be initialized by direct sampling, as in equation (3.18), to obtain a first series of $\tilde{\psi}_i^{(j)}$ and then iterated to allow the convergence of the sequences of $\{a_i\}$, which is usually obtained after 3 to 5 iterations. EIS-ML estimates are finally obtained by maximizing $\tilde{L}(\theta; x, a)$ with respect to θ.

If we adopt a Weibull distribution for ϵ_i with parameter γ (=θ_1) and a $N(0, \sigma^2)$ one for u_i, we come up with the following expressions:

$$p(x_i|\psi_{i-1}, \gamma) = \frac{\gamma}{e^{\psi_i}} \left(\frac{x_i}{e^{\psi_i}}\right)^{\gamma-1} \exp\left\{-\left(\frac{x_i}{e^{\psi_i}}\right)^{\gamma}\right\} \tag{3.23}$$

and

$$q(\psi_i|\psi_{i-1}, \theta_2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(\psi_i - \omega - \beta\psi_{i-1})^2\right\}. \tag{3.24}$$

A convenient choice for the auxiliary sampler $m(\psi_i, a_i)$ is a parametric extension of the natural sampler $q(\psi_i|\psi_{i-1}, \theta_2)$, in order to obtain a good approximation of the integrand without too heavy a cost in terms of analytical complexity. Following Liesenfeld and Richard (2003), we can start by the following specification of the function $k(\psi_i, a_i)$:

$$k(\psi_i, a_i) = q(\psi_i|\psi_{i-1}, \theta_2)\zeta(\psi_i, a_i), \tag{3.25}$$

where $\zeta(\psi_i, a_i) = \exp\{a_{1,i}\psi_i + a_{2,i}\psi_i^2\}$ and $a_i = (a_{1,i}\ \ a_{2,i})$. This specification is rather straightforward and has two advantages. Firstly, as $q(\psi_i|\psi_{i-1}, \theta_2)$ is present in a multiplicative form, it cancels out in the objective function in (3.22), which becomes a least squares problem with $\ln\zeta(\psi_i, a_i)$ that serves to approximate $\ln p(x_i|\psi_i, \theta_1) + \ln\chi(\psi_i, a_i)$. Secondly, such a functional form for k leads to a distribution of the auxiliary sampler $m(\psi_i, a_i)$ that remains Gaussian, as stated in the following theorem, whose proof is given in the appendix.

Theorem 3 *If the functional forms for $q(\psi_i|\psi_{i-1}, \theta_2)$ and $k(\psi_i, a_i)$ are as in equations (3.24) and (3.25) respectively, then the auxiliary density $m(\psi_i|\psi_{i-1}, a_i) = \frac{k(\psi_i, a_i)}{\chi(\psi_{i-1}, a_i)}$ is Gaussian, with conditional mean and variance given by:*

$$\begin{aligned} \mu_i &= v_i^2\left(\tfrac{\omega+\beta\psi_{i-1}}{\sigma^2} + a_{1,i}\right) \\ &\textit{and} \\ v_i^2 &= \tfrac{\sigma^2}{1-2\sigma^2 a_{2,i}}, \end{aligned} \tag{3.26}$$

and the function $\chi(\psi_{i-1}, a_i)$ is given by

$$\frac{1}{\sqrt{1-2\sigma^2 a_{2,i}}} \exp\left\{\frac{\sigma^2}{2(1-2\sigma^2 a_{2,i})}\left(\frac{\omega+\beta\psi_{i-1}}{\sigma^2} + a_{1,i}\right)^2 - \frac{1}{2}\left(\frac{\omega+\beta\psi_{i-1}}{\sigma}\right)^2\right\}. \tag{3.27}$$

By applying these results, it is possible to compute the likelihood function of the SCD model for a given value of θ, based upon the following steps:

Step 1. Use the natural sampler $q(\psi_i|\psi_{i-1}, \theta_2)$ to draw S trajectories of the latent variable $\left\{\tilde{\psi}_i^{(j)}\right\}_{i=1}^{n}$, as in (3.18).

Step 2. The draws obtained in step 1 are used to solve for each i (in the order from n to 1) the least squares problems described in (3.22), which takes the form of the auxiliary linear regression:

$$\begin{aligned}&\ln\gamma - \gamma\tilde{\psi}_i^{(j)} + (\gamma-1)\ln x_i - \left(\frac{x_i}{e^{\tilde{\psi}_i^{(j)}}}\right)^{\gamma} + \ln\chi(\tilde{\psi}_i^{(j)}, \hat{a}_{i+1}) = \\ &\quad a_{0,i} + a_{1,i}\tilde{\psi}_i^{(j)} + a_{2,i}(\tilde{\psi}_i^{(j)})^2 + \varepsilon_i^{(i)}, \qquad\qquad j = 1, ..., S,\end{aligned}$$

where $\varepsilon_i^{(i)}$ is the error term, $a_{0,i}$ is a constant term, and $\chi(\tilde{\psi}_i^{(j)}, \hat{a}_{i+1})$ is set equal to 1 for $i = n$ and defined by (3.27) for $i < n$.

Step 3. Use the estimated auxiliary parameters $\hat{a}_i$ to obtain S trajectories $\left\{\tilde{\psi}_i^{(j)}(\hat{a}_i)\right\}_{i=1}^{N}$ from the auxiliary sampler $m(\psi_i|\psi_{i-1}, \hat{a}_i)$, applying the result of theorem 3.

Step 4. Return to step 2, this time using the draws obtained with the auxiliary sampler. Steps 2, 3 and 4 are usually iterated a small number of times (from 3 to 5), until a reasonable convergence of the parameters $\hat{a}_i$ is obtained.

Once the auxiliary trajectories have attained a reasonable degree of convergence, the simulated samples can be plugged in formula (3.20) to obtain an EIS estimate of the likelihood. This procedure is embedded in a numerical maximization algorithm that converges to a maximum of the likelihood function. After convergence, we compute the standard errors from the Hessian matrix.

Throughout the EIS steps described above and their iterations, we employed a single set of simulated random numbers to obtain the draws from the auxiliary sampler. This technique, known as common random numbers, is motivated in Richard and Zhang (2007). The same random numbers were also employed for each of the likelihood evaluations required by the maximization algorithm.

The number of draws used (S in equation 3.20) for all estimations in this article is equal to 50.

3.4 Simulation results

In order to assess the gain in performance allowed by EIS-ML estimation in comparison with QML, we conducted several repeated simulation experiments with different parameter configurations. As the QML estimator should be consistent but inefficient in relatively small samples, trajectories of 250, 500, 1000, 5000 and 10000 observations form a SCD data generating process (DGP) were simulated 1000 times and the model was estimated both by EIS-ML and by QML. The idea to use as much as 10000 observations comes from the wish to judge the loss of efficiency of QML relative to EIS-ML. Moreover, such sample sizes are far from unusual for real data sets of durations.

The estimations were performed using the MaxSQP maximization function of Ox console 3.40, under Windows XP with a dual core Intel 2.0 Gb processor. The speed of QML estimation with Kalman filter varies from an average of 0.25 seconds for a series of 250 data to an average of 7.5 seconds for a 10000 data one. EIS-ML estimation is much slower, with a average computing time of respectively 2.5 (250 data) and 144 seconds (10000 data). This should not come as a surprise and we suspect that alternative estimation strategies, such as Bayesian MCMC, would be even slower than EIS-ML, as the results of Bauwens and Rombouts (2004) for the SV model clearly show.

The DGP is defined by equations (3.1)-(3.2), plus formula (3.23). The parameter values used in the simulations of the DGP were the following:

- $\omega = 0.0$,
- $\beta = 0.9$,
- $\sigma = 0.05$ and 0.2,
- $\gamma = 0.8$ and 1.1,

thus leading to four combinations. The starting values for likelihood optimizations were set for all estimations to $\omega = 0.0$, $\beta = 0.85$, $\sigma = 0.15$ and $\gamma = 1.05$, but we checked that other reasonable starting values provided quite similar results to those discussed below.

Tables 3.1 to 3.5 contain the means, standard deviations and mean-squared errors of the 1000 estimates for each experiment, and figures 3.1 to 3.4 display the corresponding sampling densities (obtained by kernel based smoothing). As a first remark, it can be noticed that in both estimation methods there is a tendency to underestimate the autoregressive parameter β and to overestimate the parameter σ, especially when the latter takes the low value of 0.05. Anyway, also in these cases the EIS-ML method provides estimates which in mean are closer to the DGP parameters than the QML one. The most striking result concerns the efficiency of the estimators: the EIS-ML estimated standard deviations of the estimates are always remarkably smaller than the QML ones, in particular when the

Table 3.1: Sampling means, standard deviations and mean-squared errors of 1000 estimates of the SCD model parameters for simulated series of 250 observations

250	**DGP**	*Mean EIS*	Mean QML	*StDev EIS*	StDev QML	*MSE EIS*	MSE QML
ω	**0.0000**	*0.0064*	0.0000	*0.0325*	0.0649	*0.0010*	0.0042
β	**0.9000**	*0.8262*	0.7601	*0.1482*	0.2958	*0.0274*	0.1071
σ	**0.2000**	*0.2431*	0.2771	*0.1073*	0.1939	*0.0133*	0.0435
γ	**1.1000**	*1.1261*	1.1823	*0.0792*	0.4732	*0.0069*	0.2307
ω	**0.0000**	*0.0083*	-0.0332	*0.0381*	0.1373	*0.0015*	0.0199
β	**0.9000**	*0.5963*	0.0941	*0.2784*	0.5847	*0.1697*	0.9912
σ	**0.0500**	*0.1703*	0.3521	*0.1384*	0.3583	*0.0336*	0.2197
γ	**1.1000**	*1.1415*	1.5280	*0.0801*	1.2059	*0.0081*	1.6373
ω	**0.0000**	*0.01471*	-0.0729	*0.0534*	0.2164	*0.0030*	0.0521
β	**0.9000**	*0.62392*	-0.0906	*0.2828*	0.5489	*0.1562*	1.2827
σ	**0.0500**	*0.18714*	0.5391	*0.1982*	0.5122	*0.0581*	0.5016
γ	**0.8000**	*0.83921*	1.3096	*0.0779*	1.4544	*0.0076*	2.3752
ω	**0.0000**	*0.0042*	-0.0235	*0.0482*	0.1411	*0.0023*	0.0204
β	**0.9000**	*0.7804*	0.4938	*0.2092*	0.5288	*0.0580*	0.4446
σ	**0.2000**	*0.2734*	0.4394	*0.1657*	0.3681	*0.0328*	0.1928
γ	**0.8000**	*0.8261*	0.9536	*0.0634*	0.6779	*0.0047*	0.4831

parameter σ is equal to 0.05. The combination of smaller bias and variance is reflected clearly in the mean-squared errors, which are sensibly lower across the board for the EIS-ML method. The better general performance of the EIS-ML estimator can be appreciated also by a visual inspection of the sampling densities.

Looking at the tables and at figures 3.2 and 3.3 it is easy to remark how poor the performance of the QML estimator is when the parameter σ is small. To better illustrate the issue, Figure 3.5 shows for each parameter the graph of the estimated standard deviations against the value of σ in the DGP, for a sample size of 1000. These results are based on additional simulations (for values of σ ranging from 0.02 to 0.7). This figure shows that for small values of σ in the DGP, both estimation methods tend to be imprecise. This is understandable, since as σ tends to 0, the parameter β becomes unidentified, so that the likelihood function becomes flat. We also see in figures 3.2 and 3.3 that the sampling distribution of the estimates of β has a mode at (or close to) zero. However, this problem of is far more pronounced for QML than for EIS-ML.

In order to dispel the doubt that the QML estimator might not be consistent (or that the computer program is poorly written) we present in table 3.6 the results based on ten QML estimations with one million data. From all this evidence we conclude that when the sample size is not very large (of the order of hundreds of thousands), QML estimation can be extremely inefficient if the latent factor variance is small.

Table 3.2: Sampling means, standard deviations and mean-squared errors of 1000 estimates of the SCD model parameters for simulated series of 500 observations

500	DGP	*Mean EIS*	Mean QML	*StDev EIS*	StDev QML	*MSE EIS*	MSE QML
ω	**0.0000**	*0.0038*	-0.0021	*0.0146*	0.0255	*0.0002*	0.0006
β	**0.9000**	*0.8699*	0.8467	*0.0770*	0.1476	*0.0068*	0.0246
σ	**0.2000**	*0.2186*	0.2357	*0.0706*	0.1149	*0.0053*	0.0144
γ	**1.1000**	*1.1089*	1.1202	*0.0535*	0.0778	*0.0029*	0.0064
ω	**0.0000**	*0.0043*	-0.0433	*0.0255*	0.1190	*0.0006*	0.0160
β	**0.9000**	*0.6200*	0.0981	*0.2776*	0.5902	*0.1554*	0.9913
σ	**0.0500**	*0.1430*	0.3270	*0.1132*	0.3392	*0.0214*	0.1918
γ	**1.1000**	*1.1254*	1.4719	*0.0562*	1.1315	*0.0038*	1.4187
ω	**0.0000**	*0.0068*	-0.0701	*0.0368*	0.1831	*0.0014*	0.0384
β	**0.9000**	*0.6602*	-0.0944	*0.2700*	0.5288	*0.1303*	1.2686
σ	**0.0500**	*0.1483*	0.4989	*0.1534*	0.4563	*0.0332*	0.4098
γ	**0.8000**	*0.8224*	1.2012	*0.0584*	1.2614	*0.0039*	1.7521
ω	**0.0000**	*0.0035*	-0.0221	*0.0261*	0.0994	*0.0006*	0.0103
β	**0.9000**	*0.8442*	0.6473	*0.1308*	0.4308	*0.0202*	0.2494
σ	**0.2000**	*0.2372*	0.3708	*0.1119*	0.3157	*0.0139*	0.1289
γ	**0.8000**	*0.8109*	0.8863	*0.0428*	0.4887	*0.0019*	0.2462

Table 3.3: Sampling means, standard deviations and mean-squared errors of 1000 estimates of the SCD model parameters for simulated series of 1000 observations

1000	DGP	*Mean EIS*	Mean QML	*StDev EIS*	StDev QML	*MSE EIS*	MSE QML
ω	**0.0000**	*0.0019*	0.0001	*0.0090*	0.0101	*0.0000*	0.0001
β	**0.9000**	*0.8859*	0.8803	*0.0423*	0.0626	*0.0019*	0.0043
σ	**0.2000**	*0.2099*	0.2147	*0.0428*	0.0674	*0.0019*	0.0047
γ	**1.1000**	*1.1037*	1.1098	*0.0352*	0.0414	*0.0012*	0.0018
ω	**0.0000**	*0.0012*	-0.0516	*0.0166*	0.1122	*0.0002*	0.0152
β	**0.9000**	*0.6883*	0.1722	*0.2571*	0.5722	*0.1109*	0.8570
σ	**0.0500**	*0.1074*	0.3315	*0.0882*	0.3381	*0.0110*	0.1935
γ	**1.1000**	*1.1133*	1.4830	*0.0365*	1.1450	*0.0015*	1.4579
ω	**0.0000**	*0.0018*	-0.0915	*0.0198*	0.1822	*0.0003*	0.0415
β	**0.9000**	*0.7376*	-0.0241	*0.2367*	0.5091	*0.0823*	1.1131
σ	**0.0500**	*0.1056*	0.5402	*0.1166*	0.4680	*0.0167*	0.4594
γ	**0.8000**	*0.8111*	1.2733	*0.0383*	1.3596	*0.0015*	2.0726
ω	**0.0000**	*0.0016*	-0.009	*0.0187*	0.0514	*0.0003*	0.0027
β	**0.9000**	*0.8712*	0.7818	*0.0957*	0.2838	*0.0100*	0.0945
σ	**0.2000**	*0.2183*	0.2974	*0.0718*	0.2360	*0.0055*	0.0652
γ	**0.8000**	*0.8043*	0.8285	*0.0300*	0.1498	*0.0009*	0.0232

Table 3.4: Sampling means, standard deviations and mean-squared errors of 1000 estimates of the SCD model parameters for simulated series of 5000 observations

5000	**DGP**	*Mean EIS*	Mean QML	*StDev EIS*	StDev QML	*MSE EIS*	MSE QML
ω	**0.0000**	*0.0005*	0.0002	*0.0031*	0.0034	*0.0000*	0.0000
β	**0.9000**	*0.8982*	0.8959	*0.0152*	0.0201	*0.0002*	0.0004
σ	**0.2000**	*0.1999*	0.2035	*0.0175*	0.0263	*0.0003*	0.0007
γ	**1.1000**	*1.0994*	1.1026	*0.0146*	0.0173	*0.0002*	0.0003
ω	**0.0000**	*-0.0002*	-0.0302	*0.0052*	0.0703	*0.0000*	0.0058
β	**0.9000**	*0.7793*	0.3099	*0.2142*	0.4937	*0.0604*	0.5919
σ	**0.0500**	*0.0785*	0.2934	*0.0522*	0.2516	*0.0035*	0.1225
γ	**1.1000**	*1.1048*	1.2761	*0.0162*	0.7875	*0.0002*	0.6512
ω	**0.0000**	*-0.0003*	-0.0767	*0.0068*	0.1265	*0.0000*	0.0219
β	**0.9000**	*0.7893*	0.0781	*0.2135*	0.3672	*0.0578*	0.8103
σ	**0.0500**	*0.0786*	0.5901	*0.0755*	0.3513	*0.0065*	0.4152
γ	**0.8000**	*0.8033*	1.0754	*0.0154*	1.0024	*0.0002*	1.0807
ω	**0.0000**	*0.0003*	-0.0001	*0.0047*	0.0069	*0.0000*	0.0000
β	**0.9000**	*0.8931*	0.8909	*0.0380*	0.0536	*0.0014*	0.0029
σ	**0.2000**	*0.2024*	0.2086	*0.0295*	0.0612	*0.0008*	0.0038
γ	**0.8000**	*0.8001*	0.8023	*0.0109*	0.0158	*0.0001*	0.0002

Table 3.5: Sampling means, standard deviations and mean-squared errors of 1000 estimates of the SCD model parameters for simulated series of 10000 observations

10000	**DGP**	*Mean EIS*	Mean QML	*StDev EIS*	StDev QML	*MSE EIS*	MSE QML
ω	**0.0000**	*<0.0001*	0.0001	*0.0021*	0.0024	*<0.0000*	<0.0000
β	**0.9000**	*0.9005*	0.8999	*0.0119*	0.0139	*0.0001*	0.0002
σ	**0.2000**	*0.1981*	0.1986	*0.0134*	0.0192	*0.0002*	0.0004
γ	**1.1000**	*1.0999*	1.1002	*0.0098*	0.0114	*<0.0000*	0.0001
ω	**0.0000**	*-0.0002*	-0.0307	*0.0021*	0.0278	*<0.0000*	0.0044
β	**0.9000**	*0.8561*	0.2351	*0.1145*	0.3954	*0.0150*	0.5985
σ	**0.0500**	*0.0615*	0.3606	*0.0299*	0.1944	*0.0010*	0.1342
γ	**1.1000**	*1.1018*	1.1761	*0.0092*	0.0701	*<0.0000*	0.0107
ω	**0.0000**	*-0.0001*	-0.0867	*0.0024*	0.1213	*<0.0000*	0.0222
β	**0.9000**	*0.8411*	0.0522	*0.1481*	0.3294	*0.0254*	0.8273
σ	**0.0500**	*0.0624*	0.6381	*0.0432*	0.3477	*0.0020*	0.4668
γ	**0.8000**	*0.8013*	1.0656	*0.0068*	0.8636	*<0.0000*	0.8164
ω	**0.0000**	*-0.0003*	<0.0001	*0.0025*	0.0027	*<0.0000*	<0.0000
β	**0.9000**	*0.8981*	0.9001	*0.0154*	0.0221	*0.0002*	0.0005
σ	**0.2000**	*0.2010*	0.1992	*0.0190*	0.0307	*0.0004*	0.0009
γ	**0.8000**	*0.8005*	0.7994	*0.0069*	0.0081	*<0.0000*	<0.0000

Table 3.6: Sampling means, standard deviations and mean-squared errors of 1000 estimates of the SCD model parameters for simulated series of one million observations

1000000	**DGP**	Mean QML	StDev QML	Min QML	Max QML
ω	**0.0000**	-0.0001	0.0002	-0.0004	0.0003
β	**0.9000**	0.9001	0.0007	0.8989	0.9012
σ	**0.2000**	0.1996	0.0013	0.1979	0.2013
γ	**1.1000**	1.1009	0.0008	1.0992	1.1022
ω	**0.0000**	-0.0001	0.0001	-0.0002	0.0002
β	**0.9000**	0.8997	0.0078	0.8875	0.9125
σ	**0.0500**	0.0496	0.0031	0.0445	0.0549
γ	**1.1000**	1.1008	0.0011	1.0985	1.1025
ω	**0.0000**	0.0000	0.0001	-0.0003	0.0002
β	**0.9000**	0.9006	0.0129	0.8787	0.9201
σ	**0.0500**	0.04911	0.0053	0.0413	0.0581
γ	**0.8000**	0.8007	0.0007	0.7993	0.8018
ω	**0.0000**	0.0000	0.0003	-0.0004	0.0004
β	**0.9000**	0.9002	0.0012	0.8985	0.9018
σ	**0.2000**	0.1993	0.0020	0.1966	0.2027
γ	**0.8000**	0.8007	0.0006	0.7994	0.8017

3.5 Estimation results for real data

In this section, we apply EIS-ML estimation to some data sets used by Bauwens and Veredas (2004) for QML estimation, and we compare the results. The SCD model is exactly the same as in the previous section, and as in these authors' article.

The considered data sets correspond to five stocks traded at the New York Stock Exchange (NYSE): Boeing, Coca Cola, Disney, Exxon and Ibm (the last one was not used by Bauwens and Veredas). The data were extracted from the trades and quotes database (TAQ) of the NYSE and are relative to the months of September, October and November, 1996. From the original trade and quote durations, price and volume duration were computed. Price durations measure the amount of time before observing a given cumulated variation (up or down) of the price (in this case, \$1/8). Analogously, volume durations measure the amount of time necessary to observe a cumulative traded volume of a given amount (25000 shares). As it is customary in the literature on financial durations, the durations have been purged of their seasonal component. An important feature of the trade and quote data is in fact the strong seasonality featured, both on a daily and a weekly basis, by key characteristics of the duration processes. Price and volume durations feature a strong intra-day effect, being smaller at the start and at the end of the trading day than around lunch time. Moreover, this effect may depend on the day of the week. These deterministic time-of-day

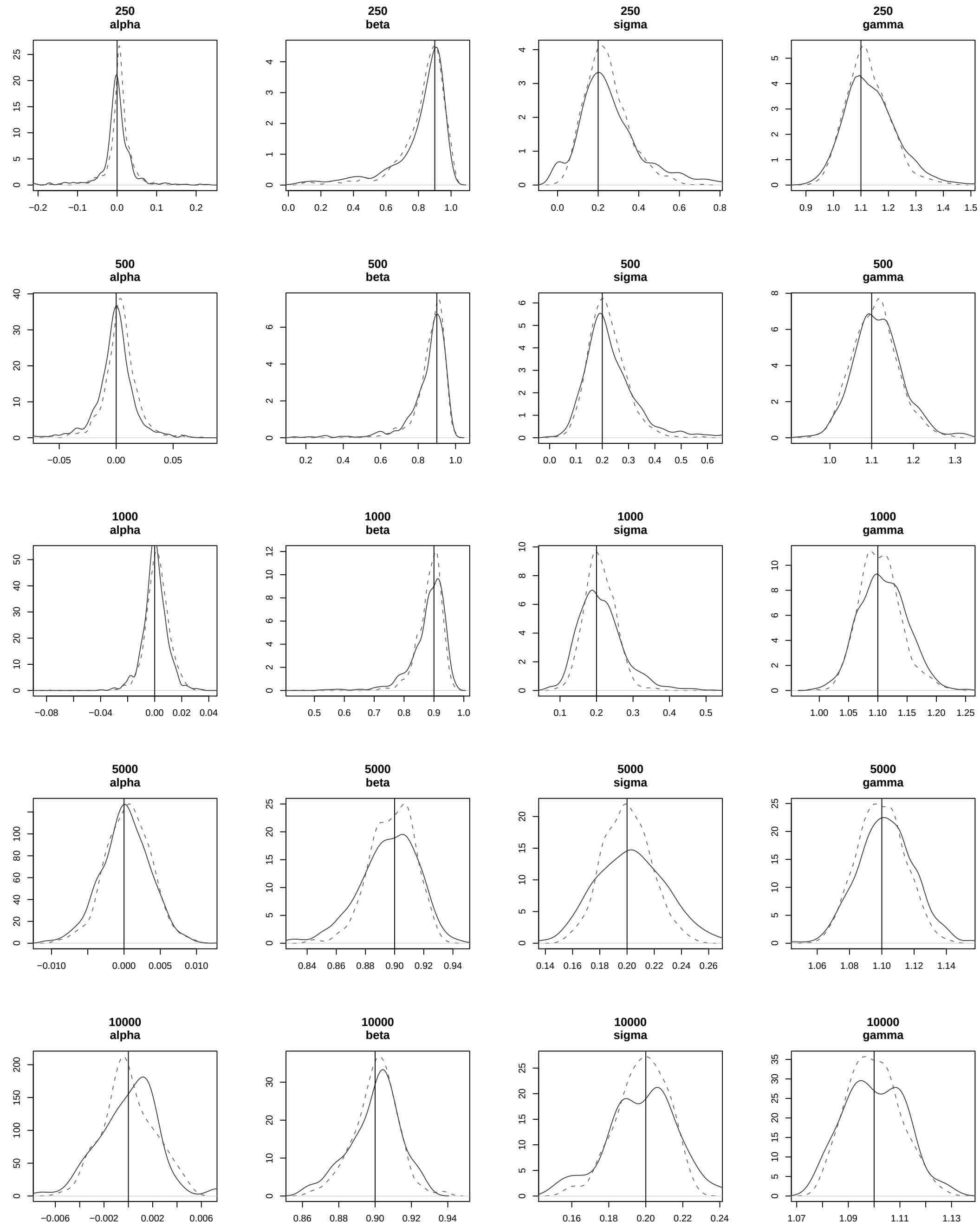

Figure 3.1: Sampling densities of 1000 EIS-ML (dashed) and QML (full) estimates of the parameters of an SCD model with parameters $\omega = 0.0$, $\beta = 0.9$, $\sigma = 0.2$, $\gamma = 1.1$.

and day-of-the week effect are controlled by regressing with a Nadaraya-Watson estimator the observed durations of each day of the week on the time of the day, and by defining the

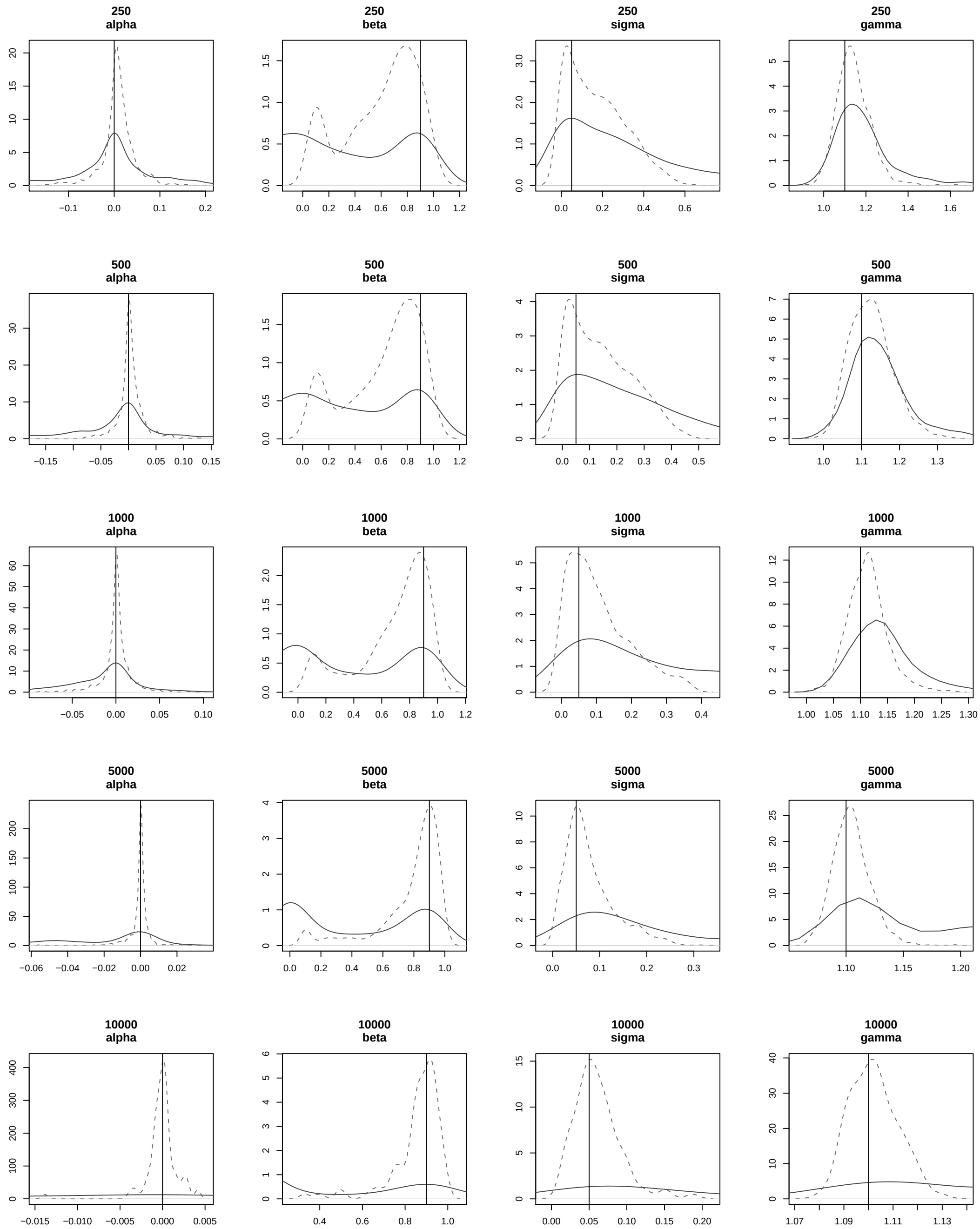

Figure 3.2: Sampling densities of 1000 EIS-ML (dashed) and QML (full) estimates of the parameters of an SCD model with parameters $\omega = 0.0$, $\beta = 0.9$, $\sigma = 0.05$, $\gamma = 1.1$.

deseasonalized durations as the original ones divided by the fitted values of the regression. For further details of the treatment of raw data, the reader can refer to the corresponding

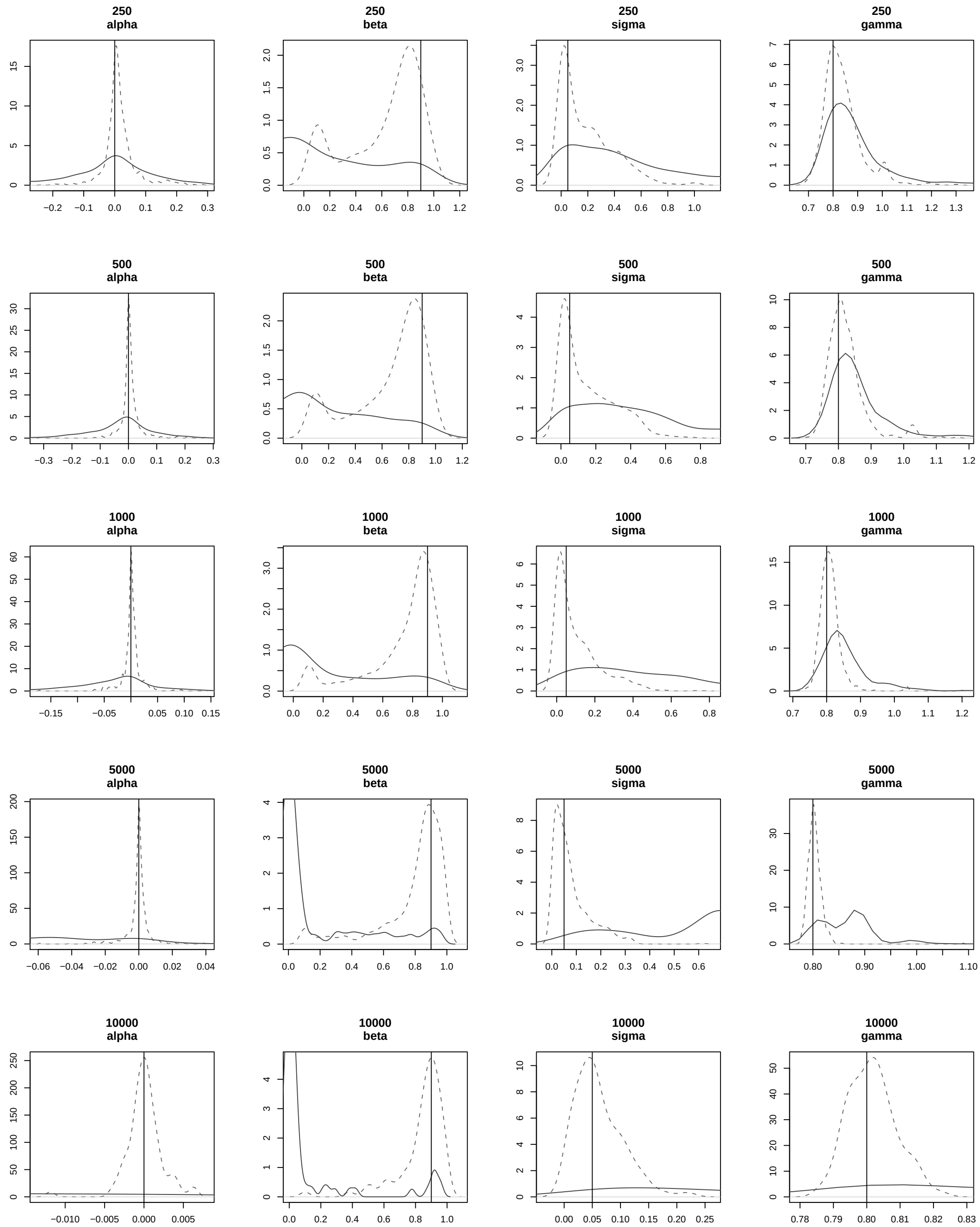

Figure 3.3: Sampling densities of 1000 EIS-ML (dashed) and QML (full) estimates of the parameters of an SCD model with parameters $\omega = 0.0$, $\beta = 0.9$, $\sigma = 0.05$, $\gamma = 0.8$.

section in Bauwens and Veredas (2004).

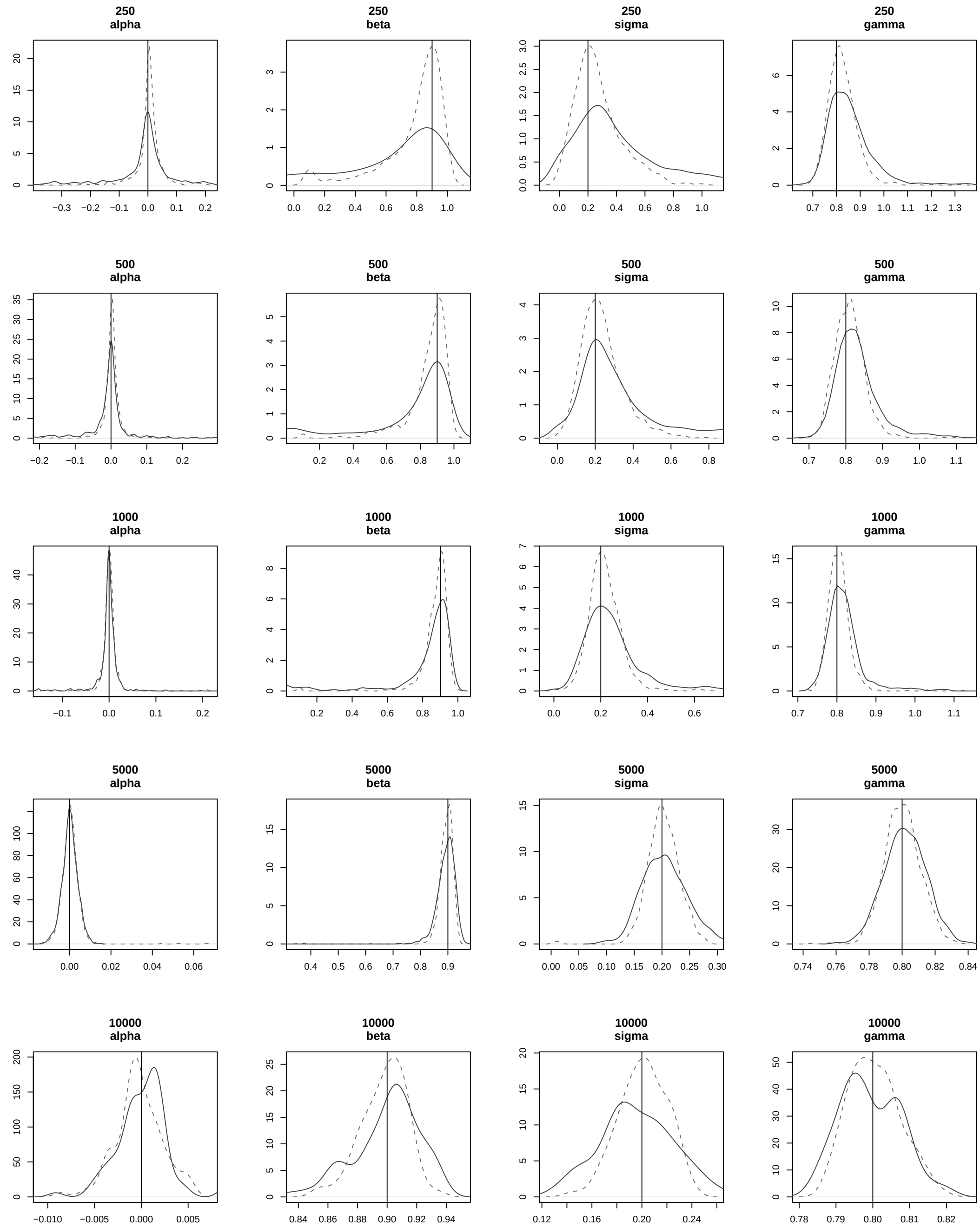

Figure 3.4: Sampling densities of 1000 EIS-ML (dashed) and QML (full) estimates of the parameters of an SCD model with parameters $\omega = 0.0$, $\beta = 0.9$, $\sigma = 0.2$, $\gamma = 0.8$.

In tables 3.7 and 3.8 we present the estimated parameters of the QML and EIS-ML estimations. For volume duration, even if the values taken by the estimates of the parameters for

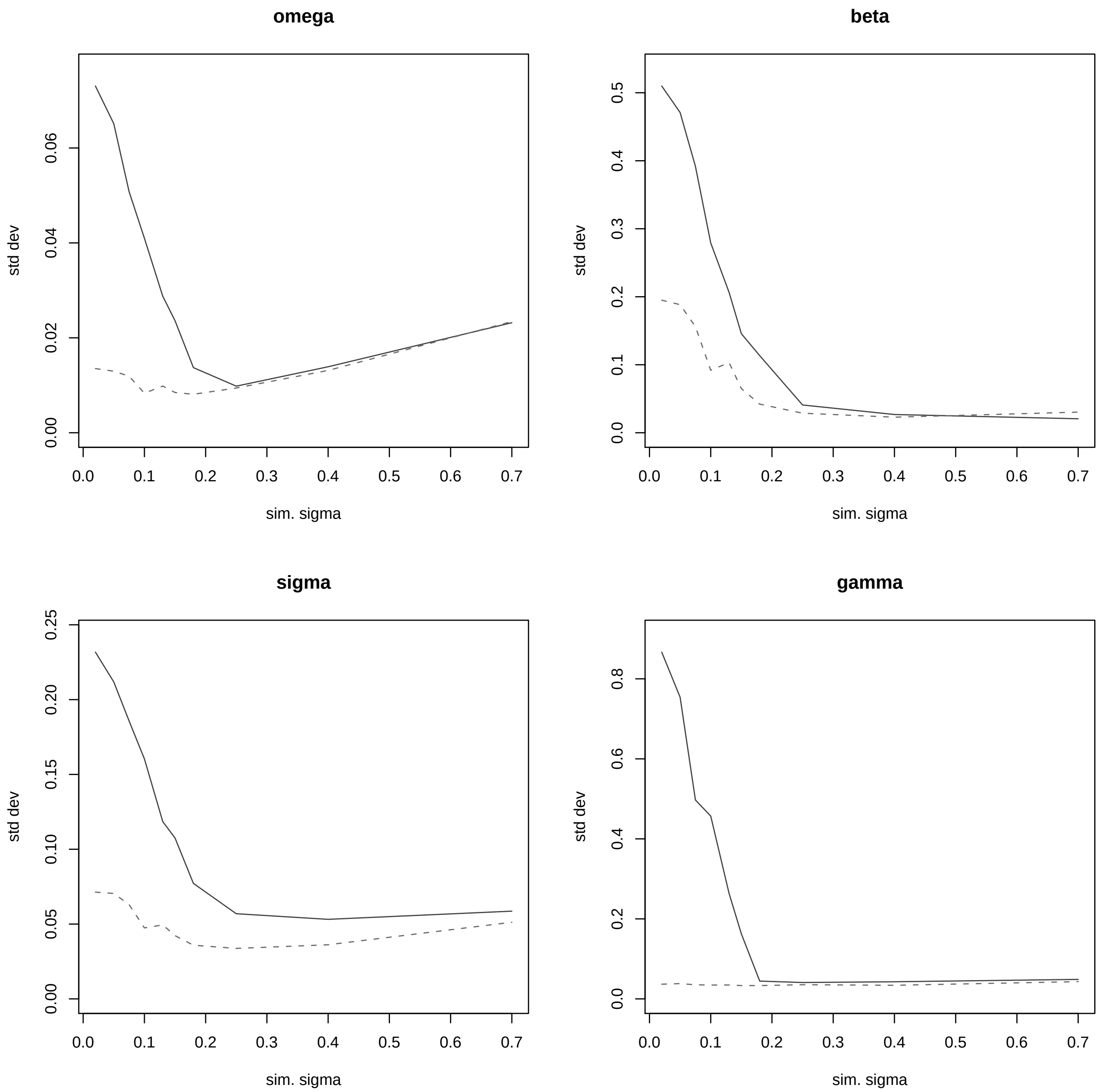

Figure 3.5: Standard errors of 1000 estimates of the SCD model parameters as a function of the DGP value of σ. Case of 1000 observations simulated from the DGP with parameters $\omega = 0.0$, $\beta = 0.9$, $\gamma = 1.1$.

the two methods considered are substantially the same, it is noticeable that the standard errors of EIS-QML are generally lower, sometimes substantially, than the QML ones. We also report in the table the computed dispersion index ($\tilde{\delta}_x$) of the durations implied by the estimates. This is computed by plugging in formula (3.2) the point estimates. We see that the data dispersion index is better approximated if we estimate with the EIS-ML algorithm. The estimates and standard errors for the price duration data are more markedly different between the two methods than for volume durations. Moreover, the improved match between the data dispersion index and the implied one of the EIS-ML estimates is even more striking in this case. To draw a conclusion, provided that the model is correctly specified, by using EIS-ML estimation rather than QML, the estimates one gets are more precise.

Finally, in both tables, in square brackets, we present the MC standard deviations of the EIS estimates. These standard deviations are calculated from ten different estimates obtained by running the algorithm using each time a different random seed for the common random numbers employed in the EIS evaluation of the likelihood. The resulting dispersion of the estimates is extremely low, which suggests that EIS is a rather robust method in this respect.

Table 3.7: Results for volume durations for an SCD model

VOLUME		Boeing	Coca Cola	Disney	Exxon	Ibm
obs.		1576	3022	1778	2045	4305
ω	*qml*	*0.002*	*-0.001*	*0.000*	*0.007*	*0.000*
		(0.0037)	*(0.0053)*	*(0.0029)*	*(0.0044)*	*(0.0016)*
	eis	0.002	0.002	0.002	0.006	-0.002
		(0.0026)	(0.0021)	(0.0018)	(0.0030)	(0.0010)
		[<0.0001]	[<0.0001]	[0.0001]	[<0.0001]	[0.0001]
β	*qml*	*0.921*	*0.865*	*0.976*	*0.901*	*0.982*
		(0.0129)	*(0.0276)*	*(0.0074)*	*(0.0237)*	*(0.0044)*
	eis	0.961	0.950	0.981	0.925	0.982
		(0.0116)	(0.0129)	(0.0061)	(0.0218)	(0.0032)
		[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]	[0.0005]
σ	*qml*	*0.116*	*0.209*	*0.102*	*0.133*	*0.101*
		(0.0168)	*(0.0278)*	*(0.0117)*	*(0.0206)*	*(0.0020)*
	eis	0.101	0.109	0.088	0.102	0.112
		(0.0143)	(0.0159)	(0.0101)	(0.0181)	(0.0081)
		[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]	[0.0002]
γ	*qml*	*1.698*	*1.401*	*1.767*	*1.812*	*1.892*
		(0.0288)	*(0.0219)*	*(0.0259)*	*(0.0310)*	*(0.0238)*
	eis	1.713	1.310	1.777	1.764	1.837
		(0.0384)	(0.0215)	(0.0368)	(0.0375)	(0.0266)
		[0.0001]	[<0.0001]	[0.0006]	[<0.0001]	[0.0005]
eis log lik		-1303.51	-2843.79	-1381.95	-1682.94	-3158.13
		[0.0186]	[0.0062]	[0.0450]	[0.0116]	[0.5885]
$\tilde{\delta}_x$	data	0.70	0.88	0.72	0.65	0.80
	qml	0.76	0.90	0.82	0.67	0.86
	eis	0.74	0.89	0.80	0.66	0.94

QML and EIS-ML estimates and standard errors in parentheses. MC standard deviations for the EIS estimates are in square parentheses. $\tilde{\delta}_x$ denotes the dispersion index. The estimated model is defined by (3.1)-(3.4) and (3.23).

Table 3.8: Results for price durations for an SCD model

PRICE		Boeing	Coca Cola	Disney	Exxon	Ibm
obs.		2620	1609	2160	2717	6728
ω	*qml*	*-0.026*	*-0.035*	*-0.005*	*0.008*	*-0.005*
		(0.0081)	*(0.0166)*	*(0.0030)*	*(0.0047)*	*(0.0020)*
	eis	-0.023	-0.027	-0.002	-0.127	-0.006
		(0.0097)	(0.0154)	(0.0033)	(0.0243)	(0.0028)
		[0.0001]	[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]
β	*qml*	*0.896*	(0.774)	*0.967*	*0.921*	*0.977*
		(0.0194)	*(0.0770)*	*(0.0103)*	*(0.0356)*	*(0.0051)*
	eis	0.876	0.733	0.960	0.179	0.962
		(0.0302)	(0.0731)	(0.0134)	(0.0564)	(0.0074)
		[0.0004]	[0.0006]	[<0.0001]	[0.0005]	[<0.0001]
σ	*qml*	*0.286*	*0.292*	*0.108*	*0.100*	*0.135*
		(0.0301)	*(0.0739)*	*(0.0181)*	*(0.0320)*	*(0.0041)*
	eis	0.332	0.377	0.136	0.674	0.192
		(0.0487)	(0.0698)	(0.0247)	(0.0344)	(0.0197)
		[0.0008]	[0.0006]	[<0.0001]	[0.0002]	[0.0002]
γ	*qml*	*1.149*	*1.113*	*1.177*	*1.161*	*1.244*
		(0.0200)	*(0.0308)*	*(0.0192)*	*(0.0175)*	*(0.0131)*
	eis	1.067	1.113	1.056	1.344	1.130
		(0.0284)	(0.0402)	(0.0208)	(0.0493)	(0.0159)
		[0.0004]	[0.0003]	[<0.0001]	[0.0003]	[<0.0001]
eis log lik		-2371.96	-1561.44	-2056.21	-2635.10	-5760.00
		[0.0336]	[0.0192]	[0.0041]	[0.0663]	[0.0404]
$\tilde{\delta}_x$	data	1.36	1.21	1.23	1.23	1.43
	qml	1.29	1.07	1.03	0.93	1.21
	eis	1.42	1.21	1.19	1.23	1.40

QML and EIS-ML estimates and standard errors in parentheses. MC standard deviations for the EIS estimates are in square parentheses. $\tilde{\delta}_x$ denotes the dispersion index. The estimated model is defined by (3.1)-(3.4) and (3.23).

3.6 Extensions

In order to illustrate the flexibility of EIS-ML as a numerical tool, we estimate two extensions of the SCD model described and used in the previous sections. The first extension consists in the introduction of a "leverage" term in the mean of the latent factor. For a motivation of this effect, we refer the reader to Feng et al. (2004), but we slightly differ from these authors by letting the value of ψ_i to depend on the lagged duration x_{i-1} (rather than ϵ_{i-1}), such that equation (3.2) becomes

$$\psi_i = \omega + \beta\psi_{i-1} + \alpha x_{i-1} + u_i. \tag{3.28}$$

The introduction of the lagged observed duration requires just a slight modification of the code and the effect on the speed of the algorithm is negligible: the EIS based computation of the likelihood takes almost exactly the same time while of course the introduction of an extra parameter slows down the maximization routine. Table 3.9 presents the results of the estimation with the data sets employed in the previous section. The introduction of the lagged duration as an explanatory variable can be tested by the likelihood ratio for the null hypothesis $\alpha = 0$. The p-values are reported in the table in curled brackets before the value of the likelihood at its maximum.

For volume durations, the estimated leverage coefficients do not display a clear sign pattern. Moreover, the results of the LR tests are mixed: we clearly reject the null only for Exxon, with a positive leverage effect, while in three other cases the evidence is mixed (p-value around 0.05) and the effect is negative. In the case of Boeing, the leverage effect is clearly not significant. For price durations the results are clearly in favor of a negative leverage effect, except for the puzzling case of the Exxon stock, where the estimates are somewhat unusual (low β and high σ).

In the second variant of the model we change the distribution of variable ϵ_i, representing the baseline duration. The Weibull distribution is replaced by a generalized gamma one with parameters (ν, γ, c). The third parameter is a location parameter and, like in the case of the Weibull, it is chosen so that the random variable has a unitary mean (so that c is a function of ν and γ and therefore does not appear in the parameters to estimate). The density function of the generalized gamma is as follows:

$$f_{GG}(\epsilon) = \frac{\gamma}{c^{\nu\gamma}\Gamma(\nu)}\epsilon^{\nu\gamma-1}\exp\left[-\left(\frac{\epsilon}{c}\right)^{\gamma}\right], \tag{3.29}$$

and it can be easily seen that the Weibull density is a particular case, arising when $\nu = 1$. Further information about this distribution is available in Bauwens and Giot (2001), who provide a detailed description of its characteristics.

The modifications in the computer code that were required to use this extension were even simpler than in the case of the leverage effect. We did not observe any speed impact on the likelihood computation, while, of course, its maximization was a tad slower because of the introduction of an extra parameter. The estimation results are available in table 3.10. Volume durations modeling does not appear to improve consistently with the introduction of this richer baseline density. The p-values of the LR tests for the null hypothesis $\nu = 1$ are seldom low (except for Ibm) as the values taken by the parameter ν tend to be rather close to 1. Different results are obtained with price durations. A significant departure from the Weibull is observed (estimates of ν are between 4.7 to 7.2) and gains in likelihood are consistent, to the point that the p-values of the LR tests for $\nu = 1$ are always smaller than 0.001. For the price durations of the Exxon stock the EIS-ML algorithm delivers a value for the likelihood but the maximization routine failed to achieve strong convergence, regardless of the vector of initial parameters chosen as a starting point.

Table 3.9: Results for volume and price durations for an SCD model with leverage

VOLUME	Boeing	Coca Cola	Disney	Exxon	Ibm
obs.	1576	3022	1778	2045	4305
ω	-0.008	0.028	0.012	-0.065	0.004
	(0.009)	(0.016)	(0.007)	(0.021)	(0.002)
β	0.954	0.972	0.991	0.867	0.988
	(0.016)	(0.014)	(0.005)	(0.031)	(0.002)
σ	0.096	0.117	0.088	0.043	0.115
	(0.014)	(0.018)	(0.010)	(0.030)	(0.010)
γ	1.700	1.330	1.78	1.674	1.847
	(0.039)	(0.025)	(0.037)	(0.040)	(0.028)
α	0.011	-0.028	-0.013	0.079	-0.007
	(0.010)	(0.016)	(0.007)	(0.024)	(0.002)
LR p-value	0.289	0.068	0.041	<0.001	0.049
EIS log-lik	-1302.96	-2842.12	-1379.86	-1676.29	-3156.17

PRICE	Boeing	Coca Cola	Disney	Exxon	Ibm
obs.	2620	1609	2160	2717	6728
ω	0.049	0.119	0.041	-0.120	0.009
	(0.011)	(0.034)	(0.007)	(0.077)	(0.002)
β	0.937	0.883	0.986	0.193	0.971
	(0.023)	(0.055)	(0.005)	(0.153)	(0.006)
σ	0.331	0.392	0.166	0.673	0.205
	(0.043)	(0.056)	(0.038)	(0.038)	(0.018)
γ	1.098	1.192	1.095	1.344	1.143
	(0.027)	(0.042)	(0.022)	(0.050)	(0.015)
α	-0.063	-0.133	-0.048	-0.005	-0.016
	(0.011)	(0.029)	(0.009)	(0.049)	(0.003)
LR p-value	<0.001	<0.001	<0.001	0.924	<0.001
EIS log-lik	-2361.79	-1552.83	-2041.77	-2635.10	-5748.14

EIS-ML estimates of the parameters and standard errors in parentheses. The LR p-value is for the hypothesis $\alpha = 0$. The estimated model is defined by (3.1), (3.28), (3.3), (3.4) and (3.23).

Table 3.10: Results for volume and price durations for an SCD model with a generalized gamma baseline duration

VOLUME	Boeing	Coca Cola	Disney	Exxon	Ibm
obs.	1576	3022	1778	2045	4305
ω	0.006 (0.005)	-0.011 (0.008)	0.002 (0.002)	-0.001 (0.008)	-0.006 (0.002)
β	0.960 (0.012)	0.953 (0.012)	0.981 (0.006)	0.930 (0.021)	0.982 (0.004)
σ	0.104 (0.016)	0.104 (0.016)	0.089 (0.010)	0.098 (0.018)	0.104 (0.008)
γ	1.900 (0.198)	1.126 (0.092)	1.819 (0.163)	1.633 (0.155)	1.437 (0.088)
ν	0.854 (0.130)	1.274 (0.166)	0.964 (0.133)	1.128 (0.165)	1.489 (0.149)
LR p-value	0.307	0.054	0.792	0.405	<0.001
EIS log-lik	-1302.99	-2841.94	-1381.92	-1682.59	-3149.46
δ_x data	0.70	0.88	0.72	0.65	0.80
δ_x eis	0.74	0.91	0.81	0.67	0.90

PRICE	Boeing	Coca Cola	Disney	Exxon	Ibm
obs.	2620	1609	2160	2717	6728
ω	-0.377 (0.151)	-0.666 (0.292)	-0.170 (0.077)	na	-0.105 (0.029)
β	0.923 (0.020)	0.828 (0.044)	0.970 (0.010)	na	0.982 (0.004)
σ	0.235 (0.367)	0.250 (0.046)	0.106 (0.019)	na	0.123 (0.013)
γ	0.389 (0.061)	0.440 (0.080)	0.0365 (0.057)	na	0.373 (0.031)
ν	5.631 (1.665)	4.717 (1.583)	6.894 (2.058)	na	7.167 (1.151)
LR p-value	<0.001	<0.001	<0.001	na	<0.001
EIS log-lik	-2335.20	-1542.15	-2009.12	na	-5617.91
δ_x data	1.36	1.21	1.23	1.23	1.43
δ_x eis	1.57	1.32	1.33	na	1.52

EIS-ML estimates of the parameters and standard errors in parentheses. The LR p-value is for the hypothesis $\nu = 1$. The estimated model is defined by (3.1)-(3.4) and (3.29). For Exxon price durations, results are not available (na).

3.7 Conclusion

This paper describes a new approach, the EIS-ML, for the numerical estimation of SCD models, showing its capability to deliver more precise estimates than the approximate QML method. The performance EIS-ML is tested both in a simulated and a real case, giving good results in terms of precision of estimation at a cost of an acceptable loss in rapidity of the computation. The evidence from the estimation of simulated series suggests an uncomfortably poor performance of the QML estimator when the latent factor variance is low, while EIS-ML appears to be much more robust. The algorithm is applied to the estimation of a set of volume and price durations showing a remarkably low MC variance. Moreover, we observed a significant gain in the capability of the SCD model in reproducing the first two empirical moments of the data when QML is replaced by EIS-ML estimation. Further interesting extensions are explored in the use of EIS-ML for the estimation of richer specifications, such as the presence of a leverage effect in the autoregressive component or a more flexible distribution for the baseline duration.

3.8 Appendix

Proof: [Theorem 3]
Given (3.24) and (3.25), the function $k(\psi_i, a_i)$ can be written as follows.

$$
\begin{aligned}
k(\psi_i, a_i) &= p(\psi_i|\psi_{i-1}, a_i)\zeta(\psi_i, a_i) \\
&= \frac{1}{\sigma\sqrt{2\pi}} \exp\Bigg\{ -\frac{1}{2}\Bigg[\left(\frac{1}{\sigma^2} - 2a_{2,i}\right)\psi_i^2 - 2\left(\frac{\omega + \beta\psi_{i-1}}{\sigma^2} + a_{1,i}\right)\psi_i \\
&\quad + \left(\frac{\omega + \beta\psi_{i-1}}{\sigma}\right)^2 \Bigg]\Bigg\}.
\end{aligned} \tag{3.30}
$$

Integrating $k(\psi_i, a_i)$ with respect to ψ_i, we obtain the function $\chi(\psi_{i-1}, a_i)$, as in (3.21).

$$
\begin{aligned}
\chi(\psi_{i-1}, a_i) &= \int_{-\infty}^{\infty} k(\psi_i, a_i) d\psi_i \\
&= \frac{1}{\sqrt{1 - 2\sigma^2 a_{2,i}}} \exp\Bigg\{ \frac{\sigma^2}{1 - 2\sigma^2 a_{2,i}} \left(\frac{\omega + \beta\psi_{i-1}}{\sigma_2} + a_{1,i}\right)^2 \\
&\quad - \frac{1}{2}\left(\frac{\omega + \beta\psi_{i-1}}{\sigma_2}\right)^2 \Bigg\}
\end{aligned} \tag{3.31}
$$

If we combine (3.30) and (3.31), as in (3.21) again, we can find the functional form for $m(\psi_i|\psi_{i-1}, a_i)$, that can be easily rewritten as a normal with mean and variance equal to μ_i and v_i^2 as in (3.26).

$$
\begin{aligned}
m(\psi_i|\psi_{i-1}, a_i) &= \frac{k(\psi_i, a_i)}{\chi(\psi_{i-1}, a_i)} \\
&= \frac{\sqrt{1 - 2\sigma^2 a_{2,i}}}{\sigma\sqrt{2\pi}} \exp\Bigg\{ -\frac{1}{2}\Bigg[\left(\frac{1}{\sigma^2} - 2a_{2,i}\psi_i^2\right) \\
&\quad - 2\left(\frac{\omega + \beta\psi_{i-1}}{\sigma^2} - a_{1,i}\right)\psi_i \\
&\quad + \frac{\sigma^2}{1 - 2\sigma^2 a_{2,i}} \left(\frac{\omega + \beta\psi_{i-1}}{\sigma_2} + a_{1,i}\right)^2 \Bigg]\Bigg\} \\
&= \frac{1}{v_i\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left[\frac{1}{v_i^2}\psi_i^2 - 2\frac{\mu_i}{v_i^2}\psi_i + \frac{\mu_i^2}{v_i^2}\right]\right\}.
\end{aligned} \tag{3.32}
$$

◇

Chapter 4

A Nonparametric ACD Model

4.1 Introduction

In the ACD literature, the variety of parametric specifications has not been matched, so far, by any attempt to provide a generic form for the autocorrelated factor, which would have the advantage to be robust to misspecification and to be able to provide an estimation which is sufficiently reliable in most cases. The aim of the work in this chapter is to introduce a generic form for the ACD family model, where the autocorrelated factor is expressed as a function of the lagged observation and of the factor itself, and it is nonparametrically estimated. Moreover, the hypotheses of the model that we propose are not particularly strict even on the functional form of the distribution of the conditional duration, that is implicitly estimated in a non parametric way, yielding a more generic form than any parametric one employed in the literature. This could be helpful because, as it has been noticed by Bauwens et al. (2004), more complex specifications of the autocorrelated factor do not seem to provide substantial improvements in the goodness of fit, raising therefore the suspicion that it is in the conditional duration that improvements could be sought for.

Up to the our, the ACD literature does not provide examples of semi or nonparametric analysis. Departures from a full parametric specification exist instead in the literature about ARCH processes, whose modeling shares a great deal of commonalities with the ACD framework. Engle and Gonzalez-Rivera (1991), for instance, propose a semiparametric estimation of GARCH models, and the autocorrelated component retains its linear original form.

The main difficulty of estimating GARCH or ACD models in a fully nonparametric way resides in the unobservability of one or some regressors. In order to overcome this difficulty, various solutions have been proposed. Hafner (1997), proposes to replace the unobservable regressor with a function of several lagged values of the observations only. This approach yields an easy approximated model, but because of the large number of regressors it is computationally heavy and severely suffers from the curse of dimensionality. Another interesting solution comes from Franke and Muller (2002) and Franke et al. (2004), who employ a deconvolution kernel estimator, that relies strongly on normality of the innovations (which means that it would be hardly adaptable to an ACD framework) and has a rather slow rate of convergence. A solution more easily adaptable to the ACD structure consists instead in the iterative scheme proposed by Bühlmann and McNeil (2002). Under a central, and albeit rather restrictive, contraction hypothesis, the estimation algorithm can be proven to be consistent and to have a rate of estimation accuracy of the order of a usual bivariate nonparametric regression technique, which means that it performs better than the deconvolution kernel and does not represent an approximation.

The advantages of the specification and estimation technique that are proposed here, and that rely strongly on the results of Bühlmann and McNeil (2002), are, in principle, rather significant. Apart from the great flexibility that it guarantees for the autocorrelated factor functional form, this specification is also less prone to suffer from an incorrect hypothesis on the distribution of the conditional duration, as the only hypothesis it relies on is that its realizations are independent and have mean equal to one. On the other hand, its main cost is that the exact role played by an independent variable in the model cannot be summarized in a single vector of parameters, and this limits the scope for inference.

The outline of this chapter [1] is as follows: Section 2 will display the main characteristics and properties of the specification and of the estimation techniques that are used, a Monte Carlo experiment is conducted in Section 3 on a series of simulated processes, to compare the performance of the nonparametric estimator and of the ML one employed in parametric formulations under both correct and incorrect specification, Section 4 is characterized by the estimation of a financial dataset that is commonly used in the ACD literature, followed by some goodness-of-fit comparisons, Section 5 presents the evaluation of the shocks impact curve calculatet on the basis of a nonparametric estimation and Section 6 concludes.

4.2 The Theoretical framework

4.2.1 The Model

We introduce in this section the ACD model in the form that can be usually found in the literature, and then rewrite it in a way that will allow us to estimate it nonparametrically. Let $\{X_t\}$ be a nonnegative stationary process adapted to the filtration $\{\mathcal{F}_t,\, t \in \mathbb{Z}\}$, with

[1] This chapter is the result of a joint work with Antonio Cosma, Université du Luxembourg, published as CORE discussion paper, 2006/67.

$\mathcal{F}_t = \sigma(\{X_s; s \leq t\})$, and such that:

$$X_t = \psi_t \epsilon_t,$$
$$\psi_t = f(X_{t-1}, \ldots, X_{t-p}, \psi_{t-1}, \ldots, \psi_{t-q}), \tag{4.1}$$

where $p, q \geq 0$ and where $\{\epsilon_t\}$ is an *iid* nonnegative process with mean 1 and finite second moment. We assume $f(\cdot)$ to be a strictly positive function. Since $f(\cdot)$ is $\mathcal{F}_{t-1}$-measurable, we have that $\mathrm{E}(X_t|\mathcal{F}_{t-1}) = \psi_t$, i.e. ψ is the conditional mean of the process. We focus on the case where $p = q = 1$, this restriction being widely justified by empirical works. Several parameterizations of (4.1) have been introduced, the first one being the linear specification:

$$\psi_t = \omega + \alpha\psi_{t-1} + \beta X_{t-1}, \tag{4.2}$$

being followed by more complicated functional forms allowing also for nonlinearity in the response of the conditional mean to the realizations of X_t or in the autoregressive part. Most of the generalizations have been introduced in order to provide more flexibility in fitting the stylized facts of financial duration data, but not always have proven to be sufficiently general to tackle data series differing in their features. In our setup, $f(\cdot)$ is allowed to be any function of the past realization X_{t-1} and of the lagged conditional mean ψ_{t-1}. Moreover, parametric specifications of the ACD family often make use of highly parameterized functions for the distributions of the innovations ϵ_t, while here we only ask for the mean of the ϵ_t's to be one and for the variance to be finite. We expect our estimation to outperform parametric models in the case were the 'real' f shows some accentuated nonlinearity like in the *threshold* models:

$$\psi_t = h(\psi_{t-1}, X_{t-1}) + \sum_i \beta_i \mathbb{I}_{[X_{t-1} \in B_i]} \psi_{t-1},$$

where B_i are disjoint subsets of $\mathbb{R}_+$ and $h(x)$ again a strictly positive function.

In order to estimate f, we rewrite (4.1) in the additive form:

$$X_t = f(X_{t-1}, \psi_{t-1}) + \eta_t, \tag{4.3}$$
$$\eta_t = f(X_{t-1}, \psi_{t-1})(\epsilon_t - 1).$$

η_t is a white noise, since $\mathrm{E}(\eta_t) = \mathrm{E}(\eta_t|\mathcal{F}_{t-1}) = 0$ and $\mathrm{E}(\eta_s\eta_t) = \mathrm{E}(\eta_s\eta_t|\mathcal{F}_{t-1}) = 0$ for $s < t$. The conditional variance of X_t is $\mathrm{Var}(X_t|\mathcal{F}_{t-1}) = f^2(X_{t-1}, \psi_{t-1})(\mathrm{E}(\epsilon_t^2) - 1)$. Thus, formally, $f(X_{t-1}, \psi_{t-1})$ could be estimated by regressing X_t on $f(X_{t-1}, \psi_{t-1})$. In practice, the ψ's are unobserved variables. To overcome the problem, we adapt the recursive algorithm suggested by Bühlmann and McNeil (2002).

4.2.2 The estimation procedure

The algorithm is built as follows. Let $\{X_t; 1 \leqslant t \leqslant n\}$ be the data set. We assume that that the data generating process is of the type described by (4.1) with $p = q = 1$. The steps of the algorithm are indexed by j.

Step 1. Choose the starting values for the vector of the n conditional means. Index these values by a 0: $\{\psi_{t,0}\}$. Set j = 1.

Step 2. Regress nonparametrically $\{X_t; 2 \leqslant t \leqslant n\}$ on $\{X_{t-1}; 2 \leqslant t \leqslant n\}$ and on the conditional means computed in the previous step: $\{\psi_{t-1,j-1}; 2 \leqslant t \leqslant n\}$, to obtain an estimate $\hat{f}_j$ of f.

Step 3. Compute $\{\hat{\psi}_{t,j} = \hat{f}_j(X_{t-1}, \hat{\psi}_{t-1,j-1}); 2 \leqslant t \leqslant n\}$; remember to choose some sensible value for $\hat{\psi}_{1,j}$, that cannot be computed recursively.

Step 4. Increment j, and return to step two to run a new regression using the $\{\psi_t\}$ computed in Step 3.

A further improvement of the algorithm can often be achieved by averaging the estimates obtained in the last steps, when the algorithm becomes more stable.

A justification and theoretical discussion of the algorithm can be found in Bühlmann and McNeil (2002). We state here from the work just cited the main theorem that allows determining the convergence rates of the estimates provided by the algorithm. We first need some notation. From now on $\|Y\|$ denotes the $\mathcal{L}_2$ norm of Y: $\|Y\|^2 = \mathbb{E}(Y^2)$. Let:

$$\tilde{f}_{t,j}(x, \psi) = \mathrm{E}(X_t | X_{t-1} = x, \hat{\psi}_{t-1,j-1} = \psi),$$

$$\tilde{\psi}_{t,j} = \tilde{f}_{t,j}(X_{t-1}, \hat{\psi}_{t-1,j-1});$$

That is, $\tilde{\psi}_{t,j}$ is the true conditional expectation of X_t given the value of $\hat{\psi}_{t-1,j-1}$ estimated at the previous step of the algorithm. So the quantity:

$$\Delta_{t;j,n} \equiv \tilde{\psi}_{t,j} - \hat{\psi}_{t,j}, \;\; j = 1, 2, \ldots, t = j+2, \ldots, n,$$

gives us the estimation error introduced at the j-th step solely due to the estimation of f. In the nonparametric language, $\|\Delta\|$ is the stochastic component of the risk of the estimator $\hat{\psi}_{t,j}$ of $\mathrm{E}(X_t | X_{t-1}, \psi_{t-1,j-1})$.

Theorem 4 (Theorems 1 and 2 in Bühlmann and McNeil (1999)) *Assume that:*

1. $\sup_{x \in \mathbb{R}} |f(x, \psi) - f(x, \varphi)| \leqslant D|\psi - \varphi|$ *for some* $0 < D < 1$, $\forall \psi, \; \varphi \in \mathbb{R}_+$.
2. $\mathrm{E}|\psi_t|^2 \leq C_1, \quad \mathrm{E}|\psi_{t,0}|^2 \leq C_2, \quad \max_{2 \leq t \leq n} \mathrm{E}|\hat{\psi}_{t,0}|^2 \leq C_3, \quad C_{1,2,3} < \infty,$ $\|\psi_j - \psi_{j,0}\| < \infty, \quad \|\hat{\psi}_{j,0} - \psi_{j,0}\| < \infty \;\; \forall j.$
3. $\mathrm{E}(\{\tilde{\psi}_{t,j} - \psi_{t,j}\}^2) \leqslant G^2 \mathrm{E}(\{\hat{\psi}_{t-1,j-1} - \psi_{t-1,j-1}\}^2)$ *for some* $0 < G < 1$, *for* $t = j+2, j+3, \ldots$ *and* $j = 1, 2, \ldots$
4. $\Delta_n \doteq \sup_{j \geqslant 2} \max_{j+2 \leq t \leq n} \|\Delta_{t;j,n}\| \to 0,$ *as* $n \to \infty$ *for* $j = 1, 2, \ldots, t = j+2, \ldots, n$.

Then, if $\{X_t\}_{t\in\mathbb{N}}$ *is as in (4.1) with* $p = q = 1$*, and choosing* $m_n = C\{-\log\Delta_n\}$*:*

$$\max_{m_n+2\leqslant t\leqslant n} \|\hat{\psi}_{t,m_n} - \psi_t\| = O(\Delta_n), \ \textit{as } n \to \infty.$$

The theorem tells us that if all the assumptions hold, then the upper bound on the quadratic risk of the estimates of the $\{\psi_t\}$ is of the same order as Δ_n, that is the error of a one step nonparametric regression to estimate $\psi_{t,j}$ from (X_{t-1}, ψ_{t-1}). That is, in a bivariate nonparametric regression with an appropriate choice of the kernel function and of the smoothing parameter, and assuming for instance that $f(X_{t-1}, \psi_{t-1})$ is twice continuously differentiable, the convergence rates are $O(n^{-1/3})$. A practical choice of m_n of about $3\log(n)$ is suggested by the authors.

We briefly discuss the assumptions of the theorem. For more insights, refer to Bühlmann and McNeil (2002). First let us write:

$$\|\hat{\psi}_{t,j} - \psi_t\| \leq \|\hat{\psi}_{t,j} - \tilde{\psi}_{t,j}\| + \|\tilde{\psi}_{t,j} - \psi_{t,j}\| + \|\psi_{t,j} - \psi_t\|. \tag{4.4}$$

The first two components of the risk (4.4) are the usual quadratic risk of an estimator $\hat{\psi}_{t,j}$ of $\psi_{t,j}$. The additional component $\|\psi_{t,j} - \psi_t\|$ is due to the fact that we do not observe ψ_t. Assumption 1 controls this last part of the risk. If there were no estimation error in passing from one step to the following of the algorithm, Assumption 1 jointly with the recursive form of the algorithm would be enough to assure the convergence of $\psi_{t,m}$ to the true value ψ_t. Assumption 2 is technical and is needed to give an upper bound to the estimation error of the first step of the algorithm. Assumption 3 is used to control the second component of (4.4). It can be written in the following way:
$\|\tilde{\psi}_{t,j} - \psi_{t,j}\| = \|\mathrm{E}(X_t|X_{t-1}, \hat{\psi}_{t-1,j-1}) - \mathrm{E}(X_t|X_{t-1}, \psi_{t-1,j-1})\|$ so Assumption 3 is a contraction property of the conditional expectation with respect to $\|\hat{\psi}_{t-1,m-1} - \psi_{t-1,j-1}\|$. It is again a technical property that Bühlmann and McNeil are obliged to impose on the process in order to prove the consistency of the estimates delivered by the algorithm.

4.2.3 The practical implementation

In our application to simulated and real data we use the following settings. For the initial values of the $\{\psi_t\}$ to use in the first step of the algorithm, we choose a constant vector of the unconditional mean of the data series $\{X_t\}$. Bühlmann and McNeil suggested using a parametric estimate, to be improved in the following steps of the algorithm. Since our goal is to compare parametric with nonparametric estimates, we think that challenging the nonparametric procedure by giving dull initial values would make the competition fairer, and the results more reliable. Moreover the algorithm gives almost the same outcome in both cases, that is when providing the unconditional mean or the parametric estimate as starting values. We can say that the algorithm is quite insensitive to changes in the choice of the initial values, providing that these are sensible.

As far as the choice of the nonparametric technique is concerned, a local linear nonparametric regression is used. For an introduction to local linear fitting, one can refer to Fan and Gijbels (1996). In the bivariate case, suppose that $\{Y_t\}_{t=1,\dots,n}$ are observations of a random variable Y, and $\{(X_1, \psi_1\}, \dots, (X_n, \psi_n)\}$ the realizations of a predictor (X, ψ). Then the conditional mean of Y given (X, ψ) is given by setting $m(Y|(X, \psi) = (x, \psi)) = \hat{a}$, where $\hat{a}$ and $\hat{b}_1$, $\hat{b}_2$ minimize:

$$\sum_{t=1}^{n}(Y_t - a - b_1(X_t - x) - b_2(\psi_t - \psi))^2 \cdot K_{h_1}(X_t - x)K_{h_2}(\psi_t - \psi), \tag{4.5}$$

where K is a kernel function, $K_h(\cdot) = \frac{1}{h}K(\cdot/h)$ and h is the bandwidth. The choice of this specific method is due to the peculiar features of our data set. In our application the predictors are the lagged durations X_{t-1} and the conditional means at the j-th step of the algorithm, $\psi_{t-1,j}$. As can be seen in Figure (4.1), they form a non uniform random design in the $x\psi$ plane and are visibly more dense in the region next to the axes, drawing in the $x\psi$ plane a "falling star" pattern. So we need a design-adapted nonparametric method for estimating regression functions. Moreover both the regressors are nonnegative, and we needed an estimator free of boundary bias effect when evaluating the response function $m(Y|X, \psi)$ close to the axes x, ψ. The bias of the local linear estimator (4.5) does not depend on the marginal density of the predictors, and this holds both for points at the interior and at boundary of the predictor domain. This makes the estimator (4.5) a good candidate for our problem. Another good candidate method to estimate the model (4.3) could have been a Nadaraya-Watson type estimator built using asymmetric kernels. Fernandes and Grammig (2005) use asymmetric kernels for building specification tests for duration models, and Scaillet (2004) introduces two new families of asymmetric kernels, the Inverse Gaussian (IG) and the Reciprocal Inverse Gaussian (RIG) that share with the local linear estimator the property of having the second derivative of the conditional density $g(y|x\psi)$ in the expression of the bias instead of the first derivative. The choice was finally made in favor of the local linear method since for data sets with high noise to signal ratio, as in our data, the RIG kernel showed a higher variance when approaching the border. It is not possible to use a robust version of (4.5), consisting in minimizing $\sum_{t=1}^{n} |Y_t - a - b_1(X_t - x) - b_2(\psi_t - \psi)| \cdot K_{h_1}(X_t - x)K_{h_2}(\psi_t - \psi)$. This last method would give an estimate of the conditional median instead of the conditional mean, but since the innovations η_t in model (4.3) are asymmetric, we would introduce a systematic bias.

A practical rule for the choice of the bandwidths is needed. Simulations show that this choice can be simplified in our case. We initially fix $h_x = h_\psi = h$. Plotting against h the mean square error of the estimator obtained from simulations, we can see that the curve decreases rapidly at the beginning, to reach a minimum and remain quite flat before increasing again for very large h. This is due to the fact that for small h the variance is large like in all nonparametric regression, then it decreases as h increases until it reaches

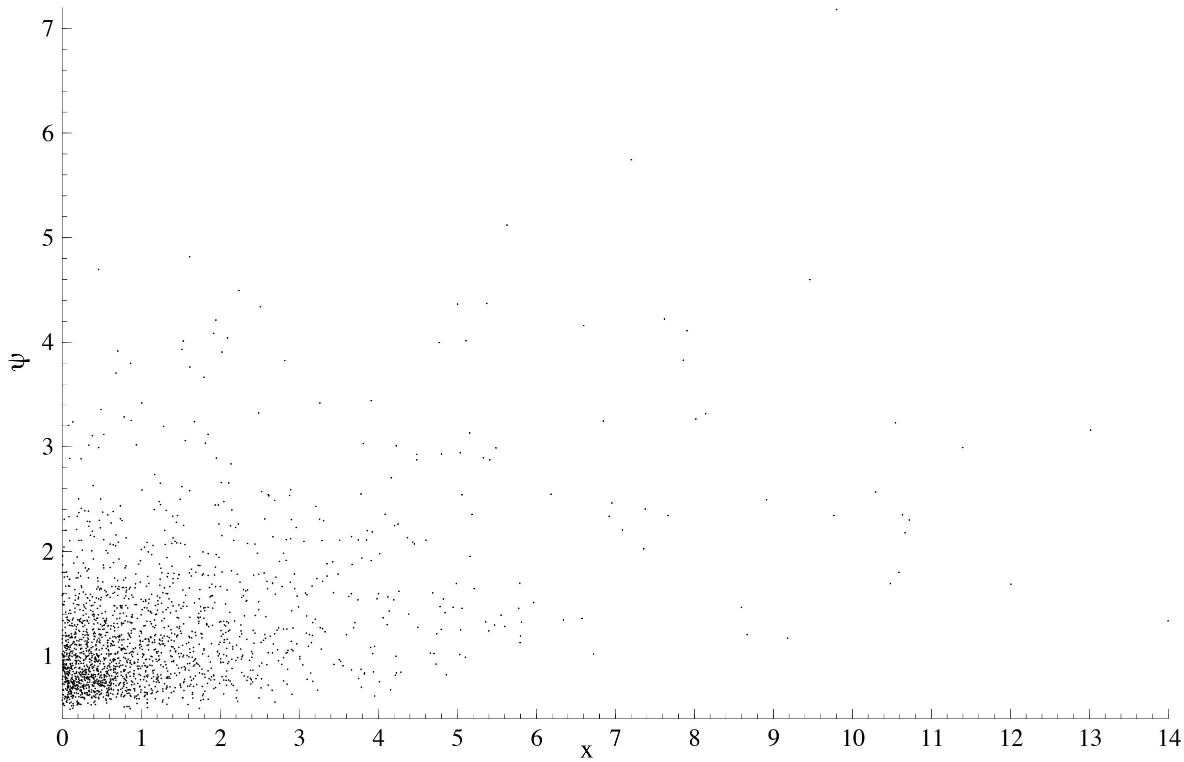

Figure 4.1: Scatterplot of a typical $x\psi$ domain

a minimum. At this point the bias should kick in, but the extra parameters b_1 and b_2 in (4.5) mitigate its growth . So the important point is to stay out of the zone where the variance is high, at the risk of oversmoothing a little. What we do is to estimate a standard ACD(1,1) model with exponential innovations. We then bootstrap 50 series of T data points, and compute the h that minimizes the MSE in the bootstrapped series. This choice is usually slightly oversmoothing with respect to the one that minimizes the MSE directly on the simulated series, but this should keep us on the save side of the minimum of the MSE curve, so we retain $h_x = h$. Since the $\{\psi_t\}$ are more dense than the $\{X_t\}$, we choose $h_\psi = 0.8h$. This too is an heuristic choice that worked well in practice.

4.3 Estimation of simulated processes

In this section an assessment of the performance of the nonparametric specification is performed via a comparison with the estimates of a linear ACD model on different simulated series. The first simulated series is characterized by an asymmetry in the conditional mean equation, which has the following form:

$$f(x_{t-1},\psi_{t-1}) = 0.2 + 0.1x_{t-1} + (0.3\mathbb{I}_{[x_{t-1}\leqslant 0.5]} + 0.85\mathbb{I}_{[x_{t-1}>0.5]})\psi_{t-1}, \tag{4.6}$$

and the conditional duration is exponentially distributed, with scale parameter such that its mean is equal to one. The size of the generated sample is of 2000 observations. We simulate 50 series from model (4.6). Figure 4.2 illustrates a window of 200 data as an example of the general appearance of the series. The simulated series have been estimated by ML with a linear ACD(1,1) specification and by the nonparametric smoother described in Section 4.2.2, with smoothing parameter $h_x = 5$, $h_\psi = 4$, 8 basic iterations and performing a final smooth based on the arithmetic mean of the last $K = 4$ iterations. The performance of the parametric and nonparametric estimators were compared by computing three widely used measures of estimation errors. The first one is the Mean Square Error (MSE), based on a quadratic loss function:

$$MSE = \frac{1}{nM}\sum_{l=1}^{M}\sum_{i=1}^{n}(\hat{\psi_{il}} - \psi_{il})^2, \tag{4.7}$$

where $i = 1,\dots,n = 2000$ denotes the $i-$th estimated conditional mean within the series, and $l = 1,\dots,M = 50$ labels the 50 series simulated from (4.6).
The second measure is the trimmed version of MSE (TMSE):

$$TMSE = \frac{1}{nM}\sum_{l=1}^{M}\sum_{i=1}^{n}\delta_i\cdot(\hat{\psi_{il}} - \psi_{il})^2. \tag{4.8}$$

$\delta_j = 0$ if $(\hat{\psi_j} - \psi_j)^2$ is in the 5% of the biggest realizations of $\{(\hat{\psi_{il}} - \psi_{il})^2, i = 1,\dots,n\}$ within one series, and $\delta_j = 1$ otherwise. TMSE eliminates the contribution of outliers to the MSE.
The third measure is the Mean Absolute Error (MAE):

$$MAE = \frac{1}{nM}\sum_{l=1}^{M}\sum_{i=1}^{n}|\hat{\psi_{il}} - \psi_{il}|. \tag{4.9}$$

Table 4.1 shows a comparison of the performance of the nonparametric and of the parametric estimators in terms of the measures just introduced. The third line reports the percentage difference between the two estimators. What is remarkable is that even if we cannot use the robust version of estimator (4.5), the nonparametric estimates have a relatively better performance in terms of MSE than in terms of TMSE. This means that the nonparametric estimator (4.5) is robust.

Figure 4.3 displays in a 200 data window an example of the evolution of the simulated ψ (hence the true dgp), and of the ones estimated parametrically and nonparametrically. We can remark that the parametric estimator seems to overreact, and make big mistakes in a

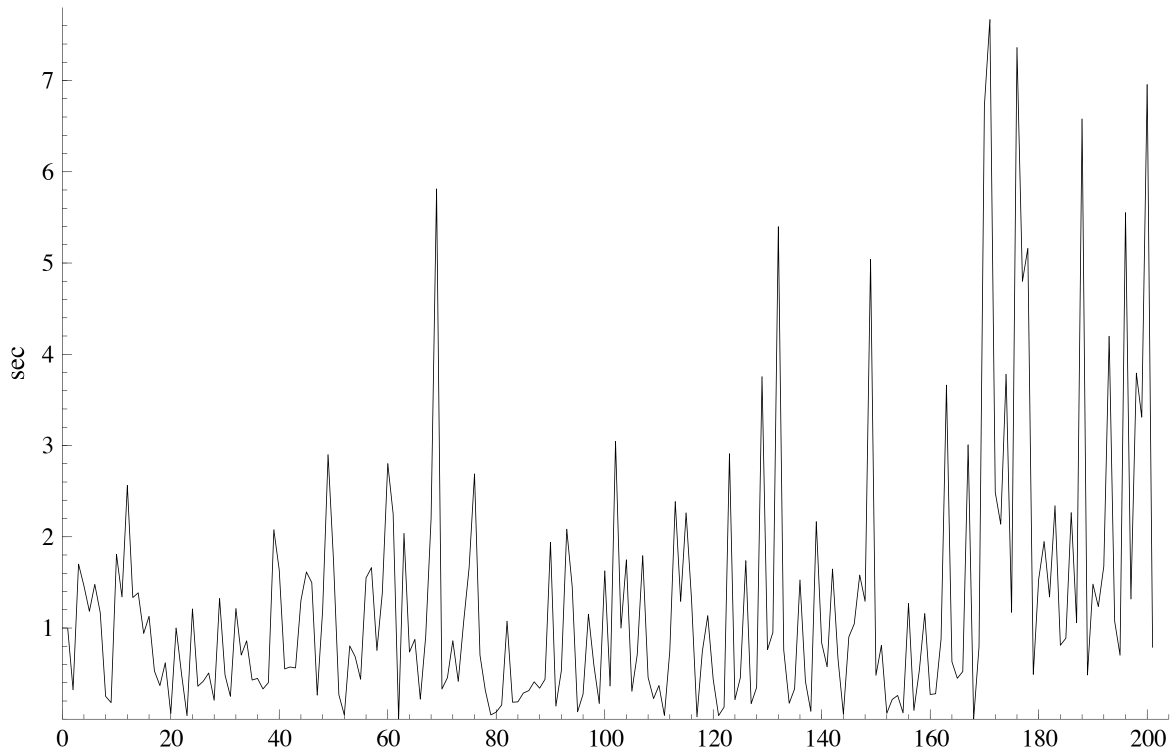

Figure 4.2: Nonlinear ACD, simulated series

small number of points. This characteristic had already been captured by the difference between the MSE and TMSE of the parametric estimator. Figure 4.4 shows the surface $f(x_{t-1}, \psi_{t-1})$ as estimated nonparametrically from one series simulated from model (4.6). It is not possible to remark in this figure an abrupt change in the slope of $f = \hat{\psi}_t(x_{t-1}, \psi_{t-1})$ as a function of ψ_{t-1} for $x \leq 0.5$ and $x > 0.5$ as specified in (4.6). Yet, it is clear that the slope increases as x increases. To make the analysis easier, in Figure 4.5 we plot the function $\hat{f}_t(x_{t-1}, \psi_{t-1})$ for two fixed values of x, $x = 0.2$ and $x = 1.6$. Even though this graph too does not display a great difference between the two slopes, if we compute them for $x = 0.2$ and $x = 1.6$, we obtain the values respectively of 0.52 and 0.58. So, even if the difference is not big, the slope is steeper for high x, as imposed in (4.6). To complete the analysis on this group of simulations, we give an example of how the nonparametric estimation evolves with the steps of the algorithm, see Table 4.2.

As it can be seen, the algorithm stabilizes already from the third step of the loop. In this particular series the last step does not display a significant improvement in the quality of the fit with respect to the previous ones. We remember that in the last step of the loop we use as regressor $\{\psi_{t-1}\}$ the arithmetic mean of the estimates $\{\hat{\psi}_{t-1}\}$ computed in the previous 4 steps of the algorithm.

	MSE	TMSE	MAE
Nonpar	0.072125	0.03098	0.195972
Par	0.100114	0.033575	0.2113
	28%	8%	7%

Table 4.1: MSE, TMSE, MAE for the Nonparametric and Parametric estimations on the asymmetric ACD simulated data.

	Nonparametric estimation	
Loop	MSE	MAE
1	0.13503	0.26955
2	0.085269	0.20679
3	0.068307	0.1834
4	0.06157	0.17614
5	0.05974	0.1752
6	0.059301	0.17494
7	0.059627	0.17591
8	0.060783	0.17618
9	0.05974	0.1752
Parametric	0.091169	0.19781

Table 4.2: Evolution of MSE and MAE for one series of simulated nonlinear ACD.

	MSE	TMSE	MAE
Nonpar	0.0098	0.006533	0.076
Par	0.002	0.00178	0.007
	79.59%	72.76%	90.79%

Table 4.3: MSE, TMSE, MAE for the Nonparametric and Parametric estimations on the standard ACD(1,1) simulated data.

We carried out the same kind of analysis on series simulated from a standard ACD(1,1) model, with no asymmetric component in the specification of the conditional mean equation. The functional form is of the conditional mean is

$$f(x_{t-1}, \psi_{t-1}) = 0.1 + 0.1x_{t-1} + 0.75\psi_{t-1}, \tag{4.10}$$

and the conditional distribution and the sample size are the same as in the first group of simulated series. The settings of the parametric and nonparametric estimators do not change from the first example. In particular, we will estimate a parametric ACD(1,1) model which this time is correctly specified. In Table 4.3 we report the values of MSE, TMSE and MAE of the nonparametric and parametric estimators. The superior performance of parametric estimates is evident, which is not surprising, as the model is correctly specified. It is not too interesting to discuss further this set of simulations, since they only confirm that a well specified parametric model clearly outperforms the nonparametric estimator.

4.4 Estimation of a financial data set

In this section, the nonparametric specification of the ACD model is tested on a set of financial data that is typically subject to analysis in the ACD family literature. The estimated series consists in a 2000 observations excerpt on the trade durations of the IBM stock traded on the New York Stock Exchange. Figure 4.6 shows a window of 200 durations of the time series considered. In Figures 4.7 and 4.8 are displayed the estimated surface $\hat{\psi}_t = f(x_{t-1}, \hat{\psi}_{t-1})$ and the curves $\hat{f}(x, \hat{\psi}_{t-1})$ at fixed $x = 0.2, 1.6, 3, 4$. Already at a first glance we can see that the slope of the surface as a function of ψ_{t-1} changes for high or low x. This is even clearer in Figure 4.8. The four curves are less apart then in the simulated case, but the difference in the slopes is evident. We have a slope of $1, 0.96, 0.88$ and 0.7 respectively for $x = 0.2, 1.6, 3, 4$. Moreover, the curves seem to have a slightly concave shape. This visual analysis suggests us that the real data generating process in the conditional mean equation puts a low weight (given by the intercepts in Figure 4.8 on the lagged observation X_{t-1}, and that the dependence of ψ_t on ψ_{t-1} diminishes with growing x. This is a reasonable feature. Let us think about a regime switching model,

	Nonparametric	Parametric	Difference
MPE	1.416277	1.415312	0.07%

Table 4.4: MPE of the nonparametric and parametric estimators.

dependent on whether the market speeds up or slows down. When the market speeds up (short durations), we are more likely to observe bunching in the data, that is, there is a bigger autocorrelation component in the conditional mean equation, and so a stronger dependence of ψ_t on ψ_{t-1}. When the market cools down, we observe less clustering in the duration data, and the autocorrelated component in the conditional mean is weaker.

We now proceed to compare the forecasting performances of the nonparametric and parametric estimators. Like all throughout this work, we use a standard ACD(1,1) specification for the parametric model. The estimation of the parametric model is carried out by maximum likelihood, using an exponential density for the innovations. We consider a series of 2100 durations of IBM. We estimate both the nonparametric and parametric models on the first 2000 observations, and forecast the conditional mean for the first observation that is not in the estimating sample. We compare the observed duration with the forecasted conditional mean, then we incorporate it in our information set to make a new one step ahead forecast. The models are not estimated again when a new duration is observed.

We compute the (one step ahead) mean prediction error (MPE) as:

$$MPR^{np,p} = \frac{1}{L}\sum_{i=1}^{L}(X_t - \hat{\psi}_t^{np,p})^2,$$

where $\hat{\psi}_t^{np,p}$ are the nonparametric and parametric forecasts of the conditional mean. The sample length for the one step ahead forecasts is $L = 100$. Table 4.4 reports the result. We can see that the parametric model has a slightly better performance, but which only amounts to a 0.7% improvement in the predicting power. A close inspection of the evolution of the MPE (not reported here) tells us that the parametric model makes big forecasts error after a change in regime, that is after the durations increase or decrease rapidly. Especially, like in the simulated series, the parametric model seems to overreact to long durations. This is a hint that a threshold ACD specification of the kind (4.6) would probably be more suited to describe the process generating the IBM data. This is in accord with the considerations made when commenting Figures 4.7 and 4.8.

4.5 Evaluation of the shocks impact curve

Engle and Russell (1998a) noticed that the ACD model had the tendency to overpredict after very long or very short durations. This would make a model with a concave shocks impact function (the ACD one is linear) better suited as a forecasting tool. The desirability for this feature was explicitely acknowledged in the subsequent literature and the Box-Cox transformation-based ACD family of specifications proposed by Fernandes and Grammig (2006) indeed show concavity in the shape of the curve. The model proposed in this paper has not an *a-priori* form for the shock impact curve given that, depending on the resulted estimated surface, the the response of the expected conditional duration to a shock in the baseline duration can vary. As an experiment, we estimate our model with the same data (quote durations for the IBM stock) used in Fernandes and Grammig (2006) and compute the resulting shocks impact curve by fixing ψ_{i-1} at 1 and letting ϵ_{t-1} vary in order to evaluate its impact on the value of the expected conditional duration ψ_i.

Figure 4.9 displays the curve resulting from the nonparametric estimation along with the one resulting from the estimation of a parametric ACD model. The result seems to confirm the hypothesis of Engle and Russell (1998a). The nonparametric estimatro in fact seems to benefit from its better flexibility and to produce a slightly concave response curve. It can be noticed too that the concavity resulting from our estimator seems less pronounced than the one observed in the estimations of the modes proposed by Fernandes and Grammig (2006), this at least on the basis of a simple visual evaluation.

4.6 Conclusion

The nonparametric specification of the ACD model encompasses most of the parametric forms so far introduced to study high frequency transaction data, the only exception being constituted by the models with two stochastic components, such as the SCD. The model can be easily estimated by standard nonparametric techniques, though a recursive approach is necessary to deal with the fact that some regressors are not directly observable. The simulated examples show that in the presence of asymmetry in the specification of the conditional mean equation the nonparametric estimator easily outperforms the symmetric parametric one. An estimation of a financial data set does not show a better performance of the nonparametric model in terms of forecasting power, since its prediction error is the same as the one of a probably misspecified standard ACD(1,1) model. Still, though not providing a specification test for parametric models, the nonparametric analysis can be useful as a benchmark in the choice of the right parametric specification. The graphical study of the dependence of the conditional mean on its lags that we carried out can give valuable information on the kind of parametric specification to choose. Also comparing the predicting performances of the nonparametric and parametric estimators can help evaluating ex post the choice of the parametric specification. For instance our analysis showed that the linear ACD(1,1) could not adapt rapidly to changes in regime of the duration process.

A final word can be spent for what could be the use of this estimation strategy in empirical analysis (for a very brief account of some of the literature see Section 1.4). Though the advantages of using a consistent estimator that is encompassing most of the specifications currently used would allow for several applications, we think that two could be the immediate fields where the nonparametric ACD could be applied. The first one would be the inclusion of calendar time as a further explanatory variable. So far, the attention in the literature on the effect of the seasonality and its possible link with other determinants of durations has been rather limited (the only exception probably being Espasa et al. (2007)) but the analysis of possible links between the determinants of durations could be an easy and interesting extension for this model. A second possible application could be the inclusion in the regression of market microstructure variable, such as volume, prices, bid-ask spread or, if available dummies about the arrival of news in the market. These variables have often been used in ACD estimations, but not always their impact on the frequency of trading stands clear and they could be easily the subject of a nonparametric or eventually, a semiparametric analysis.

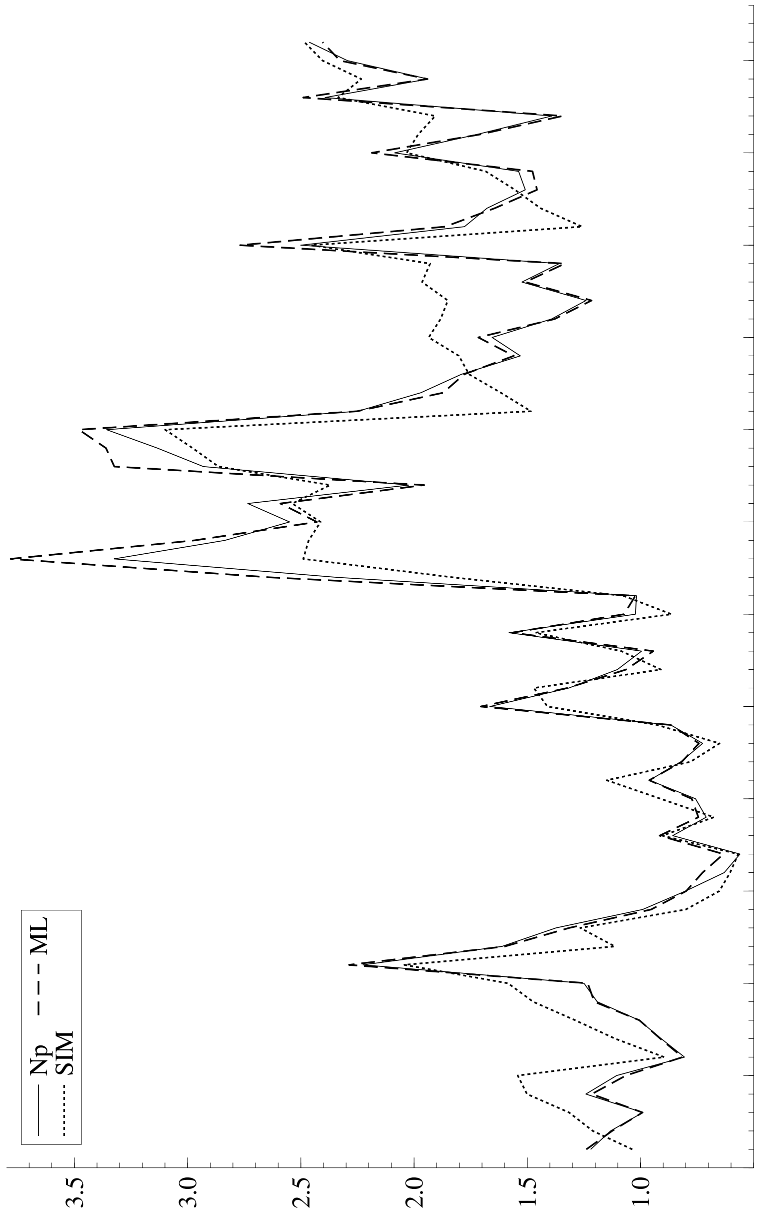

Figure 4.3: Nonlinear ACD, simulated conditional mean, parametric estimate (ML) and nonparametric estimate

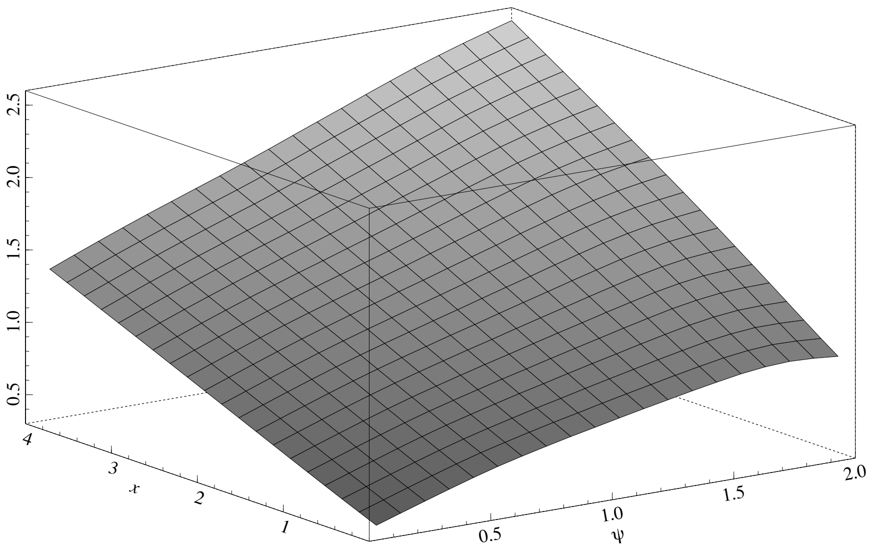

Figure 4.4: Nonparametric estimate of the surface $\hat{f}_t(x_{t-1}, \psi_{t-1})$ for a nonlinear simulated process.

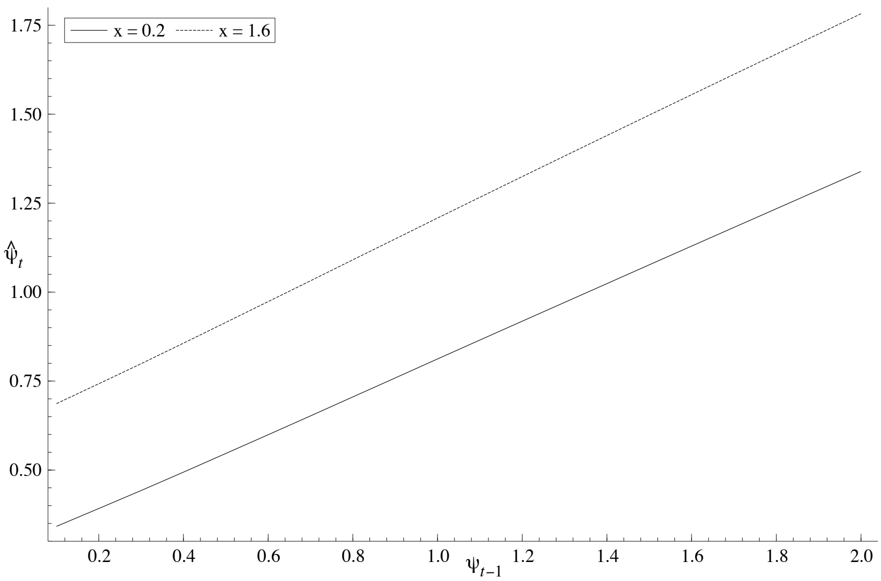

Figure 4.5: Nonparametric estimate of the curves $\hat{f}_t(x = 0.2, \psi_{t-1})$ and $\hat{\psi}_t(x = 1.6, \psi_{t-1})$ for a nonlinear simulated process.

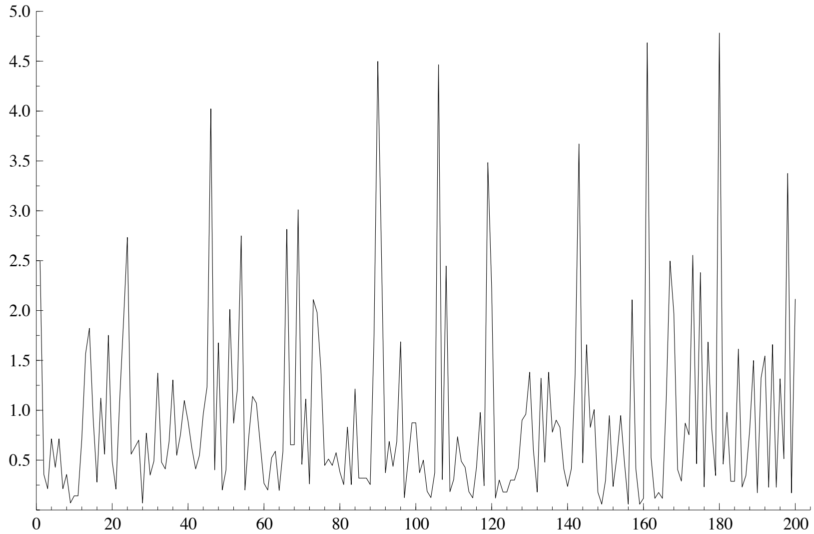

Figure 4.6: IBM trade durations.

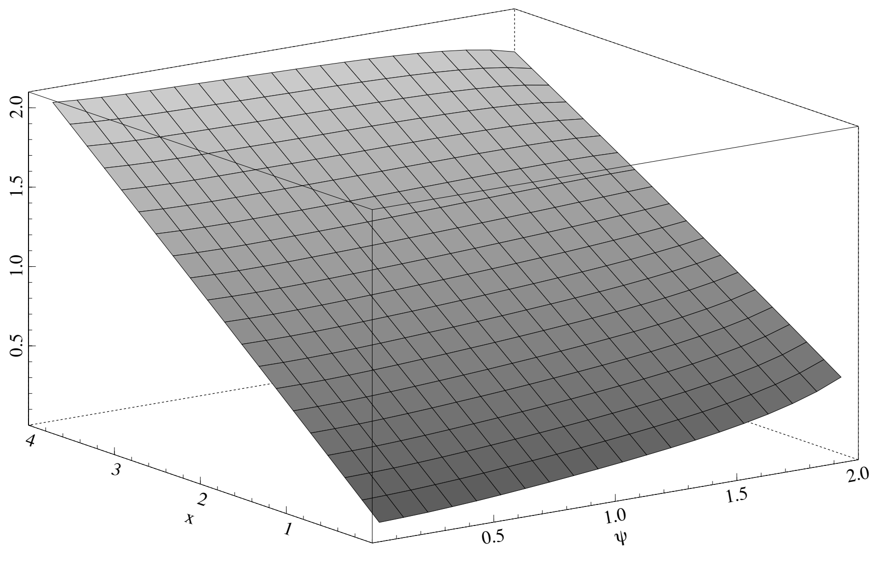

Figure 4.7: Nonparametric estimate of the surface $\hat{f}_t(x_{t-1}, \psi_{t-1})$ for the IBM durations data.

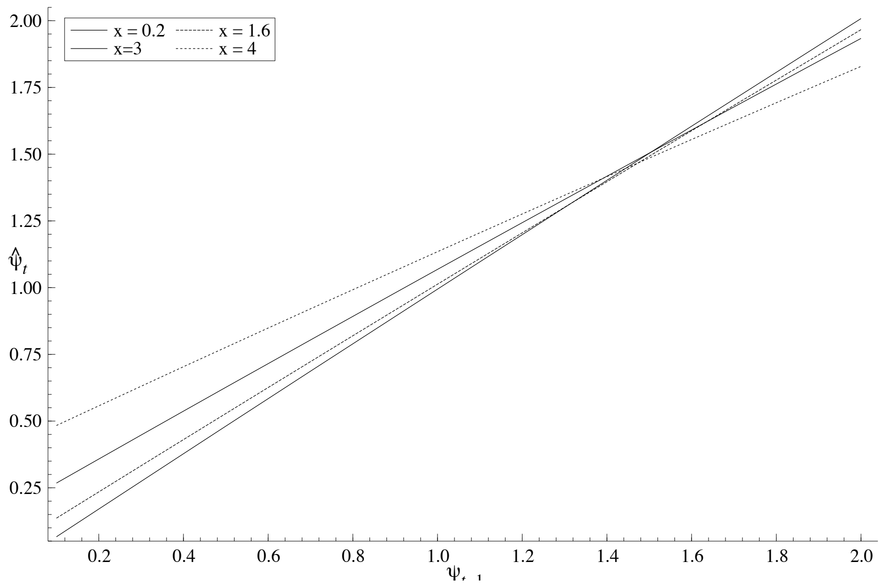

Figure 4.8: Nonparametric estimate of the curves $\hat{f}_t(x, \psi_{t-1})$, $x = 0.2, 1.6, 3, 4$, for the IBM durations data.

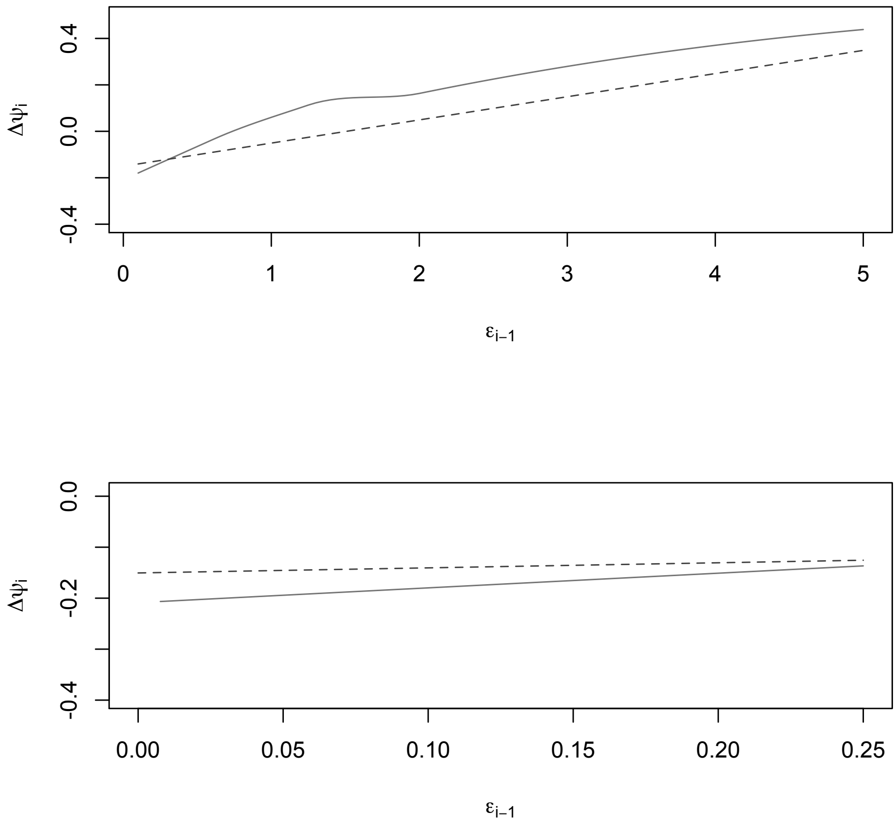

Figure 4.9: Empirical shocks impact curves of a parametric (dashed) ACD estimation and a nonparametric (solid) one.

Bibliography

Admati, A. and Pfeiderer, P. (1988). A theory of intraday patterns: Volume and price variability. *Review of Financial Studies*, 1:3–40.

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 Statistique*, 60, Special Issue, "Financial Market Microstructure":117–149.

Bauwens, L. and Giot, P. (2001). *Econometric modelling of stock market intraday activity*. Kluwer Academic Publishers.

Bauwens, L., Grammig, J., Giot, P., and Veredas, D. (2004). A comparison of financial duration models with density forecasts. *International Journal of Forecasting*, 20:589–609.

Bauwens, L. and Hautsch, N. (2006). Stochastic conditional intensity processes. *Journal of Financial Econometrics*, 4:450–493.

Bauwens, L. and Rombouts, J. (2004). Econometrics. In Gentle, J., Haerdle, W., and Mori, Y., editors, *Handbook of Computational Statistics*. Springer Verlag.

Bauwens, L. and Veredas, D. (2004). The stochastic conditional duration model: a latent factor model for the analysis of financial durations. *Journal of Econometrics*, 119(2):381–412.

Box, G. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, B 26:211–243.

Brunetti, C. and Lidholdt, P. (2002). Time series modelling of daily log-price ranges for sf/usd and usd/gbp. *Pier Working Paper 02-017*.

Bühlmann, P. and McNeil, A. J. (2002). An algorithm for nonparametric garch modelling. *Computational Statistics and Data Analysis*, 40:665–683.

Chou, H.-C. and Wang, D. (2007). Forecasting volatility of the u.k. stock market: a test of the contitional autoregressive range model. *International Research Journal of Finance and Economics*, 10:7–13.

Chou, R. Y. (2005). Forecasting financial volatilities with extreme values: The conditional autoregressive range (carr) model. *Journal of Money, Credit, and Banking*, 37(3).

Christoffersen, P. and Pelletier, D. (2004). Backtesting value-at-risk: a duration-based approach. *Journal of Financial Econometrics*, 2:84–108.

Cox, D. R. and Isham, V. (1980). *Point Processes*. Chapman and Hall, 1992 reprint.

DeLuca, G. and Gallo, G. (2004). Mixture processes for financial intradaily durations. *Studies in Nonlinear Dynamics & Econometrics*, 8(2).

DeLuca, G. and Gallo, G. (2009). Time-varying mixing weights in mixture autoregressive conditional duration models. *Econometric Reviews*, 28(1-3).

Diamond, D. and Verrecchia, R. (1987). Constraints on short- selling and asset price adjustment to private information. *Journal of Financial Economics*, 18:277–311.

Dufour, A. and Engle, R. (2000). Time and the price impact of a trade. *Journal of Finance*, 55:2467–2498.

Easley, D. and O'Hara, M. (1992). Time and the process of security price adjustment. *Journal of Finance*, 47(577-606).

Engle, R. (2000). The econometrics of ultra-high frequency data. *Econometrica*, 68:1–22.

Engle, R. and Gonzalez-Rivera, G. (1991). Semiparametric arch models. *Journal of Business and Economic Statistics*, 9:345–360.

Engle, R. and Lunde, A. (2003). Trades and quotes: a bivariate point process. *Journal of Financial Econometrics*, 1(2):159–188.

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

Engle, R. and Russell, J. R. (1998b). Econometric analysis of discrete-valued irregularly spaced financial transaction data using a new autoregressive conditional multinomial model. *Discussion Paper, Department of Economics, University of California, San Diego*, 98(10).

Espasa, A., Rodriguez-Poo, J., and Veredas, D. (2007). Semiparametric estimation for financial durations. In Bauwens, L., Pohlmeier, W., and Veredas, D., editors, *Recent developments in high frequency financial econometrics*. Springer Verlag.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and its Applications*. Chapman and Hall.

Feng, D., Jiang, G., and Song, P. (2004). Stochastic conditional duration models with leverage effect for financial transaction data. *Journal of Financial Econometrics*, 2:390:421.

Fernandes, M. and Grammig, J. (2005). Non-parametric specification tests for conditional duration models. *Journal of Econometrics*, 127:35–68.

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

Fischer, A. and Zurlinden, M. (2004). Are interventions self-exciting? *Open Economies Review*, 15(223-237).

Focardi, S. and Fabozzi, F. (2005). An autoregressive conditional duration model of credit-risk contagion. *Journal of Risk Finance*, 6(3):208–225.

Franke, J., Hardle, W., and Kreiss, J. (2004). Non parametric estimatio in a stochastic volatility model. *Recent advances and trends in Nonparametric Statistics*.

Franke, J. H. and Muller, M. (2002). Nonparametric estimation of garch processes. *Applied Quantitative Finance*, pages 367–383.

Ghysels, E., Harvey, A., and Renault, E. (1996). Stochastic volatility. In Maddala, G. and Rao, C., editors, *Handbook of Statistics*, volume 14. Elsevier Sciences, Amsterdam.

Ghysels, E. and Jasiak, J. (1998). Garch for irregularly spaced financial data: the acd-garch model. *Studies in Nonlinear Dynamics and Econometrics*, 2:133–149.

Giot, P. (2000). Time transformations, intraday data, and volatility models. *Journal of Computational Finance*, 4(2).

Grammig, J. and Wellner, M. (2002). Modeling the interdependence of volatility and inter-transaction duration processes. *Journal of Econometrics*, 106:369–400.

Hafner, C. (1997). *Nonlinear time series analysis with applications to foreign exchange rate volatility*. Physica-Verlag, Heidelberg.

Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.

Hasbrouck, J. (1991). Measuring the information content of stock trades. *Journal of Finance*, 46:179–207.

He, C. (2000). Moments and the autocorrelation structure of the exponential garch(p,q) process. *WP Series in Economics and Finance. Stockholm School of Economics*, 359.

He, C., Terasvirta, T., and Malmsten, H. (2002). Fourth moment structure of a family of first-order exponential garch models. *Econometric Theory*, 18:868–885.

Ivanov, V. and Lewis, C. (2008). The determinants of market-wide issue cycles for initial public offerings. *Journal of Corporate Finance*, 14(5):567–583.

Jasiak, J. (1998). Persistence in intertrade durations. *Finance*, 19:166–195.

Knight, J. and Ning, C. Q. (2008). Estimation of the stochastic conditional duration model via alternative methods - ecf and gmm. *Econometrics Journal*, 11(3):593–616.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge University Press.

Liesenfeld, R. and Richard, J. F. (2003). Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance*, 10(4):505–531.

Lo, I. and Sapp, S. (2008). The submission of limit orders or market orders: the role of timing and information in the reuters d2000-2 system. *Journal of International Money and Finance*, 27(7):1056–1073.

Manganelli, S. (2005). Duration, volume and volatility impact of trades. *Journal of Financial Markets*, 8:377–399.

Meitz, M. and Terasvirta, T. (2006). Evaluating models of autoregressive conditional duration. *Journal of Business and Economic Statistics*, 24:104–124.

Mosconi, R. and Olivetti, F. (2005). Bivariate generalization of the acd models. *Working paper. Dipartimento di Economia e Produzione. Politecnico di Milano.*

O'Hara, M. (2000). *Market Microstructure Theory*. Basil Blackwell, Oxford.

Pacurar, M. (2008). Autoregressive conditional duration (acd) models in finance: A survey of the theoretical and empirical literature. *Journal of Economic Surveys*, 22(4):711–751.

Richard, J. F. and Zhang, W. (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics*, 141(2):1385–1411.

Scaillet, O. (2004). Density estimation using inverse and reciprocal inverse gaussian kernels. *Journal of Nonparametric Statistics*, 16:217–226.

Spierdijk, L. (2004). An empirical analysis of the role of the trading intensity in information dissemination on the nyse. *Journal of Empirical Finance*, 11:163–184.

Wang, J. (2004). Intertrade durations and information-based trading on an electronic order-driven market. *mimeo.*

Zhang, M. Y., Russell, J. R., and Tsay, R. T. (1999). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics*, 104(1).