Regime-switching Pareto distributions for ACD models

Giovanni De Luca\textsuperscript{a,*}, Paola Zuccolotto\textsuperscript{b}

\textsuperscript{a}Dipartimento di Statistica e Matematica per la Ricerca Economica, Università di Napoli Parthenope, via Medina, 40, 80133 Napoli, Italy

\textsuperscript{b}Dipartimento Metodi Quantitativi, Università di Brescia, Italy

Abstract

Econometric literature on financial durations has been popularized by the introduction of the Autoregressive Conditional Duration model. In the latest years, a lot of refinements have been suggested. Following financial market microstructure arguments, it is argued that a Pareto distribution is a meaningful representation for durations. The model is analyzed under the hypothesis of regime-switching parameters. The transition is assumed governed both by an observable and a latent variable.

Key words: Autoregressive conditional duration, Markov-Switching model, market microstructure, Pareto distribution.

1 Introduction

In the latest years, the analysis of financial durations has produced a large amount of papers. We define financial duration the time between two consecutive events occurring in a financial market (a trade, a price change and so on). The analysis of financial durations involves an irregularly spaced time series so that specific statistical techniques have to be exploited. In their seminal work, Engle and Russell (1998) proposed the Autoregressive Conditional Duration (ACD) model. Since then, several authors have applied the model both in the basic form and with developments in various directions. Most of the studies have been based on data coming from the mature New York Stock Exchange...
and only a small part takes into account European exchanges. Actually, the choice of a random variable to describe the durations probability law is crucial for a satisfactory goodness-of-fit of the model. The simplest assumptions of an Exponential or Weibull random variable date back to the cited paper by Engle and Russell. However, many studies have detected a poor performance and tried to improve the model using more flexible distributions or pursuing a semiparametric point of view (Drost and Werker, 2003).

In this paper we analyze the use of the Pareto distribution by assuming constant as well as time-varying parameters. In particular, the main contribution of the article is to propose a model with regime-switching parameters in the innovation component. The paper is organized as follows. Section 2 reviews the Pareto Autoregressive Conditional Duration model. In Sections 3 and 4 the regime-switching ACD model is introduced and analyzed. An application to Italian data is presented in Section 5. Section 6 concludes.

2 The Pareto ACD model

Let $X_i$ be $i$-th duration, that is $X_i = t_i - t_{i-1}$, where $t_i$ is the time of the $i$-th market event. After removing the daily seasonal effect, we get the deseasonalized durations, denoted as $x_i$. The basic ACD model of orders $p$ and $q$ is given by

$$x_i = \Psi_i \epsilon_i, \quad \epsilon_i \sim i.i.d.$$  

with $E(\epsilon_i) = 1$ and

$$\Psi_i = f(x_{i-1}, \ldots, x_{i-p}, \Psi_{i-1}, \ldots, \Psi_{i-q}).$$

Because of the unit expected value of $\epsilon_i$, $\Psi_i$ is the expected value of the (seasonality-adjusted) durations, conditionally on past information. The attention of the researcher is focused both on the functional form of $\Psi_i$ and on the distributional assumption for the innovation $\epsilon_i$. The most popular expression for the conditional expected duration is the linear form

$$\Psi_i = \omega + \sum_{j=1}^p \alpha_j x_{i-j} + \sum_{j=1}^q \beta_j \Psi_{i-j},$$  \hspace{1cm} (1)$$

where the parameters are subject to the usual restrictions ensuring the positivity of $\Psi_i$ and the stationarity of the model ($\omega > 0, \alpha_j, \beta_j \geq 0, \sum_j \alpha_j + \sum_j \beta_j < 1$).
Throughout the paper we will keep (1), but it is worthwhile to specify that in the literature there are alternative formulations (see the exhaustive paper of Hautsch, 2002). We will focus, instead, on the choice of the distribution for the innovation component $\epsilon_i$. Engle and Russell (1998) introduced the two hypotheses of Exponential and Weibull distribution, given respectively by

$$f(\epsilon_i) = \exp \{ \epsilon_i \}$$

and

$$f(\epsilon_i; \gamma) = \gamma \Gamma \left( 1 + \frac{1}{\gamma} \right)^{\gamma} \epsilon_i^{\gamma - 1} \exp \left\{ - \left[ \epsilon_i \Gamma \left( 1 + \frac{1}{\gamma} \right) \right]^{\gamma} \right\},$$

but many empirical studies can easily convince that these basic distributional assumptions do not have a good performance. If the models parsimony is not taken into account, multiparameter distributions can reveal more effective. Three parameters distributions have been explored by some authors (e.g. Zhang et al. (2001) proposed the generalized Gamma).

The heterogeneity of the traders, due to different degrees of information, attitudes toward risk, budget constraints and so on, is a possible rationale for the use of mixture distributions, implemented by De Luca and Zuccolotto (2003), De Luca and Gallo (2004) and Hujer and Vuletić (2005). However, the higher the number of components of the mixture, the greater the problems in the estimation step. A mixture of more than three components involves a number of parameters that might be too high. Finally, if we assume that heterogeneity is such that no significant cluster can be deemed, but each trader has his own personal information set, an infinite mixture of distributions is a natural candidate (De Luca and Zuccolotto, 2003).

The application implies two choices: the distribution of the components and the distribution of the mixing weight. The number of parameters to be estimated is usually lower than in the case of a finite mixture. The general form of an infinite mixture of distributions is

$$f(\epsilon_i) = \int_{\mathcal{D}_\lambda} f(\epsilon_i|\lambda)f(\lambda)d\lambda.$$

According to the above choices, either one could handle very troublesome estimation problems or one could get simple mathematical expressions. An example of the latter is the choice of exponential distributions as component distributions

$$f(\epsilon_i|\lambda) = \frac{1}{\lambda} \exp \left\{ -\frac{\epsilon_i}{\lambda} \right\}.$$
and of an Inverse Gamma as mixing distribution

\[ f(\lambda) = \frac{\theta^\delta}{\Gamma(\delta)} \exp\left\{-\frac{\theta}{\lambda}\right\} \left(\frac{1}{\lambda}\right)^{\delta+1}, \]

with positive \( \theta \) and \( \delta \). The result is a \( \text{Pareto}(\theta, \delta) \) II Type assumption for the innovations \( \epsilon_i \), that is

\[ f(\epsilon_i; \theta, \delta) = \theta^\delta \delta (\theta + \epsilon_i)^{-(\delta+1)}, \]

and the model is denoted as Pareto ACD (PACD).

A low number of parameters is involved. Moreover, the unit expected value hypothesis, \( E(\epsilon_i) = 1 \), implies a further reduction of the free parameters. In the PACD model, the constraint \( E(\epsilon_i) = 1 \), that is \( \frac{\theta}{\sigma^2} = 1 \), implies \( \delta = \theta + 1 \). The density function is then

\[ f(\epsilon_i; \theta) = \theta^{(\theta+1)}(\theta + 1)(\theta + \epsilon_i)^{-(\theta+2)}. \]

The parameters of the PACD model, the vector \( \eta = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)' \) and \( \theta \), can be jointly estimated by maximizing the log-likelihood function

\[ l(\eta, \theta) = \sum_{i=1}^{n} \left[ (\theta + 1) \log \theta + \log \left( \frac{\theta + 1}{\Psi_i} \right) - (\theta + 2) \log \left( \theta + \frac{x_i}{\Psi_i} \right) \right]. \]

Another example of infinite mixture is provided by the the Burr distribution (Grammig and Maurer, 2000) which can be derived as a Gamma mixture of Weibull distributions. Under this hypothesis the residuals density function is given by

\[ f(\epsilon_i; \kappa, \sigma^2) = \frac{\kappa}{c} \left( \frac{\epsilon_i}{c} \right)^{\kappa-1} \left[ 1 + \sigma^2 \left( \frac{\epsilon_i}{c} \right)^\kappa \right]^{1+\frac{1}{\sigma^2}}, \]

where

\[ c = \frac{(\sigma^2)^{1+\frac{1}{\kappa}} \Gamma \left( 1 + \frac{1}{\sigma^2} \right)}{\Gamma \left( 1 + \frac{1}{\kappa} \right) \Gamma \left( \frac{1}{\sigma^2} - \frac{1}{\kappa} \right)} \]

and \( \kappa > \sigma^2 > 0 \).

The resulting model is denoted as Burr ACD (BACD).
3 Regime-switching Pareto ACD models

In this paper we argue that the parameter $\theta$ characterizing the Pareto distribution could itself be variable, according to a set $S_i$ of predetermined variables. The basic hypothesis is that the dynamics of the durations $x_i$ is not completely summarized by their conditional mean $\Psi_i$ but past events somehow affect the distribution of the innovations.

So we assume that, conditionally on $S_i$, the $\epsilon_i$’s are independently and identically distributed as Pareto II Type random variables with parameters $\theta_i$, where the relationship between $\theta_i$ and $S_i$ has to be properly formalized through the definition of a function $h$ indexed by a finite set of parameters $\tau$,

$$\theta_i = h(S_i; \tau) = h_i(\tau).$$

The novel PACD model is then formalized as follows:

$$x_i = \Psi_i \epsilon_i, \quad \epsilon_i | S_i \sim i.i.d.,$$

with $E(\epsilon_i | S_i) = E(\epsilon_i) = 1$,

$$f(\epsilon_i | S_i; \tau) = [h_i(\tau)]^{(h_i(\tau)+1)} (h_i(\tau) + 1)(h_i(\tau) + \epsilon_i)^{-(h_i(\tau)+2)}$$

and $\Psi_i$ modelled as in (1).

The log-likelihood function, conditionally on $S_i$, is given by

$$l(\eta, \tau) =$$

$$\sum_{i=1}^{n} \left[ (h_i(\tau) + 1) \log h_i(\tau) + \log \left( \frac{h_i(\tau) + 1}{\Psi_i} \right) - (h_i(\tau) + 2) \log \left( h_i(\tau) + \frac{x_i}{\Psi_i} \right) \right].$$

The choice of both the function $h(\cdot)$ and the variables in $S_i$ implies a specific departure from the basic PACD model. The common framework is a regime-switching model. The parameter $\theta_i$ changes according to the state of the world or regime. In this context, we refer to the state of the market. No change is allowed in the parameters of the expected conditional duration equation. In the following subsections, we analyze three regime-switching specifications.
3.1 The Threshold Pareto ACD

First, we describe the dynamics of the Pareto parameter \( \theta_i \) assuming that the process is governed by the existence of two regimes and one threshold variable, \( S_i = S_i \). The resulting Threshold Pareto ACD (TPACD) model is characterized by

\[
h_i(\tau) = \theta_i = \begin{cases} 
\theta_1 & \text{if } S_i \in \Omega_1 \\
\theta_2 & \text{if } S_i \in \Omega_2 
\end{cases},
\]

where \( \Omega = \Omega_1 \cup \Omega_2 \) is the domain of the predetermined variable \( S_i \) and \( \Omega_1 \cap \Omega_2 = \emptyset \). In this case, once defined the variable \( S_i \) in \((a, b)\), \( \Omega_1 \) is defined as \((a, s]\) and \( \Omega_2 \) as \((s, b)\) with \( a < s < b \).

The threshold variable should be financially meaningful, such as a lagged duration, the trading intensity or the traded volume. Zhang et al. (2001) proposed the Threshold ACD (TACD) model characterized by one-lagged duration as threshold variable affecting the parameters of the conditional expected duration equation and of the generalized Gamma distribution assumed for the innovations. Our approach characterizes for two reasons: it focuses only on the parameters of the innovation distribution (those in the \( \Psi_i \) equation would change very slightly) and the innovation distribution is chosen according to financial market microstructure arguments.

The parameters of the TPACD have to be estimated through the method of maximum likelihood jointly with a grid search for the parameter \( s \).

3.2 The Smooth-Transition Pareto ACD

The Threshold Pareto ACD model involves a sudden transition between the two regimes when the threshold variable \( S_i \) switches from \( \Omega_1 \) to \( \Omega_2 \) and vice-versa. However, if we consider a gradual adjustment to the changing market conditions, a smoother transition can be more interesting. A refined version of the model takes into account a smooth transition from one regime to the other. The model is denoted as Smooth-Transition Pareto ACD (STPACD). To achieve this, we need to introduce a function \( G(S_i; s, \gamma) \) describing the transition. We assume the logistic function

\[
G(S_i; s, \gamma) = \frac{\exp\{\gamma(S_i - s)\}}{1 + \exp\{\gamma(S_i - s)\}},
\]
where the parameter $s$ is the threshold between the two regimes and $\gamma$ is the smoothness parameter. The lower $\gamma$, the smoother the transition.

Finally, $\theta_i$ can be expressed as

$$h_i(\tau) = \theta_i = \theta_1 (1 - G(S_i; s, \gamma)) + \theta_2 G(S_i; s, \gamma).$$

When the variable $S_i$ increases, the function $G(\cdot)$ tends to unity. In particular, when $S_i = s$, $G(S_i; s, \gamma) = 0.5$. In the literature, Meitz and Teräsvirta (2006) proposed the Smooth Transition ACD model with a transition function acting on some parameters of the $\Psi_i$ equation.

The estimation of the parameters is carried out using maximum likelihood.

3.3 The Markov-Switching Pareto ACD model

In the previous subsections, the threshold variable $S_i$ is observable and it is selected such that it is suitable to represent the market conditions. However, it might be too restrictive to link the market conditions to one variable. The idea of an unobservable threshold variable, that is unobservable regimes, is more general. Hujer et al. (2002) developed the Markov-Switching ACD model which allows the conditional expected duration to be affected by an unobservable stochastic process. Following Hamilton (1989), the Markov-Switching Pareto ACD (MSPACD) model is still characterized by

$$h_i(\tau) = \theta_i = \begin{cases} \theta_1 & \text{if} \ S_i \in \Omega_1 \\ \theta_2 & \text{if} \ S_i \in \Omega_2 \end{cases},$$

but $S_i$ is now a latent variable. We assume that it follows a first-order homogeneous Markov process with two states and the probabilities of transition from one regime to the other are defined as follows,

$$P(S_i \in \Omega_1|S_{i-1} \in \Omega_1) = p_{11},$$
$$P(S_i \in \Omega_1|S_{i-1} \in \Omega_2) = p_{21},$$
$$P(S_i \in \Omega_2|S_{i-1} \in \Omega_1) = p_{12}$$

and

$$P(S_i \in \Omega_2|S_{i-1} \in \Omega_2) = p_{22}. $$
In general, $p_{hk}$ is the probability of moving from regime $h$ at the generic time $t_{i-1}$ to regime $k$ at time $t_i$. The homogeneity property involves that the transition probabilities do not change over time.

The estimate of a MSP ACD is carried out using the Expectation and Maximization (EM) algorithm (Dempster et al., 1977). We can write the complete log-likelihood

$$l = \sum_{i=1}^{n} \left( \sum_{h=1}^{2} \xi_{i}^{(h)} \log \left[ f(x_i|S_i \in \Omega_h, I_{i-1}) \right] + \sum_{h=1}^{2} \sum_{k=1}^{2} \xi_{i}^{(hk)} \log \left[ p_{hk} \right] \right),$$

where $\xi_{i}^{(h)} = 1$ if $S_i \in \Omega_h$ (and zero otherwise) and $\xi_{i}^{(hk)} = 1$ if $S_i \in \Omega_h$ and $S_{i-1} \in \Omega_k$ (and zero otherwise), but we cannot work with it because the state variable $S_i$ is not observed.

After defining the variable

$$\xi_i = \begin{cases} (1, 0)' & \text{if } S_i \in \Omega_1 \\ (0, 1)' & \text{if } S_i \in \Omega_2 \end{cases},$$

let us denote $P(S_i \in \Omega_h|I_{i-1})$ as $\hat{\xi}_{i|i-1}^{(h)}$ and $\hat{\xi}_{i|i-1} = [\hat{\xi}_{i|i-1}^{(1)}, \hat{\xi}_{i|i-1}^{(2)}]'$, then the forecast is given by

$$\hat{\xi}_{i|i-1} = P \xi_{i-1},$$

where

$$P = \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix}.$$ 

Even if we do not know $\xi_i$ at any time, we can replace it using its conditional estimate

$$\hat{\xi}_{i|i} = \left((1, 1) \left(\hat{\xi}_{i|i-1} \odot f_i\right)\right)^{-1} \left(\hat{\xi}_{i|i-1} \odot f_i\right),$$

where

$$f_i = \begin{bmatrix} f(x_i|S_i \in \Omega_1, I_{i-1}) \\ f(x_i|S_i \in \Omega_2, I_{i-1}) \end{bmatrix}.$$
and the symbol \( \odot \) indicates element-by-element multiplication.

Denoting \( t_n \) as the last observation time, let us define \( P(S_i \in \Omega_h | I_n) \) as \( \hat{\xi}_{i|n}^{(h)} \) and \( \hat{\xi}_{i|n} = [\hat{\xi}_{i|n}^{(1)}, \hat{\xi}_{i|n}^{(2)}]' \), where

\[
\hat{\xi}_{i|n} = \hat{\xi}_{i|i} \odot \left( P' \left( \hat{\xi}_{i+1|n} \div \hat{\xi}_{i+1|i} \right) \right),
\]

that is the so-called smoothed inference on the regime probabilities, where the symbol \( \div \) indicates element-by-element division. Moreover,

\[
\hat{\xi}^{(hk)}_{i|n} = \hat{\xi}^{(k)}_{i-1|i-1} \frac{p_{hk} \hat{\xi}_{i|n}^{(h)}}{\hat{\xi}^{(h)}_{n|n-1}}.
\]

The EM-algorithm can be divided into two steps. In the first step, the E-step, the expectation of the complete log-likelihood (2) is evaluated so to obtain the so-called \( Q \)-function

\[
Q = \sum_{i=1}^{n} \sum_{h=1}^{2} \hat{\xi}_{i|n}^{(h)} \log [f(x_i|S_i \in \Omega_h, I_{i-1})] + \sum_{i=1}^{n} \sum_{h=1}^{2} \sum_{k=1}^{2} \hat{\xi}_{i|n}^{(hk)} \log [p_{hk}].
\]

The second step, the M-step, consists in the maximization of the \( Q \)-function (5) with respect to the parameters using standard numerical routines. The estimates of the transition probabilities \( \hat{p}_{hk} \) are then used to recompute the formulas (3) and (4) so that new estimates \( \hat{\xi}^{(h)}_{i|n} \) and \( \hat{\xi}^{(hk)}_{i|n} \) are inserted into (5). This procedure goes on until convergence is reached.

4 Tranforming innovations with regime-switching Pareto distribution

With the assumed regime-switching Pareto distribution, the innovations \( \epsilon_i \) are i.i.d. conditionally on the regime, but not unconditionally, so that usual diagnostics based upon the estimated residuals \( \hat{\epsilon}_i = x_i/\hat{\Psi}_i \) are not allowed.

The problem could be solved by properly transforming the innovations, in order to obtain a standardized distribution not depending on the parameter \( \theta_i \). In other words, given the general ACD model with regime-switching Pareto distribution

\[
x_i = \Psi_i \epsilon_i, \quad \epsilon_i|S_i \sim i.i.d., \quad \text{Pareto}(\theta_i) \quad E(\epsilon_i) = 1,
\]
we identify a monotonic transformation $g$

$$\xi_i = g(\epsilon_i, \theta_i),$$

with

$$\xi_i \sim i.i.d..$$

The resulting model is thus given by

$$x_i = \Psi_i g^{-1}(\xi_i, \theta_i)$$

and is characterized by $i.i.d.$ innovations $\xi_i$.

In order to define a proper function $g$, a useful result concerning Pareto distributions has to be recalled: if $X$ has a Pareto distribution with parameters $\theta$ and $\delta$, then the variable $W = -\log \left( \frac{X}{\theta} \right)^{\delta} - 1$ has a standard logistic distribution.

In this framework we are interested in the analogous property relative to the Pareto II Type distribution, obtained through the transformation $Y = X - \theta$. We have

$$f_Y(y) = \theta^\delta \delta (\theta + y)^{-(\delta+1)}. $$

It is easily shown that the variable

$$Z = -\log \left( \frac{\theta + Y}{\theta} \right)^\delta - 1$$

has a standard logistic distribution, that is

$$f_Z(z) = \frac{e^{-z}}{(1 + e^{-z})^2}.$$

From (6) we have:

$$e^{-Z} = \left( \frac{\theta + Y}{\theta} \right)^\delta - 1,$$

$$Y = \theta \left( 1 + e^{-Z} \right)^{\frac{1}{\delta}} - \theta$$
\[
\frac{dY}{dZ} = \frac{\theta}{\delta} \left(1 + e^{-Z}\right)^{\frac{1-\delta}{\delta}} e^{-Z}.
\]

Hence

\[
f_Z(z) = \theta^\delta \frac{\delta}{\theta + \theta\left(1 + e^{-z}\right)^{\frac{1}{\delta}}} \left(\theta + \theta\left(1 + e^{-z}\right)^{\frac{1}{\delta}} - \theta\right)^{-(\delta+1)} \left|\frac{dy}{dz}\right|,
\]

\[
f_Z(z) = \theta^\delta \delta^{-(\delta+1)} \left(1 + e^{-z}\right)^{-\frac{\delta+1}{\delta}} \left(1 + e^{-z}\right)^{-\frac{1-\delta}{\delta}} e^{-z}
\]

and

\[
f_Z(z) = \frac{e^{-z}}{(1 + e^{-z})^2},
\]

as stated above. In the context of ACD models, if the innovations \(\epsilon_i\) are, conditionally on \(\theta_i\), independently distributed as a Pareto II Type random variable with parameter \(\theta_i\), then the variables

\[
\xi_i = g(\epsilon_i, \theta_i) = -\log \left(\frac{\theta_i + \epsilon_i}{\theta_i}\right)^{(\theta_i+1)} - 1
\]

are (unconditionally on the regime) i.i.d. with a standard logistic distribution.

From an operative point of view, the estimated transformed residuals can be obtained by

\[
\hat{\xi}_i = -\log \left(\frac{\hat{\theta}_i + \hat{\epsilon}_i}{\hat{\theta}_i}\right)^{(\hat{\theta}_i+1)} - 1
\]

and usual diagnostics such as the Ljung-Box and other goodness-of-fit statistics can be computed. In this context the residuals distributional fit plays a critical role, because the adherence to the assumed distribution supports the whole regime-switching apparatus, validating both the estimates of the parameters and the identification of the time one regime or the other is
present. The distributional goodness-of-fit can be detected using the well-known Kolmogorov-Smirnov (KS) and Cramer-Von Mises (CVM) statistics, given respectively by

\[ D_n = \sup_{\epsilon} |F_n(\epsilon) - F(\epsilon)| \]

and

\[ W_n^2 = n \int_{-\infty}^{\infty} [F_n(\epsilon) - F(\epsilon)]^2 dF(\epsilon), \]

where \( F(\epsilon) \) is the assumed cumulative distribution function and \( F_n(\epsilon) \) is the empirical cumulative distribution function, estimated with a sample of size \( n \). The critical values of \( D_n \) and \( W_n^2 \) are tabulated and can be used to test the similarity between the empirical and the assumed cumulative distribution functions.

5 Case study

As a case study, (seasonality-adjusted) durations between consecutive price changes for the stock Comit (Milan Stock Exchange) are analyzed. The observation period is February 2000 (21 trading days - sample size 8222). All the computations are made in Gauss using the Newton-Raphson algorithm implemented in the CML (Constrained Maximum Likelihood) library.

A preliminary analysis is carried out on data in order to check the performance of two basic ACD models, respectively with exponential (EACD) and Weibull (WACD) innovations (Table 1). Residuals exhibit a remarkable distributional lack of fit, highlighted both by the estimated variance (appreciably different from the innovation variance under the two distributional hypotheses) and by the KS and CVM statistics.

In a second step two infinite mixture models (PACD and BACD) are used in order to compare the performance of the proposed regime-switching Pareto distributions. Results are in Table 2.

As usual, estimates of the ACD parameters do not significantly differ from the traditional models, but the distributional goodness-of-fit is considerably improved. The parameters of the Burr and the Pareto distributions are highly significant and the diagnostics for the distributional goodness-of-fit confirm a good adherence of residuals to both assumptions. Nonetheless on these data the PACD model exhibits a slightly better performance from the point of view.
Table 1
Parameter estimates (standard errors in parentheses) and diagnostics on residuals - EACD and WACD model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>EACD</th>
<th>WACD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega$</td>
<td>0.0019 (0.0007)</td>
<td>0.0018 (0.0007)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0317 (0.0036)</td>
<td>0.0317 (0.0037)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.9667 (0.0038)</td>
<td>0.9667 (0.0040)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>-</td>
<td>0.9576 (0.0081)</td>
</tr>
<tr>
<td>mean log-lik</td>
<td>-0.9353</td>
<td>-0.9337</td>
</tr>
</tbody>
</table>

Diagnostics on residuals

<table>
<thead>
<tr>
<th></th>
<th>EACD</th>
<th>WACD</th>
</tr>
</thead>
<tbody>
<tr>
<td>LB(20) (p-value)</td>
<td>0.1071</td>
<td>0.1065</td>
</tr>
<tr>
<td>Residuals mean</td>
<td>1.0003</td>
<td>1.0012</td>
</tr>
<tr>
<td>Mean of the assumed distribution</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Residuals variance</td>
<td>1.2321</td>
<td>1.2346</td>
</tr>
<tr>
<td>Variance of the assumed distribution</td>
<td>1</td>
<td>1.0911</td>
</tr>
<tr>
<td>KS statistic</td>
<td>0.0289</td>
<td>0.0308</td>
</tr>
<tr>
<td>critical value ($\alpha = 0.05$)</td>
<td>0.015</td>
<td>0.015</td>
</tr>
<tr>
<td>CVM statistic</td>
<td>2.3399</td>
<td>2.6795</td>
</tr>
<tr>
<td>critical value ($\alpha = 0.05$)</td>
<td>0.461</td>
<td>0.461</td>
</tr>
</tbody>
</table>

of the variance of residuals and the CVM statistic. The good fit of the PACD model encourages the exploration of the proposed further development in that direction, in order to refine the distributional hypothesis. In the next sections the three regime-switching distributional assumptions introduced in the paper are applied to Comit data.

5.1 TPACD and STPACD models

The results of estimation of TPACD and STPACD models with two regimes are displayed in Table 3. The trading intensity at time $t_{i-1}$, $TI_{i-1}$, given by the ratio between the number of trades between $t_{i-1}$ and $t_i$ and the duration $x_i$, is used as variable determining the switching, after a transformation aimed at obtaining unit mean. So we have $S_i = TI_{i-1}/\bar{TI}$. It is our opinion that the trading intensity summarizes the conditions of the market. A high trading
### Table 2
Parameter estimates (standard errors in parentheses) and diagnostics on residuals - BACD and PACD model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>BACD</th>
<th>PACD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega$</td>
<td>0.0016 (0.0007)</td>
<td>0.0016 (0.0007)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0322 (0.0039)</td>
<td>0.0321 (0.0039)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.9666 (0.0041)</td>
<td>0.9665 (0.0041)</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>1.0457 (0.0152)</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.1443 (0.0218)</td>
<td>-</td>
</tr>
<tr>
<td>$\theta$</td>
<td>-</td>
<td>9.7711 (1.4021)</td>
</tr>
<tr>
<td>mean log-lik</td>
<td>-0.9301</td>
<td>-0.9307</td>
</tr>
</tbody>
</table>

### Diagnostics on residuals

<table>
<thead>
<tr>
<th></th>
<th>BACD</th>
<th>PACD</th>
</tr>
</thead>
<tbody>
<tr>
<td>LB(20) (p-value)</td>
<td>0.1157</td>
<td>0.1140</td>
</tr>
<tr>
<td>Residuals mean</td>
<td>0.9990</td>
<td>1.0001</td>
</tr>
<tr>
<td>Mean of the assumed distribution</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Residuals variance</td>
<td>1.2305</td>
<td>1.2330</td>
</tr>
<tr>
<td>Variance of the assumed distribution</td>
<td>1.2640</td>
<td>1.2280</td>
</tr>
<tr>
<td>KS statistic</td>
<td>0.0092</td>
<td>0.0092</td>
</tr>
<tr>
<td>critical value ($\alpha = 0.05$)</td>
<td>0.015</td>
<td>0.015</td>
</tr>
<tr>
<td>CVM statistic</td>
<td>0.1724</td>
<td>0.0949</td>
</tr>
<tr>
<td>critical value ($\alpha = 0.05$)</td>
<td>0.461</td>
<td>0.461</td>
</tr>
</tbody>
</table>

intensity involves either a high number of trades or short durations (or both) that is a very active phase in the market.

For the TPACD model the optimal threshold $s$ is searched by iteratively estimating the model for different values of $s$, ranging from the 1st to the 99th percentile of $TI_{i-1}$, and then refining the search around the values leading to the global maximum loglikelihood (Figure 1). For the STPACD model both the threshold $s$ and the smoothing parameter $\gamma$ are estimated simultaneously to the ACD model, as described in Section 3.2. In this way the smoothing parameter could be affected by a lack of significancy (see Franses and van Dijk, 2000), so that the $t$-test should be evaluated with some elasticity.

It is apparent that the two specifications lead to substantially equal results.
Table 3
Parameter estimates (standard errors in parentheses) and diagnostics on residuals - TPACD and STPACD model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>TPACD</th>
<th>STPACD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega$</td>
<td>0.0018 (0.0008)</td>
<td>0.0019 (0.0008)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0316 (0.0038)</td>
<td>0.0317 (0.0038)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.9670 (0.0040)</td>
<td>0.9669 (0.0040)</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>11.5822 (1.9155)</td>
<td>11.6127 (1.9809)</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>2.2350 (0.5340)</td>
<td>2.2100 (0.5289)</td>
</tr>
<tr>
<td>$s$</td>
<td>2.89954 (-)</td>
<td>2.7527 (0.2339)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td></td>
<td>22.0406 (50.7943)</td>
</tr>
<tr>
<td>mean log-lik</td>
<td>-0.9295</td>
<td>-0.9295</td>
</tr>
</tbody>
</table>

Diagnostics on transformed residuals (logistic)

<table>
<thead>
<tr>
<th></th>
<th>TPACD</th>
<th>STPACD</th>
</tr>
</thead>
<tbody>
<tr>
<td>LB(20) (p-value)</td>
<td>0.0881</td>
<td>0.0920</td>
</tr>
<tr>
<td>Residuals mean</td>
<td>-0.0237</td>
<td>-0.0184</td>
</tr>
<tr>
<td>Mean of the assumed distribution</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Residuals variance</td>
<td>3.0525</td>
<td>3.0448</td>
</tr>
<tr>
<td>Variance of the assumed distribution</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>KS statistic</td>
<td>0.0090</td>
<td>0.0090</td>
</tr>
<tr>
<td>critical value ($\alpha = 0.05$)</td>
<td>0.015</td>
<td>0.015</td>
</tr>
<tr>
<td>CVM statistic</td>
<td>0.0990</td>
<td>0.1219</td>
</tr>
<tr>
<td>critical value ($\alpha = 0.05$)</td>
<td>0.461</td>
<td>0.461</td>
</tr>
</tbody>
</table>

The high value of the smoothing parameter $\gamma$ in the STPACD model describes a very sharp regime switching, so that the STPACD transition almost overlaps the TPACD pattern (Figure 2).

As expected, the ACD parameters estimates do not appreciably change with respect to the simpler assumptions, while the Pareto parameters are largely significant, so that the presence of two regimes seems to be validated. The first regime is a relatively quiet status (the variance of the Pareto random variable with $\theta \simeq 11.6$ is about 1.19), while the second regime is much more turbulent (the variance of the Pareto random variable with $\theta \simeq 2.2$ is about 2.67). The threshold $s$ approximatively corresponds to the 93th percentile of the trading
intensity, so that the second regime, characterized by high volatility, comes up with very high trading intensities.

The distribution of residuals, transformed by (7) using the estimated $\theta_i$’s, is adherent to the standard logistic distribution, as demonstrated by residuals mean and variance and by KS and CVM statistics. This result supports both the estimated value of the parameters and the exact identification of the time of occurrence of the regimes.

The average permanence in the two regimes is respectively 1160.61 and 54.63
seconds. In order to further explore the main features of the regimes, the mixing distributions leading to the two Pareto specifications and the hazard functions of the two regimes can be compared (Figure 3).

Recalling that the Pareto assumption derives from the attempt to take into account the heterogeneity of traders through an infinite mixture of exponential variables, examining the mixing distributions should give informations about the composition of traders. The mixing distribution of the second regime is overdispersed with respect to the other, meaning that in situations of market instability traders tend to be more heterogenous, probably due to the greater incidence of different degrees of information during these short turbulence periods.

The examination of the hazard functions confirms the tendency to turbulence situations in the second regime, characterized by high hazard associated to short durations.

5.2 MSPACD model

The MSPACD model with two regimes, obtained with Markov-Switching Pareto distributed innovations, is estimated via EM algorithm and results are displayed in Table 4. The outcome confirms the existence of two regimes characterized by different parameters of the Pareto distribution. The first regime, with the higher parameter, is largely prevalent, and has a marginal estimated probability $\hat{p}_1 = 0.966$ and an estimated permanence probability $\hat{p}_{11} = 0.997$. The second regime arises only sporadically (Figure 4).

The raw residuals are transformed by (7) in logistic i.i.d. residuals assigning each observation to the regime with the higher probability. The distributional goodness-of-fit is satisfactory.
Table 4  
Parameter estimates (standard errors in parentheses) and diagnostics on residuals  
- MSP ACD model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MSP ACD</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \omega )</td>
<td>0.0016 (0.0007)</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.0299 (0.0035)</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.9690 (0.0036)</td>
</tr>
<tr>
<td>( \theta_1 )</td>
<td>12.3500 (2.1174)</td>
</tr>
<tr>
<td>( \theta_2 )</td>
<td>0.5631 (0.0949)</td>
</tr>
<tr>
<td>mean log-lik</td>
<td>-0.9255</td>
</tr>
</tbody>
</table>

Diagnostics on transformed residuals (logistic)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LB(20) (p-value)</td>
<td>0.0639</td>
</tr>
<tr>
<td>Residuals mean</td>
<td>-0.0099</td>
</tr>
<tr>
<td>Mean of the assumed distribution</td>
<td>0</td>
</tr>
<tr>
<td>Residuals variance</td>
<td>3.0714</td>
</tr>
<tr>
<td>Variance of the assumed distribution</td>
<td>3</td>
</tr>
<tr>
<td>KS statistic</td>
<td>0.0122</td>
</tr>
<tr>
<td>critical value (( \alpha = 0.05 ))</td>
<td>0.015</td>
</tr>
<tr>
<td>CVM statistic</td>
<td>0.2624</td>
</tr>
<tr>
<td>critical value (( \alpha = 0.05 ))</td>
<td>0.461</td>
</tr>
</tbody>
</table>

With the Markov-Switching specification the variable \( S_t \) determining the regime transition is a latent variable with its own stochastic structure, differently by the TP ACD and STP ACD models, where the switching is assumed governed by one or more predetermined variables. The Markov-Switching hypothesis should be somehow more complete, as the latent variable is able to take into account a great number of observed and unobserved factors having an impact on the transition mechanism. Nevertheless, it could be interesting a comparison between the latent mechanism featured by the Markov-Switching model and the assumption about the variable determining the TP ACD and STP ACD transition. By a nonparametric estimate of the mean of \( TI_{t-1} \) conditional on the probability of the first regime, the existence of a relation between the trading intensity and the Markov-Switching regimes is detected (Figure 5). The trading intensity sharply decreases with increasing values of the probability of the first regime: also with the Markov-Switching specification the second
Fig. 4. Smoothed inference on regime 1 probability ($\hat{\xi}_i^{(1)}$).

...regime is characterized by high trading intensity, consistently with the results obtained with the TPACD and STPACD models.

Fig. 5. Mean of $T_{i-1}$ conditional on the smoothed inference on regime 1 probability (nonparametric estimate, $k$-nearest neighbor).

6 Concluding remarks

One of the main topics in the context of ACD models is the formulation of a correct distributional assumption for the innovation component. In this sense,
a microstructure argument could be the fundamental heterogeneity of traders in the financial market. Following this logic, former research proposed the use of finite or infinite mixtures of distributions (De Luca and Zuccolotto, 2003). An infinite mixture of exponentials with Inverse Gamma mixing distribution leads to the assumption of Pareto innovations, namely to the PACD model, which exhibited good performance in several empirical analyses (see for example De Luca and Zuccolotto, 2004).

In this paper a regime-switching Pareto distribution is proposed. In other words, a regime varying specification is formulated assuming a switching parameter $\theta$ for the Pareto distribution. The idea behind this proposal is that the dynamics of the durations is not completely summarized by their conditional mean, as a regime effect is present in the innovation process.

Three possible models are developed under this hypothesis. The first two (TPACD and STPACD) are based on a regime transition determined by one or more predetermined variables, the last (MSACD) treats the regimes as a latent Markov Chain process.

The performance of the proposed specifications is checked with satisfactory results on real data from the Italian financial market: the three models outcomes are consistent each other and the distributional goodness-of-fit diagnostics confirm the adherence to the assumed hypothesis.

References


