Autoregressive Conditional Duration Models: 

An Application in the Brazilian Stock Market 

Mauri Aparecido de Oliveira 

Department of Administration and Quantitative Methods 
Escola Paulista de Política, Economia e Negócios, EPPEN, Rua Angélica, nº 100 
Jd. das Flores Osasco, CEP: 06110-295 
Federal University of São Paulo, UNIFESP, Brazil 

Ricardo Luiz Pereira Bueno 

Rua Angélica, nº 100, Jd. das Flores, Osasco, CEP: 06110-295 
Federal University of São Paulo, UNIFESP, Brazil 

Laís Sayuri Kotsubo 

Rua Angélica, nº 100, Jd. das Flores, Osasco, CEP: 06110-295 
Federal University of São Paulo, UNIFESP, Brazil 

Daniel Reed Bergmann 

Department of Administration, Faculdade de Economia, Administração e 
Contabilidade, FEA, Av. Prof. Luciano Gualberto, 908 - Butantã, São Paulo - 05508-010, University of São Paulo, USP, Brazil 

Copyright © 2015 Mauri Aparecido de Oliveira et al. This article is distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

Abstract

In recent years, the number of negotiations related to the activities in the financial sector, especially in equities and money markets have grown considerably. With the development of information systems, more people are involved in negotiations with financial instruments, particularly in the form of electronic trading. Financial markets have become a source of high-frequency data. To understand and predict future market developments, the use of high-frequency data has required greater attention and research, both in academia, and financial institutions. The models
for high-frequency data, follow the concepts of autoregressive conditional duration models (ACD) created by Engle and Russell (1998). These models belong to the class of duration models. The ACD model is a duration model with time-series characteristics, since it combines features of duration models with the specifications of ARCH-type time series. The high-frequency data was obtained from the BM&FBOVESPA. For the construction of statistical models and variants of ACD was chosen using the data of the company Petrobras, which the database is coded PETR3. The results show that the Burr-ACD model contains the EACD and WACD models as special cases. While the Burr-ACD model requires more effort for implementation and assessment than the standard ACD models, the advantage is that the conditional density and survival function of the durations of transactions in the Burr-ACD model are less restrictive and can take more realistic forms.

**Mathematics Subject Classification:** 90C29, 65K05

**Keywords:** High-Frequency Data, ACD Models, Intraday Data, Mathematical Finance, Market Microstructure

### 1 Introduction

High-frequency trading extends negotiations to programs that are usually executed by computational algorithms molding a set of trading orders at high speeds and increasing market liquidity. In other words, liquidity can be defined as the instantaneous ability of buying or selling a big amount of stocks on a market with minimum significant impacts over its price [13]. Liquidity represents the process of quickly converting a financial asset in cash, or vice-versa. The term high-frequency implies the increase of market liquidity compared to manual negotiations. As an intensifier of liquidity, high-frequency trading operations can use sophisticated algorithms to analyze multiple markets and execute multiple orders arbitrage strategies. There are four advantages in a high-frequency trading system: computer algorithms; increase on market liquidity, a set of orders, and higher speed than manual executions. Among these elements, computer algorithms and increase on market liquidity are the necessary conditions to the creation of a high-frequency trading operation [1]. Computer algorithms are programs designed by financial engineers or software engineers to automate commercial activities or quantitative models that are usually carried out by operators. Some algorithms are based on mathematical models and others are not. For example, a computer algorithm able to calculate the risk using Monte Carlo simulation is based on statistical models. On the other hand, a computer algorithm designed to obtain real-time quotation of an Internet title does not require support from statistics or mathematics. A set of orders refers to a group of sale or purchase orders, usually used in arbitrage operations that, for example, seek to profit with divergences in values.
and prices in a short period of time (See Instruction CVM n° 387, from April, 28th, 2003). Another example is the hedge strategy, in which a trader can use computer algorithms to combine an already defined position of an underlying asset (stocks or obligations) and however, cover the risks of losing money by purchasing a selling option over this asset. As a result, a set of orders (purchase of the asset and of its option) turns into a part of the automated negotiation strategy.

The speed advantage was the main item for the success of operations in the beginning of high-frequency outcry. It was possible due to a recent development of computational technology and sophisticated computational algorithms. For instance, Goldman Sachs and IBM contributed to high-frequency trading by operating with the divergence in milliseconds of purchase and sale orders, which brought forth a significant profit in the first semester of 2009 when the US was still in the middle of the subprime crisis. In general, high-frequency operations are well welcomed by financial market professionals once the executions speed of an order can put traders in a setting from which they can obtain better profit chances. To make these operations come true, financial enterprises try putting their own computers as close as possible to the equipments that perform exchanges and matching – ideally in the same data center – which actually came to be known as co-location [19].

The co-location principle demands that the servers of the negotiation be located as close as possible to the switch equipments especially in high-frequency transactions. However, the execution speed is not a necessary condition to a high-frequency negotiation system because instead it can make use of advanced computational algorithms to overcome peers [35].

The main innovation that divides the high-frequency negotiation from the low frequency one is the high capital turnover in high-speed computer-generated answers based on market changes. Negotiation strategies of high-frequency are characterized by a larger number of business and operation with lower average profits. Many managers from traditional funds can keep their business positioning for weeks or even months, causing some percentage returns for each trade. In comparison, the high-frequency managers execute multiple operations each day, earning a fraction of percentage points in return for each transaction.

High-frequency trading has also advantages. High-frequency strategies have few or almost any interconnection with traditional long-term strategies, which turns them into valuable means of diversification of portfolios in a long period. High-frequency strategies also require shorter terms of evaluation due to its own statistical properties [25]. Among the benefits obtained from high-frequency social strategies the following can be noticed: increase in market efficiency and liquidity, advance in computational technology and stabilization of market systems. High-frequency strategies can find opportunities and negotiate operations, moving away market temporary inefficiencies and making fast use of information regarding pricing. Many high-frequency strategies give market liquidity, improving its operation and minimizing the friction costs to the investors. High-frequency trading encourages computational technology innovation of financial transactions and enables possible new solutions to be created.
in order to decrease Internet communication bottlenecks among market agents. Nowadays, four groups of negotiation strategies are popular in the high-frequency category: provision of automated liquidity, market microstructure negotiation, events negotiation and detour arbitrage.

The first challenge in the high-frequency trading development is to deal with a large amount of intraday data. Despite the use of diary data by multiple traditional investment analysis, intraday data constitutes a huge volume of information and can be irregularly sparse requiring new tools and methodologies. The second challenge is the signal precision (or transmissions). Since the gains can easily become losses if a signal misalignment occurs, a signal must be precise enough to trigger operations in a fraction of second. Execution speed is the third challenge [32]. In a high-frequency landscape, orders via traditional telephony system are not sustainable. The only reliable way to reach necessary speed is automated generation and execution through computers (More details about types of orders and negotiations “on line” can be obtained at CMV in http://www.cvm.gov.br/port/protinv/caderno5.asp.). High-frequency computer systems programming requires advanced knowledge in software development. Execution time mistakes can be expensive and therefore human supervision is still essentially required on negotiations to guarantee that the system keeps functioning inside the risk-borders pre-specified. Financial markets are high-frequency data source. The original form of pricing is given tick-by-tick: every “tick” is a unity of logical information, as a quotation or a transaction cost. By nature these data are irregularly spaced in time. Liquid markets create hundreds or thousands of ticks per workday. This way, high-frequency data must be the main object of investigation for those who are interested in comprehending financial markets.

Most recently, the availability of intraday financial databases had a big impact over the applied econometrics research and financial market microstructure theory. These tick-by-tick intraday databases are now available to most stock markets as New York Stock Exchange (NYSE), Bourse de Paris, Frankfurt Stock Exchange and Bolsa de Valores de São Paulo. In applied econometrics literature, the availability of new database originated the high-frequency models, which try to describe the process of pricing (e.g. the volatility or negotiation intensity) in an intraday basis. Firstly, the extensions of the standard serial weather model (GARCH, for instance) which handle data in regularly spaced time and focus on the volatility process during the day [2, 3, 4, 11, 12, 13, 14, 17, 36, 37] Due to the irregularly spaced tick-by-tick data, time changes become necessary to convert original irregularly time-spaced data in regularly time-spaced ones. This usually involves collecting data in a specific frequency. Once collected the data and intraday seasonality has been taken into account, standard GARCH models may be applied [27, 29].

Secondly, the so-called high-frequency models follow concepts of autoregressive conditional duration models (ACD) created by [22]. The ACD model is a duration model with temporal series characteristics because it combines technical features of duration models with ARCH type temporal series specifications [15, 16].
2 High-Frequency Data: Characteristics and Importance

The growing presence of high-frequency negotiations firms in the US capital market was well documented in the past few months (an example is the Aite Group high-frequency trading report: “New World Order: The High Frequency Trading Community and its Impact on Market Structure”, February 2009, and the “High-Frequency Trading: A Critical Ingredient in Today’s Trading Market”, May 2009).

In Brazil, in September 2010 the availability of the models 2, 3 and 4 of Direct Market Access (DMA) on the BM&FBOVESPA and the implementation in November 2010 of the new pricing policy to the High-Frequency Traders (HFT) in the Bovespa segment provided the conditions for boosting the growth of this type of investor in the Brazilian stock market. The first results from those initiatives can be seen with the negotiation of such investors that represented 4.5% of the total trading amount of November 2010 and 4.0% in December 2010, with daily averages (purchases and sales) of R$0.6 billion and R$0.5 billion respectively (BM&FBOVESPA S.A. – Bolsa de Valores, Mercadorias e Futuros, Demonstrações Financeiras de 2010).

In this paper, obtaining high-frequency data was carried out in a partnership with Service Development Management of BM&FBOVESPA. Figure-1 shows the high-frequency database features obtained by BM&FBOVESPA and it is important to notice its time field format referring to trading hours as HH:MM:SS.NNNNNN, which means considering time intervals of micro-seconds range (10^-6 s). The database starts with the date and then identifies the asset, using a code. As shown in Figure-2, the first company has the ABCB4 (ABC BRASIL PN) code. In Figure-2, the beginning date of the database in 2010-09-01 is presented and part (b) shows its ending date in 2010-09-30 with the second company receiving the WSON11 (WILSON SONS DR3) code.

![Figure 1. Traded high-frequency data archive layout](image-url)
The total data volume from 2010-09-01 to 2010-09-30 is of 7,343,721 values (disregarding the first and the last lines) as can be seen on the last line of part (b) of Figure-2. Since the data is available, its handling can be done by creating descriptive statistics and construction of models. The estimated models used in this work are variations of the autoregressive conditional duration models (Autoregressive Conditional Duration – ACD).

2.1 High-Frequency Data Analysis
The following are the theoretical basis on the use and high-frequency data analysis. This type of modeling takes some considerable amount of time when performed in personal computers. The reading of some CDs containing high-frequency data obtained from BM&FBOVESPA requires a specific software to access its data. In this case, we used the LTF (Large Text File) Viewer 5.2u software. This study will be conducted using active high-frequency data from PETROBRAS ON which receives the code PETR3 within the database.

2.1.1 Trade and Quote Type Database
Trade and Quote type of database, also known as TAQ, provides information on the intraday process of pricing and trading stocks quotations. Although databases containing financial information have existed for a long time, intraday databases...
providing open information became available only on the early 90s. In our times most stock markets turn public to academic community the complete (or almost complete) record of its intraday activities.

The TAQ database from New York Stock Exchange contains data on all trading and share quotations of its own, of American Stock Exchange (AMEX) and of National Association of Securities Dealers Automated Quotation (Nasdaq).

Table-1 shows trading and pricing consecutive variations to Petrobras. The information from TAQ database cannot be immediately applied in an econometric program. An extraction procedure must be performed before in order to enable the database usable. On Table-1 the positive sign (+) indicates price rise, (0) represents stable prices and (-) means price decrease. This way considering consecutive trading from data collected in PETR3 on 2010-09-01 there was 691 price rises, 2885 stable values and 638 price decreases.

<table>
<thead>
<tr>
<th></th>
<th>01/09/10</th>
<th>02/09/10</th>
<th>03/09/10</th>
<th>06/09/10</th>
<th>08/09/10</th>
<th>09/09/10</th>
<th>10/09/10</th>
<th>13/09/10</th>
<th>14/09/10</th>
<th>15/09/10</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>671</td>
<td>1228</td>
<td>1682</td>
<td>801</td>
<td>730</td>
<td>943</td>
<td>961</td>
<td>927</td>
<td>1291</td>
<td>2150</td>
</tr>
<tr>
<td>0</td>
<td>2885</td>
<td>5400</td>
<td>7820</td>
<td>2662</td>
<td>3667</td>
<td>4497</td>
<td>5368</td>
<td>5512</td>
<td>4767</td>
<td>8901</td>
</tr>
<tr>
<td>−</td>
<td>638</td>
<td>1216</td>
<td>1598</td>
<td>797</td>
<td>743</td>
<td>946</td>
<td>992</td>
<td>899</td>
<td>1459</td>
<td>2032</td>
</tr>
<tr>
<td>Total</td>
<td>4194</td>
<td>7844</td>
<td>11100</td>
<td>4260</td>
<td>5140</td>
<td>6386</td>
<td>7321</td>
<td>7338</td>
<td>7517</td>
<td>13083</td>
</tr>
</tbody>
</table>

(a) Daily trading numbers

<table>
<thead>
<tr>
<th></th>
<th>01/09/10</th>
<th>02/09/10</th>
<th>03/09/10</th>
<th>06/09/10</th>
<th>08/09/10</th>
<th>09/09/10</th>
<th>10/09/10</th>
<th>13/09/10</th>
<th>14/09/10</th>
<th>15/09/10</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>16.0%</td>
<td>15.7%</td>
<td>15.2%</td>
<td>18.8%</td>
<td>14.2%</td>
<td>14.8%</td>
<td>13.1%</td>
<td>12.6%</td>
<td>17.2%</td>
<td>16.4%</td>
</tr>
<tr>
<td>0</td>
<td>68.8%</td>
<td>68.8%</td>
<td>70.4%</td>
<td>62.5%</td>
<td>71.3%</td>
<td>70.4%</td>
<td>73.3%</td>
<td>75.1%</td>
<td>63.4%</td>
<td>68.1%</td>
</tr>
<tr>
<td>−</td>
<td>15.2%</td>
<td>15.5%</td>
<td>14.4%</td>
<td>18.7%</td>
<td>14.5%</td>
<td>14.8%</td>
<td>13.6%</td>
<td>12.3%</td>
<td>19.4%</td>
<td>15.5%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

(b) Daily trading percentage

2.1.2 Realized Volatility

Beyond the GARCH volatility estimators among the most popular volatility measurements is intra-period volatility, known as “realized volatility”. The realized volatility according to Andersen, Bollerslev, Diebold e Labys (2003) is calculated as the sum of squares of the intra-period returns obtained from the “break” of time into n small increments of equal duration.
Figure 3. Series of (a) prices series and (b) returns from Petrobras (PETR3), (c) histogram e (d) a.c.f. returns

Take $p_{i,t}$ as the price logarithm (log-price) of the $i$ asset in $t$ time in which each log-price is associated to time intervals regularly spaced (e.g. each 5 minutes or each 30 minutes). In the multivariate context, $p_{t} = (p_{i_{1},t}, \ldots, p_{i_{n},t})'$ being the vector $n \times 1$ ($n =$ number of assets) of price logarithms in the $t$ time. $\Delta$ denotes the fraction of a trading session associated with the frequency of an implicit sample, and calling $m = 1/\Delta$ the number of sample observations per trading session. As an example if the price samples are observed each 30 minutes and the negotiations happen 24 hours a day, then there are $m = 48$ time intervals of 5 minutes per day and $\Delta = 1/288 \approx 0.0035$ [13].

If the prices are sampled every 5 minutes and a negotiation occurs 6,5 hours a day (e.g. NYSE operations 9:30am EST to 4pm EST), then there is $m = 78$ time intervals of 5 minutes per day of negotiation and $\Delta = 1/78 \approx 0.0128$. Let $T$ be the number of days of the sample. Next there will be a total amount of $mT$ observations on each asset $i = 1, \ldots, n$ [13]. The intraday return continuously composed over the $i$ asset from $t$ time to $t + \Delta$ is defined as: $r_{i,t+\Delta} = p_{i,t+\Delta} - p_{i,t}, i = 1, \ldots, n$. The vector $n \times 1$ of returns $cc$ on $t$ time to $t + \Delta$ is defined as: $r_{t+\Delta} = p_{t+\Delta} - p_{t}$. For simplicity of notation, the daily returns as indicated
as a subscript $t$, such that: $r_{i,t} = r_{i,t-1+\Delta} + r_{i,t-1+2\Delta} + \ldots + r_{i,t-1+m\Delta}, i = 1,\ldots, n$ and vectorially as $\mathbf{r}_t = r_{r_{i-1+\Delta}} + r_{r_{i-1+2\Delta}} + \ldots + r_{r_{i-1+m\Delta}}$. The realized variance (RV) for the $i$ asset ($i = 1, \ldots, n$) on the $t$ day is defined as: $RV_{i,t} = \sum_{j=1}^{m} r_{i,t-1+j\Delta}^2, i = 1,\ldots, T$. The realized volatility (RVOL) for the $i$ asset on the $t$ day is defined as the square root of the realized variance: $RVOL_{i,t} = \sqrt{RV_{i,t}}$. The realized volatility logarithm (RLVOL) is the logarithm of RVOL: $RLVOL_{i,t} = \ln(RVOL_{i,t})$. The matrix $n \times n$ of realized covariance (RCOV) on the $t$ day is defined as: $RCOV_{t} = \sum_{j=1}^{m} \mathbf{r}_{i,t-1+j\Delta} \mathbf{r}_{t-1+j\Delta}', t = 1,\ldots, T$ [5].

The behavior of the PETR3 pricing series on 2010-09-01 in the intervals of: (a) 10am to 5pm; (b) first hour of trading; (c) first thirty minutes of trading; (d) first fifteen minutes; (e) first minute; e (f) first second of trading.

![Graphs showing different intervals of PETR3 pricing series](image)

**Figure 4.** Pricing series behavior on 2010-09-01 of PETR3 in dissonant time intervals

For sure $RV_{i,t} = [RCOV_{i,t}]$ and $RCOV_{i,j,t} = [RCOV_{i,j,t}]$. The $RCOV_{i}$ matrix which dimension $n \times n$ will be positive and defined when $n < m$; i.e. when the number of assets is smaller than the intraday observations [34]. The interconnection between the $i$ asset and the $j$ asset is calculated as the following:
Having the daily measures of RV and RCOV, the non-overlapped matching measures on $h$ days are discovered by:

$$RV_{i,t}^h = \sum_{j=1}^{h} [RV_{i,j}], \quad t = h, 2h, \ldots, T/h$$ \hspace{1cm} (1)

$$RCOV_{i,t}^h = \sum_{j=1}^{h} [RCOV_{i,j}], \quad t = h, 2h, \ldots, T/h.$$ \hspace{1cm} (2)

Figure 5 shows in (a) realized volatility, on part (b) the autocorrelation function (a.c.f.) of the realized volatility, in (c) is shown the realized volatility logarithm and in (d) the autocorrelation function of the realized volatility logarithm.

**2.1.3 Intraday Seasonality of Durations**

In regular economic conditions, the transactions reveal a daily seasonality factor. It seems that the number of transactions is larger right after the opening of the business than when the session is ended (when the time intervals between operations are shorter) and significantly minor during the mid-day, i.e. in the middle
of the session (a so-called “lunch effect” when the duration between operations are also longer). Therefore, there is a certain repetitious pattern of transactions intensity for each day. This is called “intraday seasonality of durations”. Consequently, [21] recommend the decomposition of the durations in a $\phi(t_i)$ deterministic component that may vary depending on the $t_i$ starting momentum of a specific duration, and a $\hat{x}_i$ stochastic component that is free from the effects of seasonality and that molds process dynamics. The pertinent literature [21] recommends that data be transformed as such: $\hat{x}_i = \frac{x_i}{\phi(t_i)}$, where $x_i = t_i - t_{i-1}$ is the duration among transactions on $t_i$ and $t_{i-1}$, $\hat{x}_i$ is the duration under no seasonality effect and $\phi(t_i)$ is the multiplicative factor of intraday seasonality in $t_i$ time [7].

The seasonal factor $\phi(t_i)$ is interpreted as an average duration of each time unity in which data observations were made (most of time denoting the average duration of each second). The diagram that illustrates the intraday seasonality pattern, also known as diurnal mode, or hour-of-the-day function usually has the shape of an inverted U. In many situations the researchers do not have enough information to completely specify an intraday seasonality parametric function. Even though intraday cyclicity does not represent a key issue of most studies, it cannot be ignored and often can be included in analysis [12]. Thus the hour-of-the-day function can be estimated by selecting non-parametrical statistic methods as such splines, Fourier series, neural networks, wavelet analysis or kernel methods. In most works on duration modeling, cubic splines can be found or estimates can be done via kernel [18].
The method using splines allows to soften the average duration between events on subsequent periods. In first place, all average values of durations are obtained during the subsequent hours of the sessions on each day. Then a cubic spline with knots for each full hour of session is determined. These knots correspond to the previously determined average duration. This daily period factor approximation version was lectured by [21]. Aiming to guarantee more elasticity the authors added a knot in the half an hour of the last hour of the session to catch the quick increase of trading activities before the end of the stock market shift. A lightly unlike approach to the cubic spline approximation can be seen in [6] and in [31].

In the first 5 days of analysis, the diurnal effect over the average durations that presents the shape of an inverted U, especially due to the market operation with no negative trading news was observed. However, the picture in which the last 5 days are displayed it appears that the inverted U pattern does not occur. To contemplate this condition it is important to apply the [30] model because of the shape of the duration diurnal curve.

Figure-6 Petrobrás (PETR3) duration series, histogram, a.c.f. and descriptive statistics
[8] observe that the intraday seasonality factor may vary from a week to another, i.e. the periodic factor shape of Monday can be different from that was foreseen for Tuesday, etc. Consequently the estimates intraday seasonality function was performed separately for each day of the week to allow the identification a potential seasonality during the week. Figure-6 shows durations data of the first 5 days of analysis. Figures 7 and 8 points the diurnal effect on the PETR3 average durations during the 10 days of analysis, 2010-09-01 to 2010-09-15.
3 ACD Models

Duration models refer to time intervals between negotiations. Longer durations point toward lack of trading activities, which means a period of new information. The duration dynamics therefore contains useful information about intraday market activities. The time interval between successive operations brings information about the potential liquidity effects. This effect differs from the commonly studied pricing impacts that determine how far a trade goes without affecting the prices. Inter negotiations duration is an indication of the ability to trade at any price. It is already known that inter negotiations duration presents persistence and temporal groupings which led to the development of autoregressive conditional duration (ACD) models by [20] and [22] and future developments by [6], [33] and [10]. Duration clashes may be caused by the arrival of new information may have longer term, but not permanent effects on the future durations [15]. In classic econometric techniques the temporal series are frequently taken as data sequences apart by regular time intervals. Opting for increasingly higher frequencies was the search for better or fuller information [23, 6]. Both techniques data frequency and modeling were developed at the same pace of information technologies advance, essential for record keeping and processing dozens of involved information pieces. However, with the advent of database transactions and quotations, researchers have run into a problem that required a new type of modeling [9].

3.1 The Standard ACD Model

The ACD model introduced by [22] can be conceived as a marginal duration model \( x_i \). Consider the expected conditional duration as:

\[
\psi_i = E \left[ x_i | x_{i-1} \right] = \psi_i \left( \bar{x}_{i-1}, \bar{z}_{i-1} \right).
\]

(3)

The main supposition of the ACD model is that the standard durations \( \varepsilon_i = x_i/\psi_i \) are iid with \( E[\varepsilon_i] = 1 \). It implies that \( g \left( x_i | \bar{x}_{i-1}, \bar{z}_{i-1}; \theta_i \right) = g \left( x_i | \psi_i; \theta_i \right) \). Being so, every single temporal dependence of the process of duration is captured by the expected conditional duration. The basic ACD model as proposed by [22] is based on a parametrization of (3) in which \( \psi_i \) depends on \( m \) past duration and on past expected durations.

\[
\psi_i = \omega + \sum_{j=1}^{m} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j}.
\]

(4)

(4) is called ACD(m, q) model. To ensure positive conditional durations for all possible realizations, necessary conditions but not sufficient are \( \omega > 0 \), \( \alpha \geq 0 \) and \( \beta \geq 0 \).
3.1.1 The ACD(1,1) Model

The supposition introduced by [22] is that the dependence can be assumed in durations in conditionals hopes $E[x_i | \mathcal{S}_i]$, as so that $x_i / E[x_i | \mathcal{S}_i]$ is independent and identically distributed. $\mathcal{S}_i$ denotes the group of available information in $t_{i-1}$ time (the beginning of the duration $x_i$), which includes past durations. The ACD model specifies the observed duration as a mixed process given by (i) $x_i = \psi_i \varepsilon_i$, in which $\varepsilon_i$ are positive random variables iid, with (ii) $E[\varepsilon_i] = 1$, (iii) $\text{Var}[\varepsilon_i] = \sigma^2$, as such (iv) $E[x_i | \mathcal{I}_i] = \psi_i$. The second equation specifies an autoregressive model for the conditional duration $\psi_i$: $\psi_i = \omega + \alpha x_{i-1} + \beta \psi_{i-1}$, with the following restrictions over the coefficients: $\omega > 0, \alpha \geq 0$ e $\beta \geq 0$ (com $\beta = 0$ se $\alpha = 0$), e $\alpha + \beta < 1$. The last restriction assures the existence of duration unconditional average. Many options are available for the distribution of $\varepsilon_i$: gamma, exponential, Weibull, Burr, Log-normal, Pareto, generalized gamma, etc., in principle any distribution with positive support. A process defined by the equations (i) to (iv) is called ACD (1, 1) process.

3.2 ACD Variant Models

If assumed that $f(\psi_i) = \psi_i$ and that the standard duration density is an exponential distribution with parameter 1, we make use of exponential distribution because it presents a monotonous risk function which makes it particularly easy to work with. The problem of an “flat” conditional intensity have already been analyzed by [22] as not having a good adjustment with semi-parametric estimates to the risk function, then it was proposed an EACD model extension, generalizing the exponential density of the standard duration to a Weibull density (1, $\gamma$). Therefore to achieve greater flexibility [22] used the Weibull standard distribution as a parameter equal in shape to $\gamma$ and a parameter equal in scale to 1. The resultant model is called WACD. [26] proposed the use of generalized gamma distribution which takes to the GACD model. [33], as well as [31] also used the generalized gamma distribution to characterize de standard durations because it can obtain risk functions in U and inverted U shapes. Either Burr distribution or the generalized gamma distribution allow that risk functions without pre-defined shapes (both depends on two parameters) describe situations in which for short durations, the risk function increases and for long durations the risk function decreases.

A single parameter $\psi^{-1}$ the conditional inverse duration controls the shape and location of the conditional density to the EACD duration models. The WACD model offers greater flexibility by the introduction of an additional parameter to be estimate, the Weibull’s $\gamma$ which affects the location and the shape of the conditional density, distribution and transactions duration costs function. The EACD is the limit of WACD if Weibull’s $\gamma$ estimate approaches the unity.
Although provides greater flexibility than the EACD specification, the Weibull-ACD model can’t be flexible enough to model the stylized fact that features the shape of transaction duration distributions. Plotting Weibull’s density to different Weibull-γ, can be verified a sharp decrease on the histogram of transactions duration that can’t be taken in account by the Weibull distribution. Consequently, modeling the empirical evidence that the mass of the transactions duration distributions is clearly focused in shorter durations and that longer durations do exist but are rarer and also restrict in the WACD specification. Therefore, to model financial transactions process a more flexible tool than WACD is desirable. [38] proposed an alternative ACD specification which is capable of provide more flexibility to the WACD specification and has the EACD model as special case. The constructed model about the Burr distribution and is called Burr-ACD or BACD.

![Graph showing Burr distribution with different values of the localization parameter](image)

The Burr distribution is not usually applied in statistics and econometrics and it is just like a mix Weibull Gamma distributions [24]. The Burr distribution contains the Log-logistic, the Weibull and the Exponential distribution as special cases [24]. The greater the flexibility of Bull distribution over Weibull’s can be seen on Figure-9. It’s obvious that for large values of $\sigma^2$ the Burr density can assume angular shape which is way more compatible with empirical evidences. Figure-9 displays the results of plotted Weibull’s and Burr’s distributions density functions with Weibull-γ and Burr-κ equal to 0.5 and Weibull-λ and Burr-μ equal to 1 with different values of the Burr localization parameter $\sigma^2$. 
Autoregressive conditional duration models

Table-2 Estimate models: EACD, WACD and GACD

<table>
<thead>
<tr>
<th>Parameters</th>
<th>EACD(1, 1)</th>
<th>WACD(1, 1)</th>
<th>GACD(1, 1)</th>
<th>EACD(2, 2)</th>
<th>WACD(2, 2)</th>
<th>Burr ACD(1, 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>0.1645</td>
<td>0.1639</td>
<td>0.1578</td>
<td>0.1465</td>
<td>0.1531</td>
<td>0.1585</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.1617</td>
<td>0.1616</td>
<td>0.1291</td>
<td>0.0028</td>
<td>0.0023</td>
<td>0.1826</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>0.1541</td>
<td>0.1018</td>
<td>-0.1103</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.8231</td>
<td>0.8198</td>
<td>0.8465</td>
<td>0.2542</td>
<td>0.3371</td>
<td>0.8392</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>0.5677</td>
<td>0.5431</td>
<td>---</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1</td>
<td>1.01</td>
<td>0.1321</td>
<td>0.9832</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>---</td>
<td>---</td>
<td>103.6734</td>
<td>---</td>
<td>---</td>
<td>0.7403</td>
</tr>
</tbody>
</table>

$Ljung-Box$ $Q(10)$ | 67.5468 | 64.8891 | 73.9944 | 56.1238 | 71.3150 | 61.7442 |
$p$-value     | 0.0002  | 0.0000  | 0.0000  | 0.0001  | 0.0031  | 0.0001 |

$Ljung-Box$ $Q(10)$ | 8.4532 | 8.9765 | 17.8422 | 9.4320 | 9.0502 | 16.0064 |
$p$-value     | 0.8657  | 0.8501  | 0.3486  | 0.8655  | 0.5639  | 0.3277 |

$Ljung-Box$ $Q(10)$ | 67.5468 | 64.8891 | 73.9944 | 56.1238 | 71.3150 | 61.7442 |
$p$-value     | 0.0002  | 0.0000  | 0.0000  | 0.0001  | 0.0031  | 0.0001 |

$Ljung-Box$ $Q(10)$ | 8.4532 | 8.9765 | 17.8422 | 9.4320 | 9.0502 | 16.0064 |
$p$-value     | 0.8657  | 0.8501  | 0.3486  | 0.8655  | 0.5639  | 0.3277 |

On Table-2 the estimate parameters of the EACD, WACD and GACD models to PETR3 series are shown. It can be verified that the sum of parameters $\alpha$ and $\beta$ is minor than one of the five different types of the estimate models. The estimation procedures were performed by R software based on [39] work on ACD models for irregularly time-spaced data. On Table-3 are also presented the estimate parameters of Burr-ACD model to the PETR3 series. It can be verified that the sum of the parameters $\alpha$ and $\beta$ is minor than the one estimate model.

Tabela-3 Estimated models: EACD, WACD e GACD

<table>
<thead>
<tr>
<th>Parameters</th>
<th>EACD(1, 1)</th>
<th>WACD(1, 1)</th>
<th>GACD(1, 1)</th>
<th>EACD(2, 2)</th>
<th>WACD(2, 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>0.1645</td>
<td>0.1639</td>
<td>0.1578</td>
<td>0.1463</td>
<td>0.1531</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.1617</td>
<td>0.1616</td>
<td>0.1291</td>
<td>0.0028</td>
<td>0.0023</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>0.1541</td>
<td>0.1018</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.8231</td>
<td>0.8198</td>
<td>0.8465</td>
<td>0.2542</td>
<td>0.3371</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>0.5677</td>
<td>0.5431</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1</td>
<td>1.01</td>
<td>0.1321</td>
<td>0.9832</td>
<td>---</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>---</td>
<td>---</td>
<td>103.6734</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

$Ljung-Box$ $Q(10)$ | 67.5468 | 64.8891 | 73.9944 | 56.1238 | 71.3150 | 61.7442 |
$p$-value     | 0.0002  | 0.0000  | 0.0000  | 0.0001  | 0.0031  | 0.0001 |

$Ljung-Box$ $Q(10)$ | 8.4532 | 8.9765 | 17.8422 | 9.4320 | 9.0502 | 16.0064 |
$p$-value     | 0.8657  | 0.8501  | 0.3486  | 0.8655  | 0.5639  | 0.3277 |

$Ljung-Box$ $Q(10)$ | 67.5468 | 64.8891 | 73.9944 | 56.1238 | 71.3150 | 61.7442 |
$p$-value     | 0.0002  | 0.0000  | 0.0000  | 0.0001  | 0.0031  | 0.0001 |

$Ljung-Box$ $Q(10)$ | 8.4532 | 8.9765 | 17.8422 | 9.4320 | 9.0502 | 16.0064 |
$p$-value     | 0.8657  | 0.8501  | 0.3486  | 0.8655  | 0.5639  | 0.3277 |

$\sum_{i}(\alpha_i + \beta_i)$ | 0.9848 | 0.9814 | 0.9756 | 0.9789 | 0.9843 |
Table-3 shows the estimated parameters of the models EACD, WACD and GACD for PETR3 series. It can be seen that the sum of the parameters $\alpha$ and $\beta$ are less than one in five types different models estimated. Estimation procedures were performed in the R software, rely on the work of [39] on ACD models to data irregularly spaced in time. Table-4 shows the estimated parameters of the Burr-ACD model for PETR3 series. It can be seen that the sum of the parameters $\alpha$ and $\beta$ is less than one on the estimated model.

Table 4. Estimated model: BACD

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Burr-ACD(1, 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_0$</td>
<td>0.1585</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.1826</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>-0.1103</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.8392</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>---</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.7403</td>
</tr>
</tbody>
</table>

Ljung-Box for $\hat{\varepsilon}_i$ Q(10) 71.315

p-value 0.0031

Ljung-Box for $\hat{\varepsilon}_i^2$ Q(10) 9.0502

p-value 0.5639

$\sum_i (\alpha_i + \beta_i)$ 0.9115

4 Conclusion

Until two decades ago, most of the empirical finance studies habitually used daily data obtained from the first and the last observation of the day as an variable of interest (i.e. the closing price), neglecting all intraday events. However, due to increasingly automation of financial markets and the quick evolution on the increase of computational power, more and more exchanges and trades created intraday databases to record each transaction as well as its characteristics (such as price, volume, etc.) The availability of those intraday data groupings at low costs boosted the development of a new area of financial research: the high-frequency data analysis. Pulling together finances, econometrics and statistics of temporal series, the high-frequency data analysis quickly emerged as a promising way of research, making it easier to understand more profoundly market activities. Curiously, this evolution has not been limited to academic works but also has affected today’s commercial environment. In the last years, the trading speed has increased consistently. One-day trades once exclusive territory of stock market traders, are now available to all investors. High-frequency hedge-funds appeared
as a new and well succeeded class of hedge fund. The intrinsic limit of high-frequency data is represented by the transaction or tick-by-tick data in which the events are registered one by one as they arise. Consequently, the distinctive feature of this kind of data is that the observations are irregularly time-spaced. This feature has challenged researchers and, as shown in the last couple years, turned traditional econometric techniques into not directly applicable anymore. Besides, the newest models of market microstructure literature reason that time can broadcast information and therefore must be modeled as well. Motivated by those considerations, Engle and Russel (1998) developed the Autoregressive Conditional Duration (ACD) model whose explicit goal is to model time and events. On the first five days of analysis of the PETR3 high-frequency data, it was verified the diurnal effect over the durations average that presents an inverted U shape, mainly because of the market operation with no negative trading news. However it was also verified on the last five days of analysis that the inverted U pattern didn’t occur. To contemplate this scenario it is important to use a model proposed by Rooy (2006) due to the diurnal time curve duration shape. Since its introduction the ACD model and its multiples extensions became an essential tool on modeling the behavior of financial data irregularly time-spaced, opening a door to empirical and theoretical development. As proposed by Engle and Russel (1998) the ACD model shares many features with the GARCH model. This theoretical structure has been supporting most of econometrics techniques on high-frequency data. The results show that the Burr-ACD model contains the EACD and WACD models as special cases. Although in the Burr-ACD model may be necessary greater efforts to implement and evaluate of than in the standard ACD, the advantage is that the conditional density and the Burr-ACD model transactions duration survival function are less restrict and could assume more realistic forms.

Acknowledgements. The authors thank Financiadora de Estudos e Projetos FINEP for the resources which enabled the deployment of MQUANT laboratory, whose activities is part of the preparation of this work.

References


Received: November 10, 2015; Published: April 30, 2016