Central European Journal of Economic Modelling and Econometrics

The UHF-GARCH-Type Model in the Analysis of Intraday Volatility and Price Durations – the Bayesian Approach

Roman Huptas*

Submitted: 8.04.2014, Accepted: 3.12.2015

Abstract

In empirical research on financial market microstructure and in testing some predictions from the market microstructure literature, the behavior of some characteristics of trading process can be very important and useful. Among all characteristics associated with tick-by-tick data, the trading time and the price seem the most important. The very first joint model for prices and durations, the so-called UHF-GARCH, has been introduced by Engle (2000). The main aim of this paper is to propose a simple, novel extension of Engle's specification based on trade-to-trade data and to develop and apply the Bayesian approach to estimation of this model. The intraday dynamics of the return volatility is modelled by an EGARCH-type specification adapted to irregularly time-spaced data. In the analysis of price durations, the Box-Cox ACD model with the generalized gamma distribution for the error term is considered. To the best of our knowledge, the UHF-GARCH model with such a combination of the EGARCH and the Box-Cox ACD structures has not been studied in the literature so far. To estimate the model, the Bayesian approach is adopted. Finally, the methodology developed in the paper is employed to analyze transaction data from the Polish Stock Market.

Keywords: intraday volatility, UHF-GARCH-type model, ACD model, transaction data, Bayesian inference

JEL Classification: C55, C58, C11, C22

*Cracow University of Economics; e-mail: huptasr@uek.krakow.pl

1 Introduction

The analysis of trade-to-trade data from financial markets has become one of the most important areas of research in financial econometrics and empirical finance. It has allowed researchers to get a better understanding of the dynamics of price formation during a trading day. The availability of data on individual transactions provides a deeper insight into the so-called market microstructure. The transaction data contains information on the time of the trade of a given asset as well as some other associated characteristics such as trading price and volume. Among all, the exact time of the trade and the price with which the trade was made are crucial. The trading time may convey information valuable to the market traders, therefore inducing and influencing their particular actions. The analysis of time intervals between successive events of the transaction process (durations) can provide one with information about the microstructure of financial markets, thereby allowing for a more detailed insight into the various types of dependencies prevailing on the market. Durations play an important role in the market microstructure theory, where they are used as a proxy variable indicative of new information arrivals on the market (see, for example, Easley and O'Hara 1992).

The econometric modelling of financial durations has been a rapidly growing field of research since the paper of Engle and Russell (1998), who introduced the autoregressive conditional duration (ACD) model. Many extensions have been suggested to improve upon the original Engle and Russell's specification; see, e.g., Bauwens and Giot (2000), Lunde (1999), Hautsch (2002), Dufour and Engle (2000), Fernandes and Grammig (2006), Hautsch (2004, 2012), Zhang, Russell and Tsay (2001).

The very first joint model for returns and price durations has been introduced by Engle (2000) and termed as the UHF-GARCH model. Engle (2000) puts the duration framework of Engle and Russell (1998) into the volatility structure and measures the impact of transaction duration on intraday price volatility. Ghysels and Jasiak (1998), Grammig and Wellner (2002), and Liu and Maheu (2012) propose some other models to investigate the relationships between price durations and return volatility, whereas Manganelli (2005), Doman and Doman (2012), and Doman (2011, in Polish) model durations, volumes and returns simultaneously.

The main objective of this paper is to propose a simple, novel extension of Engle's UHF-GARCH-type model and to develop and apply the Bayesian approach to estimate it. In the case of the UHF-GARCH-type and ACD structures, the inference about parameters is usually based on the Maximum Likelihood (ML) method. Due to nonlinearities and still not so-well known properties of the maximum likelihood estimators for the UHF-GARCH-type models and ACD specifications with conditional distributions other than the exponential one, the Bayesian approach, relying on the Monte Carlo methods, seems to provide a natural, theoretically consistent and practically valid estimation method. Moreover, a practical use of this model in the analysis of the intraday dynamics of volatility and price durations on the Polish stock exchange is presented, with the aim of studying the dependencies between volatility and contemporaneous price durations on the Warsaw Stock Exchange (WSE) in Poland.

The structure of the paper is as follows. In the following section we discuss at length the basic definition of the ACD process, and present the Box-Cox ACD specification as a particular extension of the basic ACD structure. In Section 3, we introduce the BCACD-AR-EGARCH model – a joint structure for modelling intraday returns and price durations simultaneously. Bayesian estimation of the model in question is discussed in Section 4. In Section 5, the methodology developed in the paper is illustrated with an empirical study, using tick-by-tick data of a company listed in the WIG20 index of the Warsaw Stock Exchange. The final section contains concluding remarks.

2 ACD models in duration analysis

Let us consider a sequence of moments $t_1,\ t_2,\ \ldots,\ t_i,\ \ldots$ in which events of the transaction process occur. In the presented analysis, it represents a sequence of consecutive moments of changes in the transaction prices. The time interval between successive events (price changes) of the transaction process that occur at the moments t_i and t_{i-1} is denoted as $x_i = t_i - t_{i-1}$, and is henceforth referred to as the duration. Let Ψ_i represent the conditional expectation of the duration given the information about the process available at time t_{i-1}, i.e.:

$$\Psi_i = E\left(x_i|\Im_{i-1}\right),$$

where $\Im_{i-1}$ denotes the set of information available prior to and including moment t_{i-1}. This set comprises past durations up to and including x_{i-1}, and it may also contain some pre-determined variables of the transaction process suggested by the market microstructure literature.

The autoregressive conditional duration (ACD) models are one of the primary tools used in modelling time intervals between events of the transaction process, analysing trading intensity and examining the effects of financial market microstructure. The ACD specification was proposed for modelling of the dynamics of financial durations by Engle and Russell (1998), with the main idea behind it involving a dynamic parameterisation of the conditional expected duration, Ψ_i:

$$\Psi_i = E\left(x_i|\Im_{i-1}\right) = E\left(x_i|x_{i-1},\ldots,x_1\right) = \Psi_i\left(x_{i-1},\ldots,x_1;\theta\right), \tag{1}$$

where θ denotes a certain vector of parameters.

In the ACD model, duration x_i is expressed as the following product:

$$x_i = \Psi_i \cdot \varepsilon_i, \tag{2}$$

where ε_i is an i.i.d. process defined over a positive support with density function $f_\varepsilon(\varepsilon_i)$ and expected value $E(\varepsilon_i) = 1$. The setup defined by Equations (1)-(2) is very

general and allows for a variety of models. Different types of the ACD structure may result either from various functional forms for the conditional mean function Ψ_i, or the selection of different probability distributions for the random variable ε_i. As far as the distribution of innovations ε_i in the ACD models is concerned, only probability measures defined over the set of positive real numbers are allowed (cf. Engle and Russell 1998, Lunde 1999, Grammig and Maurer 2000, Bauwens and Giot 2001, 2003, Hautsch 2002, Bauwens, Giot, Grammig and Veredas 2004, De Luca and Gallo 2004, Fernandes and Grammig 2006, De Luca and Zuccolotto 2006, Allen, Chan, McAleer and Peiris 2008). The most common and simplest distribution for ε_i is the exponential one. Other choices include, e.g., the Weibull distribution, gamma and generalized gamma distributions, and the Burr distributions.
According to the form of the conditional mean equation of the basic ACD specification, proposed by Engle and Russell (1998), the model is based on a linear parameterisation of the expected duration dynamics. The ACD(1,1) linear process is the lowest-order ACD specification considered in empirical research, and is given by the equation:

$$\Psi_i = \omega + \alpha \cdot x_{i-1} + \beta \cdot \Psi_{i-1}, \tag{3}$$

where $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, $\alpha + \beta < 1$. These inequality restrictions are intended to ensure positive conditional durations for all possible realizations. It is worth noting that these constraints are sufficient, although not necessary, for the non-negativity of the duration process.
With respect to the specification of the functional form of the conditional expected duration Ψ_i, many extensions to the original linear ACD model have been proposed in the literature, such as the logarithmic ACD models (cf. Bauwens and Giot 2000, Lunde 1999), the Box-Cox ACD models (cf. Hautsch 2002), the exponential ACD model (cf. Dufour and Engle 2000), the asymmetric logarithmic ACD model (cf. Fernandes and Grammig 2006), the augmented Box-Cox ACD model (cf. Hautsch 2004, 2012), the augmented ACD model (cf. Fernandes and Grammig 2006), the threshold ACD model (cf. Zhang, Russell and Tsay 2001) and many others.

2.1 The Box-Cox ACD model with the generalized gamma distribution

The specification of the ACD model for duration process used in this research is of the form:

$$\begin{aligned} x_i &= \Psi_i \cdot \varepsilon_i, \\ \Psi_i^{\delta_1} &= \omega + \alpha \cdot \varepsilon_{i-1}^{\delta_2} + \beta \cdot \Psi_{i-1}^{\delta_1}, \end{aligned} \tag{4}$$

where $\omega > 0$, $\alpha > 0$, $0 < \beta < 1$, $\delta_1 > 0$, $\delta_2 > 0$. Moreover, we assume that ε_i follows the generalized gamma distribution (under the assumption that $E(\varepsilon_i) = 1$) that is $\varepsilon_i \sim GG(\lambda, \gamma, \upsilon)$ with parameter $\lambda = \left(\Gamma\left(\frac{\upsilon}{\gamma}\right) \big/ \Gamma\left(\frac{1+\upsilon}{\gamma}\right)\right)^{\gamma}$ and with density function given by:

$$f_{\varepsilon}(\varepsilon_i) = \frac{\gamma}{\lambda^{\frac{\upsilon}{\gamma}} \cdot \Gamma\left(\frac{\upsilon}{\gamma}\right)} \cdot \varepsilon_i^{\upsilon-1} \cdot \exp\left[-\frac{\varepsilon_i^{\gamma}}{\lambda}\right], \quad \varepsilon_i > 0,\ \lambda > 0,\ \gamma > 0,\ \upsilon > 0.$$

The ACD specification based on the Box and Cox power transformations of Ψ_i and ε_i was proposed by Hautsch (2002) and it allows for concave and convex news impact curves depending on the values of parameters δ_2 and δ_1. Moreover, it nests logarithmic ACD specifications for $\delta_1 \to 0$, $\delta_2 \to 0$ and for $\delta_1 \to 0$, $\delta_2 = 1$. For $\delta_1 = \delta_2 = 1$ we obtain a special case of the linear ACD model. Also, for $\delta_1 \to 0$ it coincides with the Box-Cox ACD specification suggested by Dufour and Engle (2000). The generalized gamma distribution for the innovation term in the context of ACD processes was proposed by Lunde (1999) and further employed by Bauwens, Giot, Grammig and Veredas (2004), who allow for a more flexible innovation distribution, as compared with the basic choice of the exponential one.
Under the assumption that $E(\varepsilon_i) = 1$, the conditional density for price duration in the case of an ACD model with the generalized gamma innovation distribution can be written as:

$$\begin{aligned} f\left(x_i|\Im_{i-1},\theta\right) &= \frac{1}{\Psi_i} \cdot f_{\varepsilon}\left(\frac{x_i}{\Psi_i}\right) = \\ &= \frac{\gamma}{x_i \cdot \Gamma\left(\frac{\upsilon}{\gamma}\right)} \cdot \left(\frac{x_i}{\Psi_i} \cdot \frac{\Gamma\left(\frac{1+\upsilon}{\gamma}\right)}{\Gamma\left(\frac{\upsilon}{\gamma}\right)}\right)^{\upsilon} \cdot \exp\left[-\left(\frac{x_i}{\Psi_i} \cdot \frac{\Gamma\left(\frac{1+\upsilon}{\gamma}\right)}{\Gamma\left(\frac{\upsilon}{\gamma}\right)}\right)^{\gamma}\right] \end{aligned} \tag{5}$$

with $\gamma, \upsilon > 0$, whereas Ψ_i is determined by Equation (4), and $f_{\varepsilon}(\cdot)$ denotes the density function of the generalized gamma distribution.

3 The BCACD-AR-EGARCH model

In the previous section we discussed an approach to modelling price durations. In what follows, we focus solely on the Box-Cox ACD specification with the generalized gamma distribution to model the price duration process. We proceed to specification of the return dynamics, associated with the trades where there is a change in the transaction price. Some approaches to modelling the returns in the case of irregularly spaced tick-by-tick data were proposed by Engle (2000), Ghysels and Jasiak (1998) and Grammig and Wellner (2002). A formal comparison of these models can be found in Meddahi, Renault and Werker (2005), for example. In line with the papers mentioned above, there is also the article by Liu and Maheu (2012). We follow Engle (2000) and Liu and Maheu (2012) to formulate our model and present a simple, novel generalization of their proposals.
Let us consider a sequence of moments $t_1, t_2, \ldots, t_i, \ldots$ in which the price changes,

and a sequence of corresponding transaction prices $p_1, p_2, \ldots, p_i, \ldots$. Percentage logarithmic rates of return corresponding to the price changes are denoted by $\tilde{r}_i$. Engle (2000) argues that a suitable and natural measure of volatility is the variance per unit of time. Therefore, we build our model based on the return per square root of time, defined as $r_i = \tilde{r}_i/\sqrt{x_i}$. In such a situation the data are just a sequence of joint observations of the price duration and return given by $\{y_i = (r_i, x_i),\ i = 1, \ldots, T\}$, where T is the number of observations. Therefore, the joint density can be written as the product of the marginal density of price durations, $f(x_i|\Im_{i-1};\theta)$, and the conditional density of the returns given the durations, $f(r_i|x_i, \Im_{i-1};\theta)$, all conditioned upon the past of durations and returns:

$$f(r_i, x_i|\Im_{i-1};\theta) = f(x_i|\Im_{i-1};\theta) \cdot f(r_i|x_i, \Im_{i-1};\theta). \tag{6}$$

Drawing on Engle's (2000) model, we assume a simple AR(1) structure for the returns:

$$r_i - \delta = \rho \cdot (r_{i-1} - \delta) + u_i \tag{7}$$

where $-1 < \rho < 1$ and the innovation term u_i is given by

$$u_i = \sigma_i \cdot \zeta_i \tag{8}$$

with $\zeta_i \sim$ i.i.d $t(0;1;\nu)$, $\nu > 2$ and σ_i^2 is the variance of the return conditional on the past volatility and duration information. By $t(0;1;\nu)$ we denote Student's t distribution with zero mean, unit precision and an unknown number of degrees of freedom $\nu > 2$. Several specifications of the GARCH-type model to capture the dynamics of conditional variance were suggested by Engle (2000). In contrast to Engle's (2000) approach, we propose an EGARCH(1,1)-type specification of the conditional variance, the dynamics of which evolves according to the equation:

$$\begin{aligned} \ln \sigma_i^2 = {} & \omega_G + \alpha_{1G} \cdot \zeta_{i-1} + \alpha_{2G} \cdot \left(|\zeta_{i-1}| - \frac{2\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi(\nu-2)}} \right) \\ & + \beta_G \cdot \ln \sigma_{i-1}^2 + \eta_1 \cdot \frac{1}{x_i} + \eta_2 \cdot \frac{x_i}{\Psi_i} + \eta_3 \cdot \frac{1}{\Psi_i} \end{aligned} \tag{9}$$

where $\beta_G < 1$, and Ψ_i is determined by Equation (4). The impact of durations on volatility is incorporated by means of three additional microstructure variables: the reciprocal of the duration, x_i^{-1}, the reciprocal of the expected duration, Ψ_i^{-1}, and the effect of surprise in duration, $\frac{x_i}{\Psi_i}$ (cf. Engle 2000).

This specification allows for an asymmetric response of σ_i^2 to volatility shocks in the innovation term ζ_{i-1} when parameter α_{1G} differs from zero. The use of an EGARCH-type model is also justified by the advantage of keeping the volatility component positive regardless of the sign of the right-hand side components in the volatility equation. The absence of non-negativity constraints on the parameters also facilitates numerical estimation. Finally, as mentioned at the beginning of this section, the Box-Cox ACD model is employed to model the price duration process.

4 Bayesian estimation of the BCACD-AR-EGARCH model

In the case of the UHF-GARCH-type and ACD models, inference about the parameters is usually based on the Maximum Likelihood (ML) method. Due to nonlinearities and still not so-well known properties of the ML estimators for the UHF-GARCH-type structures and ACD models with conditional distributions other than the exponential one, the Bayesian approach, relying on the Monte Carlo methods, seems to provide a natural, valuable and theoretically consistent estimation method (cf. Zellner 1971, Osiewalski 2001 in Polish). Therefore, it is employed in our study and developed to estimate the proposed BCACD-AR-EGARCH structure. Within the Bayesian paradigm, all unknown quantities are treated as random variables. Let us remind that the sequence of joint observations of the price duration and return is denoted by $\{y_i = (r_i, x_i),\ i = 1, \ldots, T\}$, whereas the model parameters – by θ. The joint density of observations collected in $y = (y_1, \ldots, y_T)'$ and the parameters, determining the Bayesian model, can be factorized as follows:

$$\begin{aligned} p(y, \theta|y_{(0)}) &= p(\theta) \cdot p(y|\theta; y_{(0)}) \\ &= p(\theta) \cdot \prod_{i=1}^{T} f(r_i, x_i|\Im_{i-1}, \theta; y_{(0)}) \\ &= p(\theta) \cdot \prod_{i=1}^{T} f(x_i|\Im_{i-1}, \theta; y_{(0)}) \cdot f(r_i|x_i, \Im_{i-1}, \theta; y_{(0)}) \end{aligned} \tag{10}$$

where $f(x_i|\Im_{i-1}, \theta; y_{(0)})$ is given in Equation (5) and $y_{(0)} = (r_0, x_0, \Psi_0)$ is the vector of initial conditions which was hitherto omitted in the notation. The conditional density of returns, $f(r_i|x_i, \Im_{i-1}, \theta; y_{(0)})$, appearing in Equation (10), takes the form:

$$\begin{aligned} f(r_i|x_i, \Im_{i-1}, \theta; y_{(0)}) &= f_u(u_i|x_i, \Im_{i-1}, \theta; y_{(0)}) \\ &= f_u(r_i - \delta - \rho \cdot (r_{i-1} - \delta)|x_i, \Im_{i-1}, \theta; y_{(0)}) \\ &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\pi(\nu-2)\sigma_i^2}} \cdot \left(1 + \frac{(r_i - \delta - \rho \cdot (r_{i-1} - \delta))^2}{(\nu-2)\sigma_i^2}\right)^{-\frac{1}{2}(\nu+1)}, \end{aligned} \tag{11}$$

where $f_u(\cdot)$ refers to Student's t density function for u_i, and σ_i^2 is determined by Equation (9).

In order to complete the Bayesian model, the prior distribution of parameters needs to be specified. We assume a proper joint prior and subjectively set the marginal priors of interest so as to reflect our vague knowledge about model parameters. Assuming also prior independence among the parameters, the joint prior factorizes as:

$$\begin{aligned} p(\theta) = p(\rho)p(\delta)p(\omega_G)p(\alpha_{1G})p(\alpha_{2G})p(\beta_G)p(\eta_1)p(\eta_2)p(\eta_3)p(\nu) \cdot \\ \cdot p(\omega)p(\alpha)p(\beta)p(\delta_1)p(\delta_2)p(\gamma)p(\upsilon). \end{aligned} \tag{12}$$

The details on each marginal prior specification are provided below:

$p(\rho) = \frac{1}{2} 1_{[-1;\,1]}(\rho)$ – the uniform distribution over the interval $[-1;\ 1]$,

$p(\delta) = f_N(\delta|\mu_\delta, \sigma^2_\delta)$, $\mu_\delta = 0$, $\sigma_\delta = 5$,

$p(\omega_G) = f_N(\omega_G|\mu_{\omega_G}, \sigma^2_{\omega_G})$, $\mu_{\omega_G} = 0$, $\sigma_{\omega_G} = 5$,

$p(\alpha_{1G}) = f_N(\alpha_{1G}|\mu_{\alpha_{1G}}, \sigma^2_{\alpha_{1G}})$, $\mu_{\alpha_{1G}} = 0$, $\sigma_{\alpha_{1G}} = 5$,

$p(\alpha_{2G}) = f_N(\alpha_{2G}|\mu_{\alpha_{2G}}, \sigma^2_{\alpha_{2G}})$, $\mu_{\alpha_{2G}} = 0$, $\sigma_{\alpha_{2G}} = 5$,

$p(\beta_G) \propto f_N(\beta_G|\mu_{\beta_G}, \sigma^2_{\beta_G}) \cdot 1_{(-\infty,1)}(\beta_G)$, $\mu_{\beta_G} = 0$, $\sigma_{\beta_G} = 5$,

$p(\eta_1) = f_N(\eta_1|\mu_{\eta_1}, \sigma^2_{\eta_1})$, $\mu_{\eta_1} = 0$, $\sigma_{\eta_1} = 5$,

$p(\eta_2) = f_N(\eta_2|\mu_{\eta_2}, \sigma^2_{\eta_2})$, $\mu_{\eta_2} = 0$, $\sigma_{\eta_2} = 5$,

$p(\eta_3) = f_N(\eta_3|\mu_{\eta_3}, \sigma^2_{\eta_3})$, $\mu_{\eta_3} = 0$, $\sigma_{\eta_3} = 5$,

$p(\nu) = 0.01 \cdot 1_{[2;\,102]}(\nu)$ – the uniform distribution over the interval $[2;\ 102]$,

$p(\omega) \propto f_N(\omega|\mu_\omega, \sigma^2_\omega) \cdot 1_{(0,\infty)}(\omega)$, $\mu_\omega = 0$, $\sigma_\omega = 5$,

$p(\alpha) \propto f_N(\alpha|\mu_\alpha, \sigma^2_\alpha) \cdot 1_{(0,\infty)}(\alpha)$, $\mu_\alpha = 0$, $\sigma_\alpha = 5$,

$p(\beta) \propto f_N(\beta|\mu_\beta, \sigma^2_\beta) \cdot 1_{(0,1)}(\beta)$, $\mu_\beta = 0$, $\sigma_\beta = 5$,

$p(\delta_1) \propto f_N(\delta_1|\mu_{\delta_1}, \sigma^2_{\delta_1}) \cdot 1_{(0,\infty)}(\delta_1)$, $\mu_{\delta_1} = 0$, $\sigma_{\delta_1} = 5$,

$p(\delta_2) \propto f_N(\delta_2|\mu_{\delta_2}, \sigma^2_{\delta_2}) \cdot 1_{(0,\infty)}(\delta_2)$, $\mu_{\delta_2} = 0$, $\sigma_{\delta_2} = 5$,

$p(\gamma) \propto f_N(\gamma|\mu_\gamma, \sigma^2_\gamma) \cdot 1_{(0,\infty)}(\gamma)$, $\mu_\gamma = 0$, $\sigma_\gamma = 5$,

$p(\upsilon) \propto f_N(\upsilon|\mu_\upsilon, \sigma^2_\upsilon) \cdot 1_{(0,\infty)}(\upsilon)$, $\mu_\upsilon = 0$, $\sigma_\upsilon = 10$,

where $f_N(\cdot|\mu_0, \sigma^2_0)$ denotes the density of the normal distribution with mean μ_0 and variance of σ^2_0. We now can write the joint density function that represents our Bayesian BCACD(1,1)-AR(1)-EGARCH(1,1) model as:

$$
\begin{aligned}
p(y, \theta|y_{(0)}) = {} & p(\rho)p(\delta)p(\omega_G)p(\alpha_{1G})p(\alpha_{2G})p(\beta_G)p(\eta_1)p(\eta_2)p(\eta_3)p(\nu)\cdot \\
& \cdot p(\omega)p(\alpha)p(\beta)p(\delta_1)p(\delta_2)p(\gamma)p(\upsilon)\cdot \\
& \cdot \prod_{i=1}^{T} f(x_i|\Im_{i-1}, \theta; y_{(0)}) \cdot f(r_i|x_i, \Im_{i-1}, \theta; y_{(0)}),
\end{aligned}
$$

where $f(x_i|\Im_{i-1}, \theta; y_{(0)})$ and $f(r_i|x_i, \Im_{i-1}, \theta; y_{(0)})$ are determined by Equations (5) and (11), respectively.

5 Empirical study

5.1 Description of the data set

In the empirical study to follow we use data from the Warsaw Stock Exchange, which is currently a market with the highest capitalization in Eastern and Central Europe. Thus, it is noticed as one of the most important and the best developed market in that region. However, the literature on the ACD models, the UHF-GARCH-type models and their empirical applications to Polish stock data is only limited; see e.g. Bień (2004, 2006, 2006a, in Polish), Doman (2005), Doman (2008), Doman and Doman (2010), Doman (2011, in Polish), Bień-Barkowska (2011, 2012, 2014, 2014a), and Huptas (2014). This study attempts to fill that gap to some extent. Moreover, it also contributes to the existing knowledge about the Polish stock market microstructure.
The empirical analysis is based on the transaction data of the Polish Telecommunications (TPSA, currently Orange Polska S.A.) company listed in the WIG 20 Index, quoted at the Warsaw Stock Exchange. The data comprise tick-by-tick observations between 23 March 2009 and 19 June 2009, and have been obtained from the website of stooq.pl. The transaction data sheets contain information on the transaction closing price and volume along with the date and time of each transaction with the accuracy of one second.
The data are partially aggregated. The details of the data preparations are as follows. The analysis covers only transactions carried out in the continuous trading phase, which in the case of the Warsaw Stock Exchange falls on between 10:00 and 16:10. Following the most prevalent approach adopted in the literature, the time intervals between the closure of the session and the beginning of the next day's session are removed. Price durations are measured with the accuracy of one second. All transactions occurring at the same time but with different prices are integrated into a single transaction with the price of transaction calculated as an average weighted by volume. The volumes of transactions made at the same price are summed up. Table 1 displays the reduction in the number of observations after data filtering.

Table 1: Data filtering – the number of observations before and after data aggregation

	The number of data points
The number of the tick-by-tick observations	121782 (100%)
The number of the returns corresponding to the price durations (the number of observations with price change)	29842 (24.5%)

The basic descriptive statistics of price durations and returns corresponding to the price durations for the company surveyed are shown in Table 2. The dynamics of price durations and corresponding returns can be seen in Figure 1. We can observe some important features of the empirical distributions of price durations and corresponding returns. Analysis of the descriptive statistics of price durations distribution reveals its overdispersion, i.e., the standard deviation exceeds the mean (the coefficient of variation is very high reaching the level of 1.59). In addition, the upper plot in Figure 1 clearly indicates clustering of short and long price durations. This suggests the presence of a strong positive autocorrelation in the duration series. The intraday returns seem to oscillate around zero, featuring noticeable outliers and time-variable volatility. We can also observe very high kurtosis of returns (see Table 2) and the volatility clustering phenomenon (see Figure 1).

Figure 1: Plots of price durations and corresponding returns for TPSA

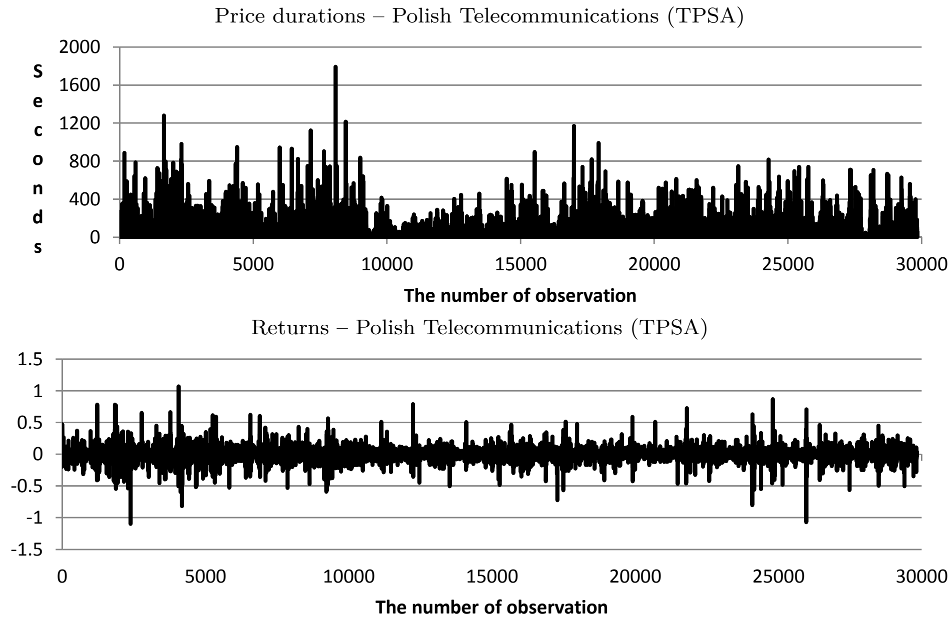
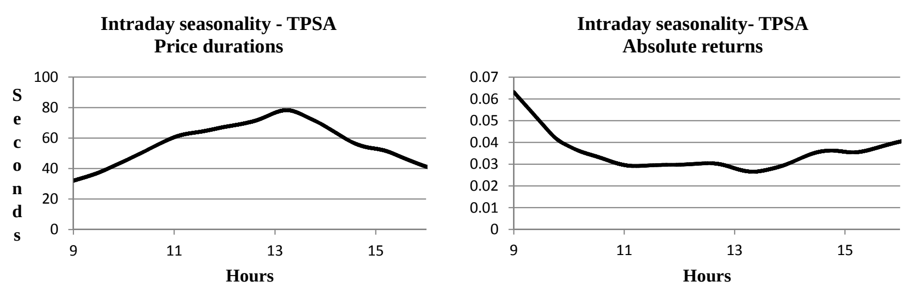


Table 2: Descriptive statistics of price durations and returns for TPSA

	Price durations	Returns
The number of observations	29842	29842
Mean	50.61	-0.0004
Standard deviation (SD)	80.78	0.064
Coefficient of variation (=SD/Mean)	1.59	–
Kurtosis	–	29.94
Minimum	1	-1.09
Maximum	1791	1.07

It is well documented in the financial literature that there exists a clear intraday seasonal (periodic or diurnal) pattern over the trading day (cf. Engle and Russell 1998, Engle 2000). We estimate the intraday seasonality pattern for price durations using the Nadaraya-Watson kernel estimator of regression of the duration on the time of the day (cf. Bauwens and Veredas 2004, Huptas 2009). In the same way we regress the absolute returns on the time of the day. We illustrate the time-of-day functions in Figure 2. It must be stressed that intraday seasonality patterns are consistent with daily information flow. Price durations are shorter just after the opening and just before the closure of the market, indicating more transactions at these times of the day. On the other hand, the durations are much longer around the lunch time. This results in an inverted U-shaped pattern. Absolute returns corresponding to the price durations are higher at the beginning of the day and then the diurnal pattern is rather flat during the rest of the day until it increases a little again before the market closure. Therefore, an L-shaped or perhaps even U-shaped seasonality pattern for the absolute returns is observed. High trading activities after the market opening are explained by the fact that traders try to incorporate information flowed over night. Then the market activity steadily declines. After the lunch time many investors tend to adjust their positions before the market closes, so the trading activity increases. Following financial literature on the subject (cf. Engle and Russell 1998), after estimating the diurnal factor we have to eliminate seasonality from the data. We remove it by dividing plain price durations by their corresponding estimated diurnal component. In turn, diurnally adjusted returns are obtained by taking the ratios of the returns and their corresponding diurnal factor.

Figure 2: Intraday seasonality patterns for the analysed data



5.2 Bayesian estimation results

Below we present the results of Bayesian estimation of the proposed BCACD-AR-EGARCH model. The joint posterior is too complicated to obtain any analytical results. In such a situation, we resort to well-known MCMC techniques in order to generate a pseudo-random sample from the posterior distribution. Since the joint and conditional posteriors do not belong to any familiar family of distributions, the Gibbs sampler cannot be used, and the Metropolis-Hastings (MH) algorithm must be adopted instead (see Hastings 1970). In our study, we employ the Metropolis and Hastings algorithm with a symmetric proposal density. As the candidate generating distribution we use the multivariate t distribution with three degrees of freedom, the expected value set to equal the previous state of the Markov chain, and the covariance matrix obtained from initial cycles, which were performed to calibrate the sampling mechanism. Our results are based on 1 million states of the Markov chain, generated after 200,000 burnt-in draws. To assess the convergence of the Metropolis and Hastings algorithm we use standardised CUMSUM plots (cf. Yu and Mykland 1994). All presented results are obtained using author's own codes implemented in the GAUSS 13.0 Mathematical and Statistical System.

Bayesian estimation results, including marginal posterior means and standard deviations (in parentheses), are reported in Table 3. Figures 3 and 4 depict the marginal posterior distributions of model parameters. It is seen that the posterior marginals are affected predominantly by the information contained in the data. For all the parameters, the posterior densities are sharply distinguishable from their prior counterparts, providing evidence of a strong data contribution to the inference. As compared with the priors, the posterior distributions are characterised by a different location and a markedly smaller dispersion.

Looking at the results for the return equation parameters, we note that the posterior mean of the autoregressive coefficient (ρ) is negative and equal to -0.1791. Moreover, the posterior distribution of ρ is well-separated from zero and features relatively little dispersion (as indicated by the standard deviation of about 0.0043). This slightly negative autocorrelation indicates the existence of the bid-ask bounce effect in the returns and is consistent with the presence of market microstructure dynamics (see Roll 1984). Moreover, as implied by the marginal posterior for the degrees of freedom, the conditional normality of the returns is strongly overridden by the data. The posterior mean and standard deviation of ν equal about 4.5179 and 0.1024, respectively. Therefore, our results confirm that allowing for fat tails of the conditional distribution may be crucial for empirically adequate statistical modelling with the use of UHF-GARCH-type processes.

Now let us focus on the parameters of the Box-Cox ACD equation. The marginal posteriors of the conditional generalized gamma distribution's parameters clearly indicate that the data disqualify the conditional exponential distribution of price durations (the latter results from the conditional generalized gamma collapses under $\gamma = 1$ and $\upsilon = 1$). It is worth noting that the results also exclude conditional Weibull distribution, the latter being a special case of the conditional generalized gamma distribution under equality constraint $\gamma = \upsilon$. The posterior distribution of γ is located on the far left of the value $\gamma = 1$ and reveals a fairly small dispersion. The posterior distribution of υ, however, is well-separated from the value $\upsilon = 1$, with the

Figure 3: Marginal posterior distributions (bars) and priors (solid lines) of parameters of the return and volatility equations within the BCACD-AR-EGARCH model

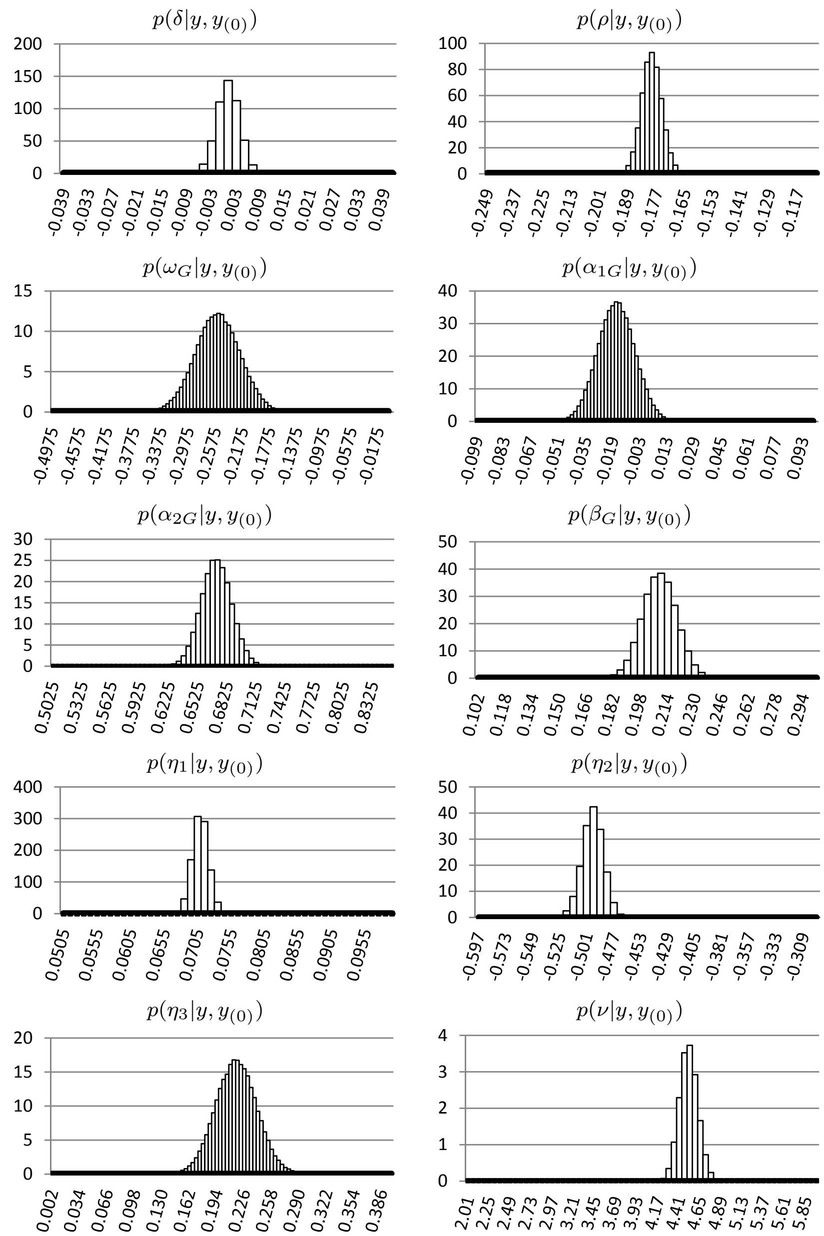
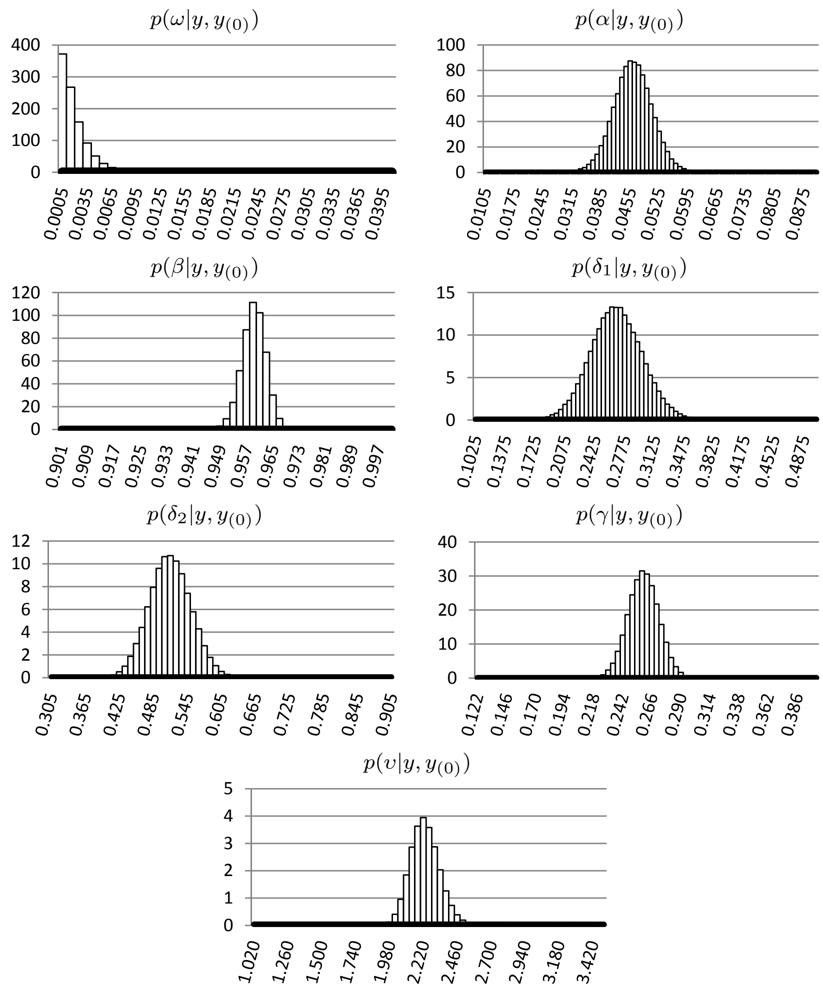


Figure 4: Marginal posterior distributions (bars) and priors (solid lines) of parameters of the Box-Cox-ACD equation within the BCACD-AR-EGARCH model



posterior mean equal to about 2.23 and a relatively small standard deviation of ca. 0.1 (see Table 3). The above-mentioned features of the pertaining marginal posteriors are also seen in the histograms presented in Figure 4.

The results clearly indicate that the effect of constant conditional price durations (taking place under $\alpha = 0$ and $\beta = 0$) is strongly rejected by the modelled data. The posterior marginal distributions of α or β are well-separated from zero. In order to fully describe the dynamics of the analysed price durations, the properties

Table 3: Posterior means and standard deviations (in parentheses) of parameters in the BCACD-AR-EGARCH model with t distribution – TPSA company

The return equation							
		δ	0.0010 (0.0027)	ρ	−0.1791 (0.0043)		
The volatility equation							
ω_G	−0.2538 (0.0327)	α_{1G}	−0.0169 (0.0109)	α_{2G}	0.6713 (0.0157)	β_G	0.2091 (0.0102)
η_1	0.0709 (0.0012)	η_2	−0.4958 (0.0094)	η_3	0.2168 (0.0239)	ν	4.5179 (0.1042)
The Box-Cox-ACD equation							
ω	0.0019 (0.0017)	α	0.0461 (0.0047)	β	0.9592 (0.0036)	δ_1	0.2677 (0.0304)
δ_2	0.5145 (0.0368)	γ	0.2591 (0.0126)	υ	2.2314 (0.1024)		

of conditional duration distribution alone are not enough. The obtained results indicate that the expected price duration is characterized by a relatively strong persistence. The posterior mean of β stands at about 0.9592. At the same time posterior distribution $p(\beta|y, y_{(0)})$ exhibits a very small dispersion, as evidenced by the standard deviation of 0.0036.

We notice that the posterior distribution of parameter δ_1 in the Box and Cox transformation is located to the right of 0 and to the left of 1. The posterior mean and standard deviation of δ_1 stand at ca. 0.2677 and 0.0304, respectively. Therefore, the location and dispersion of $p(\delta_1|y, y_{(0)})$ indicate that the data reject definitely the linear and logarithmic specifications. Next, the concavity of the shock impact curve seems to be a characteristic feature of price durations. The posterior distribution of parameter δ_2 in the Box and Cox transformation is located between 0 and 1, with the posterior mean and standard deviation equal to 0.5145 and 0.0368, correspondingly.

As far as the variance equation is concerned, we note that the posterior results for parameter β_G imply volatility persistence, though only a fairly weak one (the posterior mean of β_G stands at 0.2091). The asymmetry effect is slightly negative, but it seems to be statistically insignificant. The posterior mean of α_{1G} is equal to -0.0169 and is accompanied by a relatively large standard deviation of 0.0109. Looking at the posterior distribution of the parameter in question (see Figure 3), one can see that it contains 0 in the 95% HPD interval.

Finally, we analyse the estimation results for parameters η_1, η_2 and η_3, pertaining to the effects of duration on conditional variance. The posterior distributions of all three parameters η_1, η_2 and η_3 are well-separated from zero (see Figure 3), indicating statistical significance of all contemporaneous duration variables. The posterior mean of η_1 (related to the inverse of the duration, $\frac{1}{x_i}$) is slightly positive and equal to 0.0709, suggesting that a longer contemporaneous duration is associated with a lower volatility. It remains in agreement with predictions gathered from the theoretical microstructure model of Easley and O'Hara (1992), in which "no trade means no news", and short waiting times between transactions are associated with high volatility and high returns. Simultaneously, no trade (long price duration) is interpreted as no information so that volatility decreases. These empirical results obtained for the Polish data are in line with the ones presented by Engle (2000) for IBM Company, and Liu and Maheu (2012) for IBM and all Chinese stocks. Let us recall that the contemporaneous price duration divided by the expected price duration, i.e. $\frac{x_i}{\Psi_i}$, is interpreted as the surprise in durations. The posterior mean of η_2 (related to the duration surprise) is negative, amounting to about -0.4958. The posterior distribution of the parameter under consideration reveals a very small dispersion (as implied by the standard deviation of about 0.0094). It follows that when the actual price duration is longer than the expected one, the volatility decreases. Conceivably, according to this observation, it would appear that the uninformed traders on the Warsaw Stock Exchange are mostly risk averse. Such a result stands in contrast to the findings presented by Liu and Maheu (2012) for IBM and all Chinese stocks, but, on the other hand, is in line with the results obtained by these same Authors for Exxon Mobile Corporation and Pfizer companies, which are heavily traded stocks. The posterior mean of parameter η_3 (related to the inverse of the expected price duration $\frac{1}{\Psi_i}$) is positive and stands at about 0.2168. According to this, a higher expected transaction rate (i.e. shorter price durations) leads to higher volatility.

6 Concluding remarks

The main aim of this paper was to present a new specification of Engle's (2000) joint model to analyse the dynamics of intraday volatility and price durations and to develop and apply the Bayesian approach to estimate it. Our proposal is a novel, simple generalization of Engle's model. Specifically, we combine an EGARCH structure with duration variables for the volatility, and a Box-Cox ACD model with the generalized gamma distribution for the error term in the price duration equation. Our specification is treated within the Bayesian methodology, exploiting MCMC simulation methods to generate a pseudo-random sample from the posterior distribution.

The main finding arising from the empirical part of our research is that the proposed model provides an adequate description of the intraday volatility dynamics and price durations. However, a formal Bayesian comparison with other model specifications of the sort is required in order to examine the relative explanatory power of our proposal. Based on our results we can conclude that the Bayesian estimation approach provides a universal and convenient inference tool, especially since the properties of the maximum likelihood estimators for the UHF-GARCH-type and ACD models with conditional distributions other than the exponential one are still not well-known in the literature.

Moreover, the conditional normality of the returns is strongly rejected by the data. The volatility persistence is not strong and the asymmetry effect in volatility seems to be insignificant. In our empirical study the contemporaneous duration and expected duration terms emerge as relevant factors in modelling conditional variance. Longer price durations and longer expected price durations are associated with lower volatility, which is consistent with Easley and O'Hara's (1992) theoretical model. Furthermore, our research provided evidence that the generalized gamma distribution is much more appropriate in modelling price durations than the exponential one. The data also strongly reject the hypothesis of constant conditional price durations. It must be stressed that our analysis was limited to one stock only, so it would be of great interest to examine in further research more assets so as to better understand the differences across various stocks and market structures. Also, regarding other possible directions for future endeavours, an interesting strand of research would be Bayesian estimation of the ACD-GARCH-type models with market microstructure variables other than the ones considered in our study.

Acknowledgements

This research was realised within project number UMO-2011/01/N/HS4/03092 financed by the Polish National Science Centre based on decision number DEC-2011/01/N/HS4/03092. I gratefully acknowledge and thank the anonymous reviewers for careful assessment of my work, their valuable and meaningful remarks and comments as well as the detailed and constructive advice on improvement of this paper. Any remaining errors are the responsibility of the author.

References

[1] Allen D., Chan F., McAleer M., Peiris S., (2008), Finite sample properties of the QMLE for the Log-ACD model: application to Australian stocks, *Journal of Econometrics* 147, 163–185.

[2] Bauwens L., Giot P., (2000), The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks, *Annales d'Économie et de Statistique* 60, 117–149.

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

[4] Bauwens L., Giot P., (2003), Asymmetric ACD models: introducing price information in ACD models, *Empirical Economics* 28, 709–731.

[5] Bauwens L., Giot P., Grammig J., Veredas D., (2004), A comparison of financial duration models via density forecast, *International Journal of Forecasting* 20, 589–609.

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

[7] Bień K., (2004), Zastosowanie modeli Ultra-High-Frequency GARCH do analizy zmienności szeregów czasowych o bardzo dużej częstotliwości (UHF-GARCH Models: an Application to the Volatility Analysis of Ultra-High-Frequency Data, In Polish), *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, 1037, 38–47.

[8] Bień K., (2006), Zaawansowane specyfikacje modeli ACD – prezentacja oraz przykład zastosowania, *Przegląd Statystyczny* 53(1), 90–107.

[9] Bień K., (2006a), Model ACD – podstawowa specyfikacja i przykład zastosowania, *Przegląd Statystyczny* 53(3), 83–97.

[10] Bień-Barkowska K., (2011), Distribution Choice for the Asymmetric ACD Models, *Dynamic Econometric Models* 11, 55–72.

[11] Bień-Barkowska K., (2012), A Bivariate Copula-based Model for a Mixed Binary-Continuous Distribution: A Time Series Approach, *Central European Journal of Economic Modelling and Econometrics* 4, 117–142.

[12] Bień-Barkowska K., (2014), Explaining Liquidity Dynamics in the Order Driven FX Spot Market, *Przegląd Statystyczny* 61(3), 223–243.

[13] Bień-Barkowska K., (2014a), Capturing Order Book Dynamics in the Interbank EUR/PLN Spot Market, *Emerging Markets Finance and Trade* 50, 93–117.

[14] De Luca G.D., Gallo G., (2009), Time-varying mixing weights in mixture autoregressive conditional duration models, *Economic Review* 28, 101–120.

[15] De Luca G.D., Zuccolotto P., (2006), Regime-switching Pareto distributions for ACD models, *Computational Statistics and Data Analysis* 51, 2179–2191.

[16] Doman M., (2011), *Mikrostruktura giełd papierów wartościowych*, Poznań University of Economics Press, Poznań.

[17] Doman M., Doman R., (2010), Dependencies between price duration, volatility, volume and return on the Warsaw Stock Exchange, *Journal of Modern Accounting and Auditing* 6, 27–38.

[18] Dufour A., Engle R.F., (2000), The ACD model: predictability of the time between consecutive trades, *Discussion paper*, ISMA Centre, University of Reading.

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

[20] Engle R.F., (2000), The econometrics of ultra-high-frequency Data, *Econometrica* 68, 1–22.

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

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

[23] Ghysels E., Jasiak J., (1998), GARCH for irregularly spaced financial data: the ACD-GARCH model, *Studies in Nonlinear Dynamics and Econometrics* 2, 133–149.

[24] Grammig J., Maurer K., (2000), Non-monotonic hazard functions and the autoregrssive conditional duration model, *Econometrics Journal*, 3, 16–38.

[25] Hastings W.K., (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57, 97–109.

[26] Hautsch N., (2002), Modelling intraday trading activity using Box-Cox ACD models, *Discussion Paper* 02/05, CoFE, University of Konstanz.

[27] Hautsch N., (2004), *Modelling Irregularly Spaced Financial Data – Theory and Practice of Dynamic Duration Models*, Lecture Notes in Economics and Mathematical Systems 539, Springer, Berlin.

[28] Hautsch N., (2012), *Econometrics of Financial High-Frequency Data*, Springer-Verlag, Berlin Heidelberg.

[29] Huptas R., (2009), Intraday Seasonality in Analysis of UHF Financial Data: Models and Their Empirical Verification, *Dynamic Econometric Models* 9, 129–138.

[30] Huptas R., (2014), Bayesian Estimation and Prediction for ACD Models in the Analysis of Trade Durations from the Polish Stock Market, *Central European Journal of Economic Modelling and Econometrics* 6, 237–273.

[31] Liu C., Maheu J.M., (2012), Intraday Dynamics of Volatility and Duration: Evidence from Chinese Stocks, *Pacific-Basin Finance Journal* 20, 329–348.

[32] Lunde A., (1999), A generalized gamma autoregressive conditional duration model, *Discussion paper*, Alborg University.

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

[34] Meddahi N., Renault E., Werker B., (2006), GARCH and Irregularly Spaced Data, *Economics Letters* 90, 200–204.

[35] Osiewalski J., (2001), *Ekonometria bayesowska w zastosowaniach*, Cracow University of Economics Press, Kraków.

[36] Roll R., (1984), A Simple Implicit Measure of the Bid-Ask Spread in an Efficient Market, *Journal of Finance* 39, 1127–1139.

[37] Yu B., Mykland P., (1994), Looking at Markov samplers through CUMSUM paths plots: a simple diagnostic idea, *Technical Report* 413, Department of Statistics, University of Carolina at Berkeley.

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

[39] Zellner A., (1971), *An Introduction to Bayesian Inference in Econometrics*, John Wiley, New York.