3-2012

On the intraday periodicity duration adjustment of high-frequency data

WU Zhengxiao
Singapore Management University, zxwu@smu.edu.sg

Published in Journal of Empirical Finance, 2012 March, Volume 19, Issue 2, Pages 282-291
http://doi.org./10.1016/j.jempfin.2011.12.004

On the intraday periodicity duration adjustment of high-frequency data

Zhengxiao Wu *

Department of Statistics and Applied Probability, Risk Management Institute, National University of Singapore, 117546, Singapore

A R T I C L E I N F O

Article history:
Received 18 January 2011
Received in revised form 18 October 2011
Accepted 27 December 2011
Available online 3 January 2012

JEL classification:
C22
C41
C53

Keywords:
Autoregressive conditional duration model
High-frequency data
Intraday periodicity
Nonstationary Poisson process
Point process

A B S T R A C T

In the last decade, intensive studies on modeling high frequency financial data at the transaction level have been conducted. In the analysis of high-frequency duration data, it is often the first step to remove the intraday periodicity. Currently the most popular adjustment procedure is the cubic spline procedure proposed by Engle and Russell (1998). In this article, we first carry out a simulation study and show that the performance of the cubic spline procedure is not entirely satisfactory. Then we define periodicity point processes rigorously and prove a time change theorem. A new intraday periodic adjustment procedure is then proposed and its effectiveness is demonstrated in the simulation example. The new approach is easy to implement and well supported by the point process theory. It provides an attractive alternative to the cubic spline procedure.

1. Introduction

The recent development of computer, web and database technology facilitates both theoretical and empirical studies of high-frequency/tick-by-tick data. Since the pioneer work by Engle and Russell (1998), intensive research on modeling financial data at the transaction level has been conducted. The literature in this direction has evolved into a new area of financial research, often referred as “high-frequency finance”.

In analyzing tick-by-tick data, one challenge that a conventional econometrician would face is that transaction data inherently arrive in irregular time intervals and cannot be analyzed easily by standard fixed time interval methods in econometrics. Rather than aggregating the tick-by-tick data up to fixed time intervals so that the standard econometrics applies, Engle and Russell (1998) introduces the point process models to the econometrician community. The advantage of the point process models is that they can extract information from the irregular time intervals between transactions (which are called “durations” in Engle and Russell (1998)), hence knowledge and insight can be attained on the market activity. This is illustrated by Dufour and Engle (2000), Engle (2000, 2002) and Easley et al. (2008), among others.

The original model in Engle and Russell (1998) is called the autoregressive conditional duration (ACD) model. It can be regarded as the counterpart of the GARCH model for duration data. Since its introduction, the ACD model and its various extensions have become an important tool for modeling the high-frequency data (for alternative approaches, see for example, Zeng, 2003). Hautsch (2004) provides a thorough discussion and comparison of the ACD models. Pacurar (2008) gives a comprehensive survey on both theoretical developments in ACD modeling and empirical studies using financial data. And as noted in Pacurar (2008), the ACD literature is still rather young and not as rich. For example, the A-FIGARCH model (Baillie and Morana, 2009), which accounts both long memory and structural change in the volatility, is yet to be applied to the duration data.

To implement the ACD models, people commonly follow the two-step procedure in Engle and Russell (1998), in which the duration data are first adjusted to remove the periodicity; and then the parameters of the ACD models are estimated based on the adjusted durations. The most common periodicity in financial markets is the intraday periodicity or sometimes referred as the diurnality, which stems from market characteristics such as opening/closing of trading or lunch time for traders. Engle and Russell (1998) reports higher trading activity at the beginning and close of a trading day, and slower trading activity in the middle of the day based on an IBM stock transactions data set.

To produce the diurnally adjusted durations, Engle and Russell (1998) approximates the expected duration at each moment of a day by a cubic spline, whose nodes are set at each hour or half an hour. The diurnally adjusted durations are obtained by dividing the original durations by the corresponding value on the spline. This approach to removal of intraday periodicity is currently dominant in the empirical studies of high frequency finance (Bauwens et al., 2004; Pacurar, 2008). In this article, we refer it as the cubic spline approach.

In the literature however, there has not been any assessment of the effectiveness of the cubic spline procedure (Pacurar, 2008). An unsatisfactory periodicity adjustment procedure at the first step could lead to a bad fit of the ACD models at the second step even if the models were adequate.

Furthermore, to the best of our knowledge, point processes with periodicity have never been formulated explicitly in the literature. This is among the reasons why no assessment has been attempted for the cubic spine approach.

In this article, the performance of the cubic spline approach is evaluated by a simulation study. The result shows that it performs rather poorly. This motivates us to propose a new periodicity adjustment method for duration data. Our method is based on an explicit formulation of the periodic point processes. We call the new approach the time change procedure. Compared with the cubic spline procedure, the time change approach eliminates the needs of setting nodes. In addition, in the cubic spline approach, one also needs to decide which time point best represents the duration, so that the duration is adjusted by the spline function valued at that point. The time change approach bypasses this difficulty. We demonstrate that the time change procedure is easy to implement and it gives a better removal of the intraday periodicity in the simulation study. The data and the code used in this article are available from the author on request.

The rest of the paper is organized as follows, Section 2 describes and assesses the cubic spline approach; Section 3 defines the periodic point processes and develops the theoretical foundation of the time change procedure; Section 4 demonstrates the new procedure on the same data used in Section 2 and compares its performance with the cubic spline procedure; Section 5 summarizes the conclusion and discusses future work; Section 6 collects the proofs of the theorems; the Appendix presents an algorithm to generate transaction times following a nonstationary Poisson process.

2. Effectiveness of periodicity adjustment procedures

In this section we propose a formal method to assess the effectiveness of periodicity adjustment procedures. The idea is to simulate transaction times that follow a known nonstationary Poisson process and then carry out the periodicity adjustment. We expect the adjusted durations to be approximately independent for an effective adjustment method.

2.1. Data

We illustrate with an IBM stock transaction data set in year 2005. The data are from the Trade and Quote (TAQ) database supplied by New York Stock Exchange (NYSE). Following the practice in Engle and Russell (1998), three days are deleted from the 252 trading days in this one year sample. A halt in IBM trading occurred on June 1st, 2005 and there were scheduled partial closures on November 11th and November 25th, 2005. After deleting these three days there are a total of 2,118,217 transactions executed in 249 trading days. The time stamps of the transactions are recorded in seconds. As IBM is traded from 9:30 EST to 16:00 EST during each trading day, the daily time stamps (in seconds) take the following values: 0, 1, 2,…, 23,399, 23,400. For example, a trade occurs after 9:30:00 and before 9:30:01 will have a daily time stamp 1. Fig. 1 depicts the histogram of the daily transactions times. The "U" shape confirms the time-of-day effect reported in Engle and Russell (1998). The peculiarly frequent transactions around the open are less obvious in Fig. 1, but there is 23 s in the first minute having trading counts above 500. These extreme high frequency transactions are caused by the opening auction (see Engle and Russell, 1998 for details).

Now we simulate transaction times for 60 days with a nonstationary Poisson process, so that the empirical distribution of the simulated transaction times resembles the histogram plotted in Fig. 1. Specifically, we set the intensity function $\lambda(t)$ to be

$$\lambda(t) = c_i/249, \text{ for } i-1 < t \leq i \text{ and } i = 1, 2, \ldots, 23400, \tag{1}$$

where c_i is the number of trades that have daily transaction time i in the 249 trading days. Similar to the real data, a simulated event has time stamp i for an arrival time $t \in (i-1, i]$ and hence the simulated time stamps are also whole numbers. Denote the set of time stamps by $\{t_j\}$, where the index j starts from 0. We refer the reader to the Appendix for a simple and efficient algorithm derived for generating $\{t_j\}$.

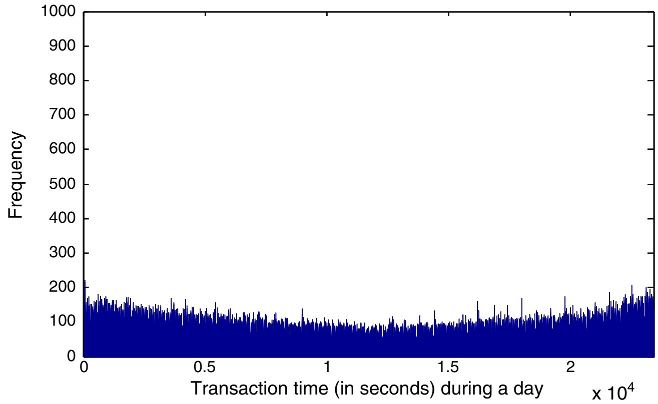

Fig. 1. Histogram of IBM stock transaction times in 249 trading days. Time 0 is to be interpreted as 9:30:00 EST and time 23,400 is 16:00:00 EST.

2.2. Cubic spline procedure

We apply the cubic spline procedure on the simulated data. Let $x_j = t_j - t_{j-1}$ be the interval between two arrival times, which is also called the raw duration. Thus the index j in the sequence $\{x_j\}$ starts from 1. The diurnally adjusted durations are given by

$$\tilde{x}_j = x_j/\varphi\left(t_{j-1}\right) \tag{2}$$

where $\varphi(t)$ is a cubic spline function. Engle and Russell (1998) suggest to set nodes at each hour from 10:00 to 16:00, with an additional node at 15:30 because activity drops off quickly at the end of the day and the additional node can provide more flexibility. The values of $\varphi(t)$ at the nodes are decided by the average duration over the 10 min prior to the node times. For instance, the value of $\varphi(t)$ at 10:00 is the average duration of trades occurred between 9:50 to 10:00. Fig. 2 plots the function $\varphi(t)$ for the simulated data. It shows shorter durations on average at the beginning and close of the trading day, and longer durations in the middle of the day. This pattern is expected, since shorter durations indicate higher trading activity, and longer durations are signs of slower trading activity.

The diurnally adjusted durations $\{\tilde{x}_j\}$ are calculated by Eq. (2). This series would have approximately zero intertemporal correlations if the cubic spline procedure has removed the periodicity successfully. We employ a Ljung–Box test to formally test the null hypothesis that the first 15 autocorrelations are 0. Under the null hypothesis, the test statistics follows asymptotically a Chi-squared distribution with 15 degrees of freedom. The test statistics produced for this series is 383.1946 with a p-value 0. Hence the null hypothesis is rejected and it implies that the adjusted durations cannot be uncorrelated.

Usually zero durations are removed when one applies ACD models to tick-by-tick data (Engle and Russell, 1998; Pacurar, 2008). Engle and Russell (1998) also states that "A model of the trading process would be contaminated by including opening trades, hence transactions occurring in the first 20 min of the day (9:30–9:50) were deleted". We clean the data by removing the zero durations and transactions between 9:30 and 9:50, and perform another Ljung–Box test on the first 15 autocorrelations. The test statistics is 59.2777 with a p-value 3.355e-07. The test statistics is less extreme, but there is still strong evidence that that the cubic spline procedure does not effectively remove the intraday periodicity.

In fact, it is not surprising that the cubic spline procedure does not perform well. There are typically hundreds of transactions each hour. A cubic spline approximation with hourly nodes is obviously too crude. Nonparametric kernel methods or a cubic spline with finer node intervals are expected to have better performances (see Veredas et al., 2007, and references therein), but they are computationally more expensive. Furthermore, a better estimate $\varphi(t)$ does not solve all the problems. In Eq. (2), the raw duration x_j is adjusted by $\varphi(t_{j-1})$, where t_{j-1} is the starting time of the duration x_j. Imagine the scenario that in two different trading days, two duration intervals have the same starting time but with different lengths. Then the duration-based adjustment procedure would modify these two intervals using a same seasonal factor. This is not a good practice: consider the extreme case when one interval is 1 s long and the other one is 1 hour long, although they share the same starting time, the diurnality definitely should have different impact on them. In addition, if $\varphi(t_{j-1})$ is an acceptable adjustment factor, it seems equally plausible to use t_j, the ending time of the duration, to represent the duration. In other words, one can alternatively use $\varphi(t_j)$ as the adjustment factor and let $\tilde{x}_j = x_j/\varphi(t_j)$. Or presumably better, one can consider $\tilde{x}_j = x_j/\varphi\left(\frac{t_{j-1}+t_j}{2}\right)$. The choice of the time point representing the duration interval seems rather subjective. It could have a non-negligible impact in the data analysis and modeling, especially when the trading is less active and the durations are long. This difficulty is inherent in all the duration based adjustment procedures. However, it will be shown in Section 4 that the time change approach circumvents this difficulty.

3. Periodic point process

3.1. Notation and definitions

To examine the periodicity adjustment procedures more rigorously, it is indispensable to work on an explicit formulation of the periodic point processes. Point processes have been a well studied subject in the statistics literature (see, e.g., Daley and Vere-Jones, 2004; Karr, 1991). In high frequency finance applications, it is the marked point process with a conditional intensity that catches the most attention of the econometricians, because of the tractability in its modeling and in its statistical inferences.

Let (Ω, F, P) be a probability space. Let $N = \sum_{i=1}^{\infty} \delta_{\{T_i\}}$ denote a simple point process on $\mathbf{R}_+$ adapted to a filtration $\{H_t\}$, where $H_t \subseteq F$, $\{T_i\}$ is the increasing sequence of the event arrival times (transaction times, for instance) and $\delta_{\{T_i\}}$ is the Dirac measure concentrated on the point T_i. Then $N_t = \sum_{i=1}^{\infty} 1_{\{T_i \le t\}}$ defines the counting process associated with N. We assume $E(N_t) < \infty$ for each t to avoid the concept of local martingales in the definition of the conditional intensity.

Definition 3.1. *A positive, H_t-predictable process λ_t is the (H_t)-conditional intensity of N (or N_t), if*

$$N_t - \int_0^t \lambda_u du$$

is a (H_t)-martingale.

In financial markets, tick-by-tick data include not only the transaction time, but also other variables such as price, volume, etc. These are modeled by the marks in a marked point process. The definition of the conditional intensity of a marked point process is more involved. Suppose that $M = \sum_{i=1}^{\infty} \delta_{\{T_i, Z_i\}}$ is a marked point process adapted to a filtration $\{H_t\}$, with mark space (E, E). In practice, usually $\{H_t\}$ is the completion of $F_t^M = \sigma(M_s(B) : 0 \le s \le t, B \in E)$, which records the internal history of M; E is often a subset of a Euclidean space and E is its Borel σ-field.

Definition 3.2. (Karr, 1991) *A stochastic process $(\gamma_t(B) : t \ge 0, B \in E)$ is the (H_t)-conditional intensity of M provided that*

1. *For each t, $B \rightarrow \gamma_t(B)$ is a random measure on E;*
2. *For each B, the process $\gamma_t(B)$ is the (H_t)-conditional intensity of the counting process $M_t(B) = \sum_{i=1}^{\infty} 1_{\{T_i \le t, Z_i \in B\}}$.*

We propose to define the periodic point processes by modeling their conditional intensities.

Definition 3.3. *A simple point process $N = \sum_{i=1}^{\infty} \delta_{\{T_i\}}$ with conditional intensity λ_t is periodic with period p if there exists a positive, deterministic, periodic function $s(t)$ with period p such that $\lambda_t^* = \lambda_t / s(t)$ is first-order stationary, i.e. $E(\lambda_t^*)$ exists and is finite and does not depend on t.*

Definition 3.4. *A marked point process $M = \sum_{i=1}^{\infty} \delta_{\{T_i, Zi\}}$ with conditional intensity $(\gamma_t(B) : t \ge 0, B \in E)$ is periodic with period p if there exists a positive, deterministic, periodic function $s(t)$ with period p such that $\boldsymbol{\gamma_t^*(B) = \gamma_t(B)/s(t)}$ **is first-order stationary for all** $\boldsymbol{B \in E}$.*

In the above definitions the scale of $s(t)$ can be selected arbitrarily. A specific choice of scale will be made in Section 3.2. Hereafter attention will be focused on the simple point process. The idea extends to the marked point process parallelly.

3.2. Time change theorem

Let $N = \sum_{i=1}^{\infty} \delta_{\{T_i\}}$ be a periodic simple point process with period p on $\mathbf{R}_+$, with H_t-conditional intensity $\lambda_t = s(t)\lambda_t^*$. Assume $s(\cdot)$ and λ^* are bounded hereafter for simplicity.

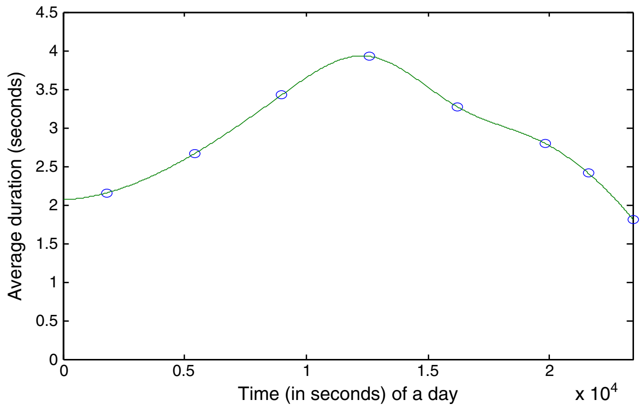

Fig. 2. Cubic spline estimate of daily pattern for transaction durations for the simulated data. Time 0 is to be interpreted as 9:30:00 EST and time 23,400 is 16:00:00 EST.

As λ_t^* is first order stationary, it follows that

$$E(N_{t+\Delta t}-N_t) = E\left(\int_t^{t+\Delta t} \lambda_u du\right) = E(\lambda_u^*)\int_t^{t+\Delta t} s(u)du.$$

The time change function *a(t)* is defined by

$$a(t) = \int_0^t s(u)du = E(N_t)/E(\lambda_u^*).$$

Since $s(\cdot)$ is positive and bounded, it follows that $a(\cdot)$ is continuous and strictly increasing. We define the time inverse of $a(\cdot)$ by

$$a^{-1}(t) = inf\{s : a(s) > t\}, t \geq 0.$$

It is easy to see that $a^{-1}(\cdot)$ is also continuous, strictly increasing and $a(a^{-1}(t)) = a^{-1}(a(t)) = t$ for $t \geq 0$. Our periodicity adjustment procedure is based on the following time change theorem.

Theorem 3.5. *Define*

$$D_t(\omega) = N_{a^{-1}(t)}(\omega), \text{ for all } t \geq 0 \text{ and } \omega \in \Omega,$$

and $G_t = H_{a^{-1}(t)}$, *then* D_t *is a simple point process with* G_t*-conditional intensity* $\lambda^*_{a^{-1}(t)}$.

Note that $E(D_{t+s} - D_t)$ does not depend on *t*. In fact, the theorem states that if we have constructed a clock which reads *a(t)* when the actual time is *t*, then according to this clock the point process has a first-order stationary conditional intensity. Hence the expected number of the events in every unit time is constant. In other words, the seasonal factor *s(t)* is removed if we time the events by the constructed clock.

We suggest to scale *s(t)* so that the clock we constructed reads *p* when the actual time is *p* (recall that *p* is the period of *s(t)*). More precisely, scale *s(t)* such that

$$p = a(p) = \int_0^p s(u)du. \tag{3}$$

In this way, the clock is as fast as a normal clock in large, which can be seen from $a(p) = a^{-1}(p) = p$, $a(2p) = a^{-1}(2p) = 2p, \ldots$. With such a *s(t)*, we denote $l = E(\lambda^*)$. Hence one expects to observe *l* events every unit time according to the clock.

3.3. Estimate

In financial applications, $a(\cdot)$ is always unknown. It is necessary to estimate $a(\cdot)$ to apply the time change theorem.

Suppose the periodic simple point process *N* is observed up to time *np*. We write

$$N_t^i = N_{(i-1)p+t} - N_{(i-1)p}, \text{ for } i = 1, 2, \ldots, n.$$

An estimate of $a(\cdot)$ is given by

$$\hat{a}_n(t) = \frac{p\sum_{i=1}^n N_t^i}{N_{np}}, \text{ for } 0 \leq t \leq p. \tag{4}$$

From the definition, $\hat{a}_n(0) = 0$ and $\hat{a}_n(p) = p$, hence $\hat{a}_n$ has the right scale as requested in Eq. (3). Now we investigate the asymptotic property of $\hat{a}_n(t)$. Note that N_t in general is not stationary, however, it is plausible to assume that $\{N^i\}$ are stationary and ergodic when they are considered as a sequence of counting-measure (on $[0,p)$)-valued random variables. The following theorem implies the strong consistency of $\hat{a}_n$.

Theorem 3.6. *If the counting-measure-valued random variables* $\{N^i\}$ *are stationary and ergodic, then*

$$\sup_{t\in[0,p]} |\hat{a}_n(t) - a(t)| \to 0 \text{ a.s.}$$

as $n \to \infty$.

One can also establish the asymptotic normality for $\hat{a}_n$ under certain mixing conditions. Bradley (2007a, 2007b, 2007c) provides a comprehensive and up-to-date survey of various mixing conditions. Let A and B be sub-σ-fields of F. Define

$$\alpha(A,B) = sup\{|P(A\cap B) - P(A)P(B)| : A\in A, B\in B\}.$$

Extend the one-sided stationary sequence $\{N^i, i=1,2,\ldots\}$ to the two-sided stationary sequence $\{N^i : i\in \mathbf{Z}\}$ and let $F_a^b = \sigma(N^i : a\le i\le b)$. Our theorem is based on a result by Ibragimov (1962). See Ibragimov and Linnik (1971) or Hall and Heyde (1980) for a proof.

Lemma 3.7. *(Ibragimov, 1962) Suppose $\{X_i, i\in\mathbf{Z}\}$ is a stationary sequence with $EX_0=0$, $E|X_0|^{2+c}<\infty$ for some $c>0$. Let $\alpha_X(n) = \alpha(\sigma(X_i : i\le -n), \sigma(X_i : i\ge 0))$ and suppose*

$$\sum_{n=1}^{\infty} \alpha_X(n)^{c/(2+c)} < \infty.$$

If $S_n = X_1 + \ldots + X_n$, then $S_n/\sqrt{n} \xrightarrow{d} N(0,\sigma^2)$, where $\sigma^2 = EX_0^2 + 2\sum_{i=1}^{\infty} EX_0X_i < \infty$.

Theorem 3.8. *If the counting-measure-valued random variables $\{N^i\}$ are stationary. Let $\alpha(n) = \alpha(F_{-\infty}^{-n}, F_0^{\infty})$ and suppose*

$$\sum_{n=1}^{\infty} \alpha(n)^{1-\delta} < \infty, \text{ for some } \delta > 0. \tag{5}$$

Then

$$\sqrt{n}(\hat{a}(t) - a(t)) \xrightarrow{d} N\left(0, \sigma_t^2\right),$$

and

$$\sigma_t^2 = \frac{p^2\sigma_{11}(t) + a^2(t)\sigma_{22}(t) - 2pa(t)\sigma_{12}(t)}{p^2 l^2}, \tag{6}$$

where

$$\sigma_{11}(t) = Var\left(N_t^0\right) + 2\sum_{i=1}^{\infty} Cov\left(N_t^0, N_t^i\right),$$
$$\sigma_{12}(t) = \sigma_{21}(t) = Cov\left(N_t^0, N_p^0\right) + \sum_{i=1}^{\infty}\left(Cov\left(N_t^0, N_p^i\right) + Cov\left(N_p^0, N_t^i\right)\right)$$

and

$$\sigma_{22}(t) = Var\left(N_p^0\right) + 2\sum_{i=1}^{\infty} Cov\left(N_p^0, N_p^i\right).$$

Condition (5) is satisfied if $\{N^i\}$ are M-dependent, that is, if $\{N_i, i\le 0\}$ and $\{N_i, i\ge M\}$ are independent. Condition (5) can be further relaxed by using a result in Doukhan et al. (1994) in place of Lemma (7). Central limit theorems with other type of mixing conditions such as ρ-mixing, φ-mixing can also be established. We will not enter into the details of these proofs since they do not give significant improvements for applications.

4. Time change procedure

Theorem 3.5 provides the theoretic basis for the following two-step intraday periodicity duration adjustment procedure:

1. Estimate of the time change function $\hat{a}(t)$ by using (4).
2. The diurnally adjusted durations are given by

$$\tilde{x}_j = \hat{a}\left(t_j\right) - \hat{a}\left(t_{j-1}\right). \tag{7}$$

Unlike the cubic spline procedure, the time change procedure (7) makes the adjustments at the transaction times. All the information of the data is preserved during the adjustment in the sense that one can recover the original data (i.e. the transaction times) by reversing the time change. In contrast, the cubic spline procedure (2) loses information because the starting time t_{j-1} does not fully characterize the duration.

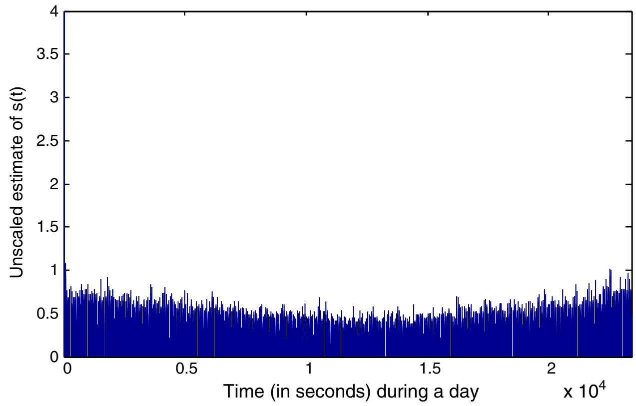

Fig. 3. Unscaled estimate of *s(t)* from the simulated transaction data. Time 0 is to be interpreted as 9:30:00 EST and time 23,400 is 16:00:00 EST.

We illustrate the time change procedure with the simulated 60 days transaction data described in Section 2. In this data set, the seasonal component $s(t)$ is proportional to the intensity function (1) of the nonstationary Poisson process and λ^* is a constant. The period p is 23,400 (seconds). The adjusted transaction times $\{a(t_j)\}$ are approximately the event arrival times of a stationary Poisson point process. Hence

$$\left\{a\left(t_j\right)-a\left(t_{j-1}\right)\right\} \tag{8}$$

are approximately i.i.d. exponential random variables. Note that $\{a(t_j)-a(t_{j-1})\}$ would be exactly i.i.d. exponential random variables, if $\{t_j\}$ were not rounded (to integers).

Before estimating $a(\cdot)$, one can estimate $s(t)$ roughly by

$$\hat{s}(t) \propto \frac{f_i}{60}, \text{ for } i-1<t\leq i \text{ and } i=1,2,\ldots,23400,$$

where f_i is the number of trades that have daily transaction time i in the simulated 60 days. We plot the unscaled $\hat{s}(t)$ in Fig. 3. It has a shape similar to the underlying intensity function $\lambda(t)$.

Let

$$\hat{a}(t) = \int_0^t \hat{s}(u)du, \tag{9}$$

and we rescale $\hat{a}(\cdot)$ so that $\hat{a}(23400) = 23400$. It is easy to see that $\hat{a}(t)$ is equal to $\hat{a}_n(t)$ given by (4) when $t=0,1,2,\ldots,23400$. Fig. 4 displays the function $\hat{a}(t)$. As expected, the curve is steeper near the open and close of the market. It is flatter in the middle because of the lower trading activity around lunch time. In other words, the clock we constructed runs faster near the open and close of the market, and it runs slower in the middle of a day.

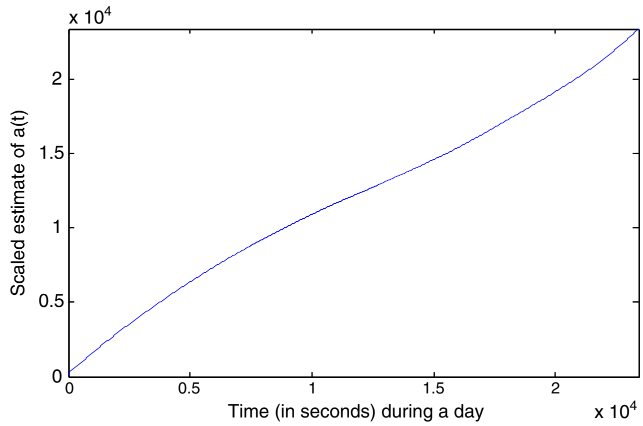

Fig. 4. Scaled estimate of *a(t)* from the simulated transaction data. Time 0 is to be interpreted as 9:30:00 EST and time 23,400 is 16:00:00 EST.

Table 1
Ljung–Box test statistics for uncleaned data and cleaned data (recall that the cleaned data are obtained by removing the zero durations and transactions occur in the first 20 min).

χ^2_{15} (p-value)	Cubic spline approach	Time change by $\hat{a}(\cdot)$	Time change by $a(\cdot)$
Uncleaned data	383.19 (0)	259.17 (0)	186.07 (0)
Cleaned data	59.28 (3.355e-07)	18.83 (0.22)	19.20 (0.20)

One can apply Theorem 3.8 to construct asymptotic confidence intervals. It is not hard to show that in this setting $\sigma_t^2 = \frac{a(t)(1-a(t)/p)}{l}$. Therefore,

$$\hat{a}(t) - z_{\alpha/2}\sqrt{\frac{\hat{a}(t)(1-\hat{a}(t)/p)}{n\hat{l}}} < a(t) < \hat{a}(t) + z_{\alpha/2}\sqrt{\frac{\hat{a}(t)(1-\hat{a}(t)/p)}{n\hat{l}}} \tag{10}$$

gives an asymptotic $100(1-\alpha)\%$ confidence interval for $a(t)$, where $\hat{l} = N_{np}/(np)$ is an estimate of l and $z_{\alpha/2}$ is the $1-\alpha/2$ percentile of the standard normal distribution. In the simulated example, $a(t)$ is known hence we can calculate the estimation error $\hat{a}(t) - a(t)$. Fig. 5 depicts $\hat{a}(t) - a(t)$ and the 95% confidence interval computed from (10). It can be seen that as expected the estimation error stays in the confidence interval most of the time.

Next, we calculate two sets of diurnally adjusted durations $\{\tilde{x}_j\}$ by Eqs. (7) and by (8) respectively. Again we are able to apply (8) because the underlying $a(t)$ is known in this simulation study. In real applications, only (7) is available to us. The same Ljung–Box tests as in Section 2 are carried out on both the uncleaned data and cleaned data, to test the null hypothesis that the first 15 autocorrelations are 0. Table 1 summarizes the results. For both uncleaned and cleaned data, the time change approach leads to much smaller Ljung–Box test statistics than the cubic spline procedure. In particular, the Ljung–Box test does not reject thenull hypothesis at a 5% significance level when we adjust the cleaned data by the time change procedure. For the uncleaned data, however, even if we apply the time change by $a(t)$ which won't be available in practice, we are not able to eliminate the autocorrelation and cannot accept the null hypothesis. We believe that this is caused by the rounding of the transaction times. In fact, if the frequency of the transactions is so high that there are typically several trades within each second, then zero durations will cluster, which introduces strong autocorrelations. This is the underlying reason why Engle and Russell (1998) suggest to delete these "contaminated" data.

5. Discussion

Various ACD models have been proposed to model the diurnally adjusted durations. However, their diagnostic tests often indicate that the models do not fit the data well. As reported in Pacurar (2008): "Interestingly, several authors (Engle, 2000; Engle and Russell, 1998; Fernandes and Grammig, 2006; Zhang et al., 2001, among others) reveal substantial difficulties in completely removing dependence in the residual series, which suggests that the question of the most appropriate model for trade durations is far from being answered." The lack of fitness of their models can be caused (partially) by the ineffective periodicity adjustment procedures that they use. After all, cubic spline procedure is only an ad hoc method. Besides the location of the nodes, one also needs to decide rather arbitrarily which time point (the starting time, the ending time, or the mid time point) best represents the duration so that the duration is adjusted by the value of the spline function at that specific time point. On the other hand, the time change approach is well supported by point process theory and its performance is superior

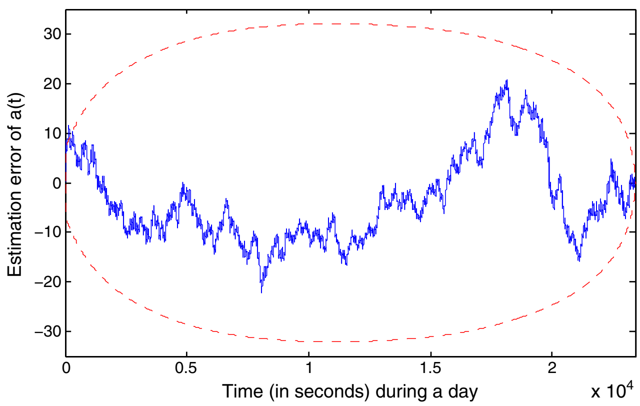

Fig. 5. The plot of $\hat{a}(t) - a(t)$ for the simulated transaction data. Time 0 is to be interpreted as 9:30:00 EST and time 23,400 is 16:00:00 EST. Dotted lines give the asymptotic 95% confidence interval computed from (10).

to the cubic spline procedure in our simulation study. The implementation is easy. Hence, we recommend the time change procedure toall the practitioners of ACD models.

In an unpublished working paper, Russell (1999) proposes an autoregressive conditional intensity (ACI) model. He estimates the seasonal intensity by linear spline with hourly nodes. Bauwens and Hautsch (2006) extends Russell's work to the so-called stochastic conditional intensity (SCI) model, and they estimate the periodicity from the durations also by the spline method. Similar spline procedure is found in Bowsher (2007). We believe that the time change approach is an attractive alternative to the spline procedure in these intensity-based models as well.

Andersen and Bollerslev (1997, 1998) investigate the intraday periodicity in the return volatility in foreign exchange and equity markets, using aggregated five minutes return data. As noted in Engle (2000), such aggregates can lead to loss of information. It would be interesting to study the problem based on the original transaction level data. The point process and the time change procedure would play a role here. Future work is to be done in this direction.

6. Proofs

Proof of Theorem 3.5. Fix $t,s \geq 0$, then it suffices to show that

$$E(D_{t+s}|G_t) = D_t + E\left(\int_t^{t+s} \lambda^*_{a^{-1}(u)} du \middle| G_t\right).$$

Since N_t is a periodic simple point process with H_t-conditional intensity λ_t, we have

$$\begin{aligned} E(D_{t+s}|G_t) &= E\left(N_{a^{-1}(t+s)}\middle|H_{a^{-1}(t)}\right) = N_{a^{-1}(t)} + E\left(\int_{a^{-1}(t)}^{a^{-1}(t+s)} \lambda_u du \middle| H_{a^{-1}(t)}\right) = D_t + E\left(\int_t^{t+s} \lambda_{a^{-1}(v)} d\left(a^{-1}(v)\right)\middle|H_{a^{-1}(t)}\right) \\ &= D_t + E\left(\int_t^{t+s} \frac{\lambda_{a^{-1}(v)}}{s(a^{-1}(v))} dv \middle| H_{a^{-1}(t)}\right) = D_t + E\left(\int_t^{t+s} \lambda^*_{a^{-1}(v)} dv \middle| G_t\right). \end{aligned}$$

We write $v = a(u)$ in the third equality. The fact that $\frac{d}{dv} a^{-1}(v) = \frac{1}{s(a^{-1}(v))}$ is used to get the fourth equality. □

Proof of Theorem 3.6. Fix t, then $\{N_t^i\}$ are stationary and ergodic (real valued) random variables with $E(N_t^i) = la(t)$. The Birkhoff's ergodic theorem implies $n^{-1}\sum_{i=1}^n N_t^i \to la(t)$ a.s. When $t = p$, $n^{-1}\sum_{i=1}^n N_p^i = \frac{N_{np}}{n} \to la(p) = lp$. Therefore, $\hat{a}(t) = \frac{p\sum_{i=1}^n N_t^i}{N_{np}} = \frac{p\left(\sum_{i=1}^n N_t^i\right)/n}{N_{np}/n} \to a(t)$ a.s. Since $\hat{a}_n$ is a sequence of increasing functions that converges pointwise to $a(t)$ which is bounded and continuous, it follows that $sup_{t\in[0,p]}|\hat{a}_n(t) - a(t)| \to 0$ a.s. (see, for example, Durrett, 2005, page 59).

Proof of Theorem 3.8. Fix t and note that N_t^i is stochastically bounded by a Poisson random variable with parameter $psup_t \{\lambda_t^* s(t)\} < \infty$ (recall that we assume that λ^* and $s(\cdot)$ are both bounded). Hence $E|N_t^0|^r < y$, for any $r > 0$.

We first show that

$$\sqrt{n}\left[\begin{pmatrix} \frac{\sum_{i=1}^n N_t^i}{n} \\ \frac{\sum_{i=1}^n N_p^i}{n} \end{pmatrix} - \begin{pmatrix} la(t) \\ lp \end{pmatrix}\right] \xrightarrow{d} N\left(0, \begin{pmatrix} \sigma_{11}(t) & \sigma_{12}(t) \\ \sigma_{21}(t) & \sigma_{22}(t) \end{pmatrix}\right)$$

by the Cramér–Wold device. Let $u, v \in \mathbf{R}$, $X_i = u(N_t^i - la(t)) + v(N_p^i - lp)$ and $S_n = X_1 + \ldots + X_n$, it suffices to show that $S_n/\sqrt{n} \xrightarrow{d} N(0, u^2\sigma_{11}(t) + v^2\sigma_{22}(t) + 2uv\sigma_{12}(t)$. Clearly $\{X_i\}$ are stationary with mean 0 and $E|X_0|^r < \infty$, for any $r > 0$. Furthermore, since $\sigma(X_i) \subset \sigma(N^i)$, it follows $\alpha_X(n) \leq \alpha(n)$. Hence $\sum_{n=1}^\infty \alpha_X(n)^{1-\delta} < \infty$. Let $c = \frac{2}{\delta} - 2$, Lemma 7 implies that $S_n/\sqrt{n} \xrightarrow{d} N(0, u^2\sigma_{11}(t) + v^2\sigma_{22}(t) + 2uv\sigma_{12}(t)$.

Next we employ the delta-method with function $f(x,y) = x/y$. It follows that

$$\sqrt{n}\left(\frac{\sum_{i=1}^n N_t^i}{N_{np}} - \frac{a(t)}{p}\right) \xrightarrow{d} N\left(0, \frac{p^2\sigma_{11}(t) + a^2(t)\sigma_{22}(t) - 2pa(t)\sigma_{12}(t)}{p^4 l^2}\right),$$

i.e.

$$\sqrt{n}(\hat{a}_n(t) - a(t)) \xrightarrow{d} N\left(0, \frac{p^2\sigma_{11}(t) + a^2(t)\sigma_{22}(t) - 2pa(t)\sigma_{12}(t)}{p^2 l^2}\right).$$ □

Acknowledgment

The author is grateful to Richard T. Bailie, the editor, an anonymous referee, Ying Chen, Yong Zeng, Sergey Utev and Jialiang Li for their helpful comments. This work is jointly supported by MOE grants R-155-000-070-133 and R-155-000-105-112.

Appendix. Simulate transaction times with periodicity

Suppose one observes the transaction times for n trading days. Let $s_{(1)} \le s_{(2)} \le \ldots \le s_{(m)}$ be the order statistics of the superposition of the n days' daily time stamps. They are integer numbers taking values in $\{0, 1, 2, \ldots, p\}$. Hence an empirical intensity function is given by

$$\lambda(t) = c_i/n, \text{ for } i-1 < t \le i \text{ and } i = 1, 2, \ldots, p, \tag{11}$$

where c_i is the number of trades that have daily transaction time i in these n days. One can see that $n = 249, m = 2118217$, $p = 23400$ in Section 2.1.

Now we would like to simulate transactions which follow a nonstationary Poisson process with intensity (11). The simulated transaction has time stamp i if its arrival time t satisfies $i-1 < t \le i$. We suggest the following algorithm:

1. set $j = 0$ and $sum = 0$,
2. while $sum < m$, do
   (a) generate $E_j \sim \mathrm{Exp}(n)$, where *Exp(n)* denotes an exponential distribution with mean n,
   (b) $sum = sum + E_j$,
   (c) $o_j = \lceil sum \rceil$, i.e., round up sum to an integer o_j.
   (d) $j = j + 1$.

Then $\{s_{o_j}\}$ give us the simulated time stamps. The validity of the algorithm is a consequence of Theorem 3.5.

References

Andersen, T., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. J. Empir. Finance 4, 115–158.
Andersen, T., Bollerslev, T., 1998. Deutsche Mark-dollar volatility: intraday activity patterns, macroeconomic announcements and longer run dependencies. J. Finance 53, 219–265.
Baillie, R., Morana, C., 2009. Modelling long memory and structural breaks in conditional variances: an adaptive FIGARCH approach. J. Econ. Dyn. Control. 33, 1577–1592.
Bauwens, L., Hautsch, N., 2006. Stochastic conditional intensity processes. J. Financ. Econ. 4, 450–493.
Bauwens, L., Giot, P., Grammig, J., Veredas, D., 2004. A comparison of financial duration models via density forecasts. Int. J. Forecast. 20, 589–609.
Bowsher, C.G., 2007. Modelling security market events in continuous time: intensity based, multivariate point process models. J. Econ. 141, 876–912.
Bradley, R.C., 2007a. Introduction to Strong Mixing Conditions, Volume 1. Kendrick Press, Heber City, Utah.
Bradley, R.C., 2007b. Introduction to Strong Mixing Conditions, Volume 2. Kendrick Press, Heber City, Utah.
Bradley, R.C., 2007c. Introduction to Strong Mixing Conditions, Volume 3. Kendrick Press, Heber City, Utah.
Daley, D.J., Vere-Jones, D., 2004. An Introduction to the Theory of Point Processes, 2nd ed. Springer, New York.
Doukhan, P., Massart, P., Rio, E., 1994. The functional central limit theorem for strongly mixing processes. Ann. Inst. H. Poincaré Probab. Statist. 30, 63–82.
Dufour, A., Engle, R.F., 2000. Time and the impact of a trade. J. Finance 55, 2467–2498.
Durrett, R., 2005. Probability: Theory and Examples, 3rd ed. Thomson, Brooks/Cole, Belmont, CA.
Easley, D., Engle, R.F., O'Hara, M., Wu, L., 2008. Time-varying arrival rates of informed and uninformed trades. J. Financ. Econ. 171–207.
Engle, R.F., 2000. The econometrics of ultra-high frequency data. Econometrica 68, 1–22.
Engle, R.F., 2002. New frontiers for ARCH models. J. Appl. Econ. 17, 425–446.
Engle, R.F., Russell, J.R., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 66, 1127–1162.
Fernandes, M., Grammig, J., 2006. A family of autoregressive conditional duration models. J. Econ. 130, 1–23.
Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Application. Academic Press, New York.
Hautsch, N., 2004. Modelling Irregularly Spaced Financial Data. Springer, Berlin.
Ibragimov, I.A., 1962. Some limit theorems for stationary processes. Theory Probab. Appl. 7, 349–382 (in Russian).
Ibragimov, I.A., Linnik, Y.V., 1971. Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff, Groningen.
Karr, A.F., 1991. Point Processes and Their Statistical Inference. Dekker, New York.
Pacurar, M., 2008. Autoregressive conditional durations in finance: a survey the theoretical and empirical literature. J. Econ. Surv. 22, 711–751.
Russell, J.R., 1999. Econometric Modeling of Multivariate Irregularly-Spaced High-Frequency Data. Working paper, University of Chicago.
Veredas, D., Rodríguez-Poo, J.M., Espasa, A., 2007. On the (Intradaily) seasonal and dynamics of a financial point process: a semiparametric approach. In: Bauwens, L., Polhmeier, W., Veredas, D. (Eds.), High Frequency Financial Econometrics: Recent Developments. Springer Verlag.
Zeng, Y., 2003. A partially-observed model for micro-movement of asset prices with Bayes estimation via filtering. Math. Finance 13, 411–444.
Zhang, M.Y., Russell, J.R., Tsay, R.S., 2001. A nonlinear autoregressive conditional duration model with applications to financial transaction data. J. Econ. 104, 179–207.