On the Intertrade Waiting-time Distribution

Enrico Scalas, a,b,* Rudolf Gorenflo, c Hugh Luckock, d
Francesco Mainardi, e Maurizio Mantelli a and Marco Raberto f

a University of Piemonte Orientale, Italy
b INFM Genova, Italy
c Free University of Berlin, Germany
d University of Sydney, Australia
e University of Bologna, Italy
f University of Genova, Italy

Abstract

Continuous-time random walks can be used as phenomenological models of high-frequency time dynamics in financial markets. Empirical analyses show that the intertrade durations (or waiting-times) are non-exponentially distributed. This fact imposes constraints on agent-based models of financial markets based on continuous-double auctions.

Key words: Duration, Intertrade duration, Waiting-times, Survival function
JEL classification: C16, C22

1. INTRODUCTION

In recent times, there has been an increasing interest in the statistical properties of high-frequency financial data related to market microstructure (Goodhart and O’Hara, 1997; O’Hara, 1999; Madhavan, 2000; Dacorogna et al., 2001; Raberto et al., 2001; Luckock, 2003). High-frequency econometrics is now well established after research on conditional duration models (Engle and Russel, 1997, 1998; Bauwens and Giot, 2000; Lo et al., 2002; Gourieroux and Jasiak, 2003).

In high-frequency financial data not only returns but also waiting times between consecutive trades are random variables (Zumbach, 1998). This remark is present in a paper by Lo and MacKinlay (1990), but it can be traced at least to papers on the application of compound Poisson processes (Press, 1967) and subordinated stochastic processes (Clark, 1973) to finance. Models of tick-by-tick financial data based on compound Poisson processes were recently studied by Rydberg and Shephard (1998, 1999, 2000).

Compound Poisson processes belong to the more general class of continuous-time random walks (CTRWs) (Montroll and Weiss, 1965). The application of CTRW to problems in economics dates back, at least, to the 1980s. Hilfer (1984) discussed the application of stochastic processes to operational planning, and used CTRWs as tools for sale forecasts. The augmented CTRW formalism has been used as a model of high-frequency price dynamics in financial markets by our research group since 2000, in a series of three papers (Scalas et al., 2000; Mainardi et al., 2000, Gorenflo et al., 2001). After us, other scholars have used this formalism (Masoliver et al., 2003a, Masoliver et al., 2003b; Kutner and Switała, 2003). However, at the beginning of the 20th century, the PhD thesis of Lundberg presented a model for ruin theory of insurance companies, later elaborated by Cramér (Lundberg, 1903; Cramér, 1930). The underlying stochastic process for claims of the Lundberg-Cramér model is another example of compound Poisson process and thus of CTRW.

* Corresponding author. Email: scalas@unipmn.it

ISSN 1740-6242 © 2005 Global EcoFinance™ All rights reserved. 38
The remainder of this paper is organized as follows. In the second section, the CTRW theory is sketched and some basic quantities are defined. The third section focuses on the survival function and its anomalous behaviour. Finally, the fourth section concludes the paper with a discussion of the anomalous behaviour.

2. THEORY

In order to model tick-by-tick data, we use the so-called continuous-time random walk (CTRW), where time intervals between successive steps are random variables, as discussed by Montroll and Weiss (1965). In physics, CTRWs are used as models of diffusion with instantaneous jumps from one position to the next. For this reason they are suitable as models of high-frequency price dynamics.

Let $S(t)$ denote the price of an asset at time $t$. In a real market with a double-auction mechanism, prices are fixed when buy orders are matched with sell orders and a transaction (trade) occurs. It is more convenient to refer to returns rather than prices. For this reason, we shall take into account the variable $x(t) = \log[S(t)]$: the logarithm of the price.

For a small price variation $\Delta S = S(t_{i+1}) - S(t_i)$, one has that the return $r = \Delta S / S(t_i)$ and the logarithmic return $r_{\log} = \log[S(t_{i+1}) / S(t_i)]$ are almost equal.

CTRWs are essentially point processes with reward (Cox and Isham, 1979). The point process is characterized by a sequence of independent identically distributed (i.i.d.) positive random variables $\tau_i$, which can be interpreted as waiting times between two consecutive events:

$$t_n = t_0 + \sum_{i=1}^{n} \tau_i; \quad \tau_n = t_n - t_{n-1}; \quad n = 1, 2, 3, \ldots; \quad t_0 = 0. \quad (1)$$

The rewards are i.i.d. random variables: $\xi_i$. In the usual physical interpretation, the $\xi_i$'s represent jumps of a diffusing particle, and they can be $n$-dimensional vectors. Here, only the 1-dimensional case is considered; however, the extension of the results to $n$ dimensions is straightforward. The position $x$ of the particle at time $t$ is given by the following random sum (with $N(t) = \max\{n : t_n \leq t\}$ and $x(0) = 0$):

$$x(t) = \sum_{i=1}^{N(t)} \xi_i \quad (2)$$

In the financial interpretation, the $\xi_i$'s have the meaning of log-returns, whereas the positions $x(t)$ represent log-prices at time $t$. The time series $\{x(t_i)\}$ is characterised by $\phi(\xi, \tau)$, the joint probability density of log-returns $\xi_i = x(t_{i+1}) - x(t_i)$ and of waiting times $\tau_i = t_{i+1} - t_i$. The joint density satisfies the normalization condition: \( \int \int \phi(\xi, \tau) d\xi d\tau = 1 \).

It must be again remarked that both log-returns and waiting times are assumed to be independent and identically distributed (i.i.d.) random variables. This strong assumption is useful to derive limit theorems for the stochastic processes described by CTRWs. However, in financial time series, the presence of volatility clustering, as well as correlations between waiting times do falsify the i.i.d. hypothesis (McCauley, 2004).

In general, log-returns and waiting times are not independent from each other (Raberto et al., 2002; Meerschaert et al., 2002). By probabilistic arguments (see Montroll and Weiss, 1965; Mainardi et al., 2000; Scalas et al., 2003; Scalas et al., 2004), one can derive the following integral equation that gives the probability density, $p(x,t)$, for the particle of being in position $x$ at time $t$, conditioned by the fact that it was in position $x = 0$ at time $t = 0$:

$$p(x,t) = \delta(x)\Psi(t) + \int_0^t \int \phi(x-x',t-t')p(x',t')dxdt' \quad (3)$$

where $\delta(x)$ is Dirac's delta function and $\Psi(t)$ is the so-called survival function. $\Psi(t)$ is related to the marginal waiting-time probability density $\psi(\tau)$. The two marginal densities $\psi(\tau)$ and $\lambda(\xi)$ are:
\[
\begin{align*}
\psi(\tau) &= \int_{-\infty}^{+\infty} \phi(\xi, \tau) d\xi \\
\lambda(\xi) &= \int_{0}^{\infty} \phi(\xi, \tau) d\tau
\end{align*}
\] (4)

and the survival function \( \Psi(t) \) is:

\[
\Psi(t) = 1 - \int_{0}^{t} \psi(\tau) d\tau = \int_{t}^{\infty} \psi(\tau) d\tau
\] (5)

Both the two marginal densities and the survival function can be empirically derived from tick-by-tick financial data in a direct way. Indeed, the next section will be devoted to the empirical survival function.

3. EMPIRICAL RESULTS ON THE SURVIVAL FUNCTION

3.1. The DJIA data

The data set consists of nearly 800,000 prices \( S(t_i) \) and times of execution \( t_i \) from the TAQ database registered at NYSE in October 1999 for the 30 stocks of the Dow Jones Industrial Average Index, namely, at that time: AA, ALD, AXP, BA, C, CAT, CHV, DD, DIS, EK, GE, GM, GT, HWP, IBM, IP, JNJ, JPM, KO, MCD, MMM, MO, MRK, PG, S, T, UK, UTX, WMT, XON.\(^1\) These data were cleaned in order to remove misprints in prices and times of execution. The choice of only one month of high-frequency data was a trade-off between the necessity of managing enough data for significant statistical analyses and the goal of minimizing the effect of external economic influences.

In order to roughly evidence intraday patterns (Dacorogna et al., 2001), the data set has been divided into three daily periods: morning (from 9:00 to 10:59), midday (from 11:00 to 13:59) and afternoon (from 14:00 to 17:00).

3.2. The Anderson-Darling test on the survival function

In the presence of constant activity in the market, the survival function is exponential:

\[
\Psi(\tau) = e^{-\tau/\tau_0},
\] (6)

where \( \tau_0 \) is the average duration. This is the only memoryless distribution for a point process (Cox and Isham, 1979). Therefore, it is natural to test the empirical survival function against the above exponential model. A good statistical test of exponentiality of a survival function is the Anderson-Darling (AD) test (Stephens, 1974).

In Table 1, the values of the AD statistic are given for all the 30 DJIA stocks traded in October 1999. In all these cases the null hypothesis of exponentiality can be rejected at the 1% significance level (the limiting value is 1.957). It is therefore safe to conclude that the survival function for waiting times is non-exponential.

4. DISCUSSION AND CONCLUSIONS

Why are the above empirical findings on the survival function important? This has to do with the market price formation mechanisms in a continuous double-auction market. \textit{A priori}, there is no strong reason for independent market investors to place buy and sell orders in a time-correlated way. This argument would lead one to expect a Poisson process for inter-order waiting times. If the trade process were a simple thinning of the order process, then exponential waiting times should be expected between consecutive trades as well (Cox and Isham, 1979). Eventually, even if empirical analyses should show that time correlations are already present at the order level, it would be interesting to understand why they are there. In other words, the empirical results on the survival probability set limits on statistical market models for price formation.

\(^1\) Company symbols are available at http://www.nyse.com
Table 1. Empirical average waiting time $\tau_0$ and AD statistics $A^2$

<table>
<thead>
<tr>
<th>Stock</th>
<th>$\tau_{01}^m$ (s)</th>
<th>$\tau_{01}^{mi}$ (s)</th>
<th>$\tau_{01}^{af}$ (s)</th>
<th>$A^2$ (mo)</th>
<th>$A^2$ (mi)</th>
<th>$A^2$ (af)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>27.1</td>
<td>40.0</td>
<td>28.8</td>
<td>29.2</td>
<td>66.0</td>
<td>44.8</td>
</tr>
<tr>
<td>ALD</td>
<td>21.2</td>
<td>30.8</td>
<td>23.4</td>
<td>21.8</td>
<td>55.5</td>
<td>33.8</td>
</tr>
<tr>
<td>AXP</td>
<td>11.8</td>
<td>18.5</td>
<td>11.7</td>
<td>81.7</td>
<td>102.5</td>
<td>130.7</td>
</tr>
<tr>
<td>BA</td>
<td>22.0</td>
<td>32.0</td>
<td>22.6</td>
<td>17.4</td>
<td>20.2</td>
<td>21.2</td>
</tr>
<tr>
<td>C</td>
<td>7.1</td>
<td>10.5</td>
<td>8.2</td>
<td>252.2</td>
<td>142.8</td>
<td>210.7</td>
</tr>
<tr>
<td>CAT</td>
<td>29.2</td>
<td>42.4</td>
<td>31.6</td>
<td>72.3</td>
<td>128.7</td>
<td>64.6</td>
</tr>
<tr>
<td>CHV</td>
<td>22.1</td>
<td>34.3</td>
<td>27.1</td>
<td>104.4</td>
<td>121.5</td>
<td>64.9</td>
</tr>
<tr>
<td>DD</td>
<td>20.3</td>
<td>30.8</td>
<td>22.1</td>
<td>22.9</td>
<td>44.3</td>
<td>36.1</td>
</tr>
<tr>
<td>DIS</td>
<td>15.2</td>
<td>20.8</td>
<td>16.6</td>
<td>53.4</td>
<td>53.4</td>
<td>74.7</td>
</tr>
<tr>
<td>EK</td>
<td>34.1</td>
<td>51.2</td>
<td>36.3</td>
<td>24.8</td>
<td>34.8</td>
<td>44.3</td>
</tr>
<tr>
<td>GE</td>
<td>7.0</td>
<td>11.3</td>
<td>7.9</td>
<td>351.9</td>
<td>284.7</td>
<td>445.6</td>
</tr>
<tr>
<td>GM</td>
<td>24.6</td>
<td>36.6</td>
<td>27.0</td>
<td>22.4</td>
<td>60.8</td>
<td>40.9</td>
</tr>
<tr>
<td>GT</td>
<td>34.3</td>
<td>55.5</td>
<td>37.9</td>
<td>73.7</td>
<td>95.7</td>
<td>54.1</td>
</tr>
<tr>
<td>HWP</td>
<td>10.4</td>
<td>16.1</td>
<td>12.7</td>
<td>94.8</td>
<td>77.8</td>
<td>100.8</td>
</tr>
<tr>
<td>IBM</td>
<td>8.9</td>
<td>10.0</td>
<td>9.2</td>
<td>409.6</td>
<td>472.5</td>
<td>489.5</td>
</tr>
<tr>
<td>IP</td>
<td>24.8</td>
<td>36.3</td>
<td>27.0</td>
<td>25.0</td>
<td>37.2</td>
<td>19.4</td>
</tr>
<tr>
<td>JNJ</td>
<td>16.1</td>
<td>23.0</td>
<td>17.7</td>
<td>30.4</td>
<td>35.6</td>
<td>38.0</td>
</tr>
<tr>
<td>JPM</td>
<td>17.0</td>
<td>29.5</td>
<td>19.0</td>
<td>33.0</td>
<td>85.2</td>
<td>85.8</td>
</tr>
<tr>
<td>KO</td>
<td>12.9</td>
<td>18.3</td>
<td>14.4</td>
<td>44.5</td>
<td>37.8</td>
<td>44.1</td>
</tr>
<tr>
<td>MCD</td>
<td>19.4</td>
<td>29.3</td>
<td>22.1</td>
<td>40.9</td>
<td>72.7</td>
<td>44.1</td>
</tr>
<tr>
<td>MMM</td>
<td>30.1</td>
<td>42.0</td>
<td>30.4</td>
<td>80.1</td>
<td>86.8</td>
<td>37.5</td>
</tr>
<tr>
<td>MO</td>
<td>11.4</td>
<td>15.6</td>
<td>12.9</td>
<td>74.2</td>
<td>89.0</td>
<td>75.2</td>
</tr>
<tr>
<td>MRK</td>
<td>11.7</td>
<td>16.8</td>
<td>13.2</td>
<td>133.1</td>
<td>136.0</td>
<td>189.8</td>
</tr>
<tr>
<td>PG</td>
<td>16.2</td>
<td>23.6</td>
<td>17.9</td>
<td>43.5</td>
<td>37.2</td>
<td>48.8</td>
</tr>
<tr>
<td>S</td>
<td>23.4</td>
<td>38.8</td>
<td>28.6</td>
<td>40.1</td>
<td>23.0</td>
<td>41.6</td>
</tr>
<tr>
<td>T</td>
<td>8.8</td>
<td>12.2</td>
<td>10.6</td>
<td>193.2</td>
<td>179.1</td>
<td>208.9</td>
</tr>
<tr>
<td>UK</td>
<td>40.4</td>
<td>69.1</td>
<td>46.7</td>
<td>33.8</td>
<td>72.4</td>
<td>47.2</td>
</tr>
<tr>
<td>UTX</td>
<td>28.5</td>
<td>39.3</td>
<td>29.0</td>
<td>33.7</td>
<td>62.9</td>
<td>58.0</td>
</tr>
<tr>
<td>WMT</td>
<td>12.5</td>
<td>18.2</td>
<td>14.9</td>
<td>105.2</td>
<td>110.6</td>
<td>139.1</td>
</tr>
<tr>
<td>XON</td>
<td>12.0</td>
<td>19.6</td>
<td>14.1</td>
<td>104.8</td>
<td>121.4</td>
<td>129.0</td>
</tr>
</tbody>
</table>

For each daily period, the values of the empirical average waiting time $\tau_0$ and the AD statistics $A^2$ are given (Stephens, 1974).

A simple explanation of the anomalous behaviour of durations can be given in terms of exponential mixtures due to variable activity during the trading day.

Suppose that a trading day can be divided into $N$ subintervals where waiting times follow an exponential distribution with different average waiting times $\tau_{01}, \ldots, \tau_{0N}$. Using the rate (the inverse of the average waiting time: $\mu_i = 1/\tau_{0i}$), the survival function can be written as:

$$
\Psi(\tau) = \sum_{i=1}^{N} a_i e^{-\mu_i \tau}
$$

where $a_i$ are suitable weights whose sum must be 1, as $\Psi(0) = 1$. This sum of exponential components is itself non-exponential. For illustrative purposes, in Figure 1, there is a comparison between eq. (7) and simulated data in which the day has been divided into 10 intervals of equal weight. In each interval the average waiting time between trades was a constant and the waiting times followed an exponential distribution. The value of the constant increased from 10 to 50 seconds in the first five intervals and then decreased from 40 to 5 seconds in the last five intervals, so that the sequence of waiting times (in seconds: 10,20,30,40,50,40,30,20,10,5) is a rough representation of the activity in a real financial market. The open circles are the survival function of the Monte Carlo simulation, the solid line represents the single exponential fit of the survival function, whereas, the crosses are values of the survival function computed according to eq. (7) with $a_i = 1/10$ for each $i$. Even if, for long waiting times, the tail of the distribution is again exponential with rate equal to $\mu_i = 1/5$, the exponential mixture can describe deviations from the single exponential law for short and intermediate waiting times.
In conclusion, we have shown that, in October 1999, waiting times between consecutive trades in the 30 NYSE DJIA stocks were non-exponentially distributed. We have argued that this fact has implications for market microstructural models that should be able to reproduce such a non-exponential behaviour to be realistic. Finally, we have offered a possible explanation of the anomaly in terms of variable trading activity during the day.

Figure 1. Survival function for simulated data (open circles) compared to a simple exponential fit (solid line) and to a mixture of exponentials (crosses)

REFERENCES


Lundberg, F. (1903) *Approximerad Framställning av Sannolikehets-funktionen. Återförsäkring av Kollektivrisker*. Almqvist & Wiksell, Uppsala.


