Exploratory analysis of intradaily stock returns: A semi Markov chain approach

Jonas Andersson
Anders Christoffersson

Division of Statistics
Research Report 2004:4
Exploratory analysis of intradaily stock returns: A semi Markov chain approach

Jonas Andersson* and Anders Christoffersson†

September 27, 2004

Abstract

In this paper, we study the intradaily distributional and temporal properties of the IBM stock in the time period November 1990 until February 1991. We do this by exploratory analysis by means of semi Markov chains, i.e. a Markov chain where time between events is considered random. Furthermore, the behaviour of the method is studied under some commonly used models.

Keywords: Intraday data, duration, volatility, exploratory analysis, semi Markov chain.

1 Introduction

Many approaches exist for analyzing the changes of the distributional properties of financial returns, see e.g. Bollerslev et al. (1992) or Shephard (1996) for a review of the area. A common feature of most of these approaches is that this time dependence is assumed to be present in the mean or, even more frequently, in the variance of the distribution, conditioned on all previous observations. The explanation for this, we believe, is twofolded. Firstly, using daily data, which has been the most commonly used data frequency in recent years, does not encourage non-parametric methods, because of the large datasets required for such analyses. Secondly, the focus has often been on the volatility, which is most often measured as the variance of the conditional distribution. It is, however, not clear why this should be the best measure of volatility. If we are looking for a measure of risk, it is the tails of the conditional distribution that is of interest and since this distribution is invariably non-normal, the variance does not reveal all information of the probability mass of the tails.

Intradaily data become more and more available as computational and data storage capacity is getting cheaper. We can, therefore, study the time dynamics on an intraday level, and we have, for this purpose, obviously more data points than for e.g. daily data. A common way to use these data is to average squares over one day, obtaining the so called

*Corresponding author. Department of Finance and Management, Norwegian School of Economics and Business Administration, N-5045 Bergen, Norway
†Department of Information Science, Division of Statistics, Box 513, SE-751 20 Uppsala, Sweden.
realized volatility, see e.g. Andersen et al. (2001). This quantity can then be modeled basically by using time series models for the conditional mean. However, non-parametric methods for the analysis of the time dynamics of the entire distribution is now more likely to give usable information. In Andersson and Newbold (2002) this is done by functional data analysis where the time dynamics of the estimated distribution function of each day is analyzed rather than just the realized volatility. In the rest of this paper we will use the theory of Markov chains to model the conditional distribution of asset returns. This idea was also exploited by Russel and Engle (1998) but then in a parametric setting. Another related paper, also a parametric variant, is Rydberg and Shephard (2003).

As opposed to the literature on financial economics, we focus entirely on the data analysis of intradaily data. Thus, we do not address the problem of removing market microstructure effects which is a very important topic when one would like to test economic theories using intradaily data. Rather, we develop a method aimed at analyzing temporal- and distributional aspects of intradaily data. The reader interested in cleaning and filtering high frequency data is referred to Dacorogna et al. (2001).

In the next section we outline the approach and describe how to form a discretely valued time series from the observed one. In Section 3 the properties of these discrete valued time series, originally created from data generated by some specific time series models, will be derived. We will however limit our discussion concerning this to regularly spaced time series. Applications of the methods to the IBM stock are presented in Section 4. Finally, conclusions are presented.

2 The general setup

With the purpose of making the paper as self contained as possible we will in this section in detail explain the approach taken to model the irregularly spaced time series at hand.

In order to get a unified way of expressing all the quantities we are interested in we define the matrix

$$ N = \begin{bmatrix} n_{1,1} & n_{1,(0|1)} & n_{1,2} & n_{1,(0|2)} & \ldots & n_{1,k} & n_{1,(0|k)} \\ n_{(0|1),1} & n_{(0|1),(0|1)} & n_{(0|1),2} & n_{(0|1),(0|2)} & \ldots & n_{(0|1),k} & n_{(0|1),(0|k)} \\ n_{2,1} & n_{2,(0|1)} & n_{2,2} & n_{2,(0|2)} & \ldots & n_{2,k} & n_{2,(0|k)} \\ n_{(0|2),1} & n_{(0|2),(0|1)} & n_{(0|2),2} & n_{(0|2),(0|2)} & \ldots & n_{(0|2),k} & n_{(0|2),(0|k)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ n_{k,1} & n_{k,(0|1)} & n_{k,2} & n_{k,(0|1)} & \ldots & n_{k,k} & n_{k,(0|k)} \\ n_{(0|k),1} & n_{(0|k),(0|1)} & n_{(0|k),2} & n_{(0|k),(0|2)} & \ldots & n_{(0|k),k} & n_{(0|k),(0|k)} \end{bmatrix} $$

(1)

where $k$ is the number of states and $n_{i,j}$ is the number of times a transition from state $i$ to state $j$ occurs without delay. The classes $(0|j)$ are defined as states where no trade occurs after the last trade has been of type $j$. By considering these latter states we can conclude that the matrix is defined in such a way that some classes are impossible to enter after a visit in some of the other classes, i.e. some states do not communicate. In fact, the matrix
can be written

\[
N = \begin{bmatrix}
n_{1,1} & n_{1,(0),1} & n_{1,2} & 0 & \ldots & n_{1,k} & 0 \\
n_{(0),1,1} & n_{(0),1,2} & 0 & \ldots & n_{(0),1,k} & 0 \\
n_{2,1} & 0 & n_{2,2} & n_{2,(0),2} & \ldots & n_{2,k} & 0 \\
n_{(0),2,1} & 0 & n_{(0),2,2} & n_{(0),2,(0),2} & \ldots & n_{(0),2,k} & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
n_{k,1} & 0 & n_{k,2} & 0 & \ldots & n_{k,k} & n_{k,(0),k} \\
n_{(0),k,1} & 0 & n_{(0),k,2} & 0 & \ldots & n_{(0),k,k} & n_{(0),k,(0),k}
\end{bmatrix}
\]

(2)

If time is measured in seconds, as is the case in our dataset, the matrix will have their largest values for transitions of the type \((0|j), (0|j)\) since trades are occurring rather infrequently in relation to the number of seconds during a trading day. With an analogue notation the matrix \(P\) with elements \(p_{i,j}\) defines the transition probability matrix.

\[
P = \begin{bmatrix}
p_{1,1} & p_{1,(0),1} & p_{1,2} & 0 & \ldots & p_{1,k} & 0 \\
p_{(0),1,1} & p_{(0),1,2} & 0 & \ldots & p_{(0),1,k} & 0 \\
p_{2,1} & 0 & p_{2,2} & p_{2,(0),2} & \ldots & p_{2,k} & 0 \\
p_{(0),2,1} & 0 & p_{(0),2,2} & p_{(0),2,(0),2} & \ldots & p_{(0),2,k} & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
p_{k,1} & 0 & p_{k,2} & 0 & \ldots & p_{k,k} & p_{k,(0),k} \\
p_{(0),k,1} & 0 & p_{(0),k,2} & 0 & \ldots & p_{(0),k,k} & p_{(0),k,(0),k}
\end{bmatrix}
\]

(3)

Another conditional probability matrix of interest is the one describing the sequence of trade states ignoring the size of the durations between trades. We therefore define

\[
N^H = \begin{bmatrix}
n_{1,1}^H & n_{1,2}^H & \ldots & n_{1,k}^H \\
n_{2,1}^H & n_{2,2}^H & \ldots & n_{2,k}^H \\
\vdots & \vdots & \ddots & \vdots \\
n_{k,1}^H & n_{k,2}^H & \ldots & n_{k,k}^H
\end{bmatrix}
\]

(4)

where \(n_{i,j}^H\) is the number of times a trade in class \(i\) is followed by a trade in class \(j\). The corresponding transition probability matrix is denoted \(P^H\) with elements \(p_{i,j}^H\).

\[
P^H = \begin{bmatrix}
p_{1,1}^H & p_{1,2}^H & \ldots & p_{1,k}^H \\
p_{2,1}^H & p_{2,2}^H & \ldots & p_{2,k}^H \\
\vdots & \vdots & \ddots & \vdots \\
p_{k,1}^H & p_{k,2}^H & \ldots & p_{k,k}^H
\end{bmatrix}
\]

(5)

The elements of \(N\) and \(N^H\) are related according to

\[
n_{i,j}^H = n_{i,j} + n_{(0|i),j}
\]

(6)

and the elements of \(P\) and \(P^H\) by

\[
p_{i,j}^H = \frac{p_{i,j}π_i + p_{(0|i),j}π_{(0|i)}}{\sum_{j=1}^{k}(p_{i,j}π_i + p_{(0|i),j}π_{(0|i)})},
\]

(7)
where the \( \pi \)'s are the stationary probabilities for the different states, see appendix. The first quantity we are interested in are estimates of the transition probabilities in the matrix \( P = [p_{i,j}], i,j = 1,(0|1), ..., k,(0|k). \)

\[
\hat{p}_{i,j} = \frac{n_{i,j}}{n_{i,\cdot}} \tag{8}
\]

where

\[
n_{i,\cdot} = \sum_{j=1}^{k} n_{i,j} \tag{9}
\]

Conditional on the number observations in a state, \( n_{i,\cdot} \), the variance of \( \hat{p}_{i,j} \), is known to be

\[
Var(\hat{p}_{i,j}) = \frac{p_{i,j}(1-p_{i,j})}{n_{i,\cdot}} \tag{10}
\]

so an approximate confidence interval around \( \hat{p}_{i,j} \) can be calculated by

\[
\hat{p}_{i,j} \pm z_\alpha \sqrt{\frac{\hat{p}_{i,j}(1-\hat{p}_{i,j})}{n_{i,\cdot}}} \tag{11}
\]

where \( \alpha \) is the chosen confidence level.

Within this framework we can also consider \( D_j \), the duration to the next trade after a trade in class \( j \). This quantity is, under the Markovian assumption, geometrically distributed with probability \( 1-p(0|j),(0|j) \).

\[
D_{ij} \sim Ge(1-p(0|j),(0|j)) \tag{12}
\]

We would like to make inference about the expected duration between trades, estimated by \( \overline{D}_j = \frac{\sum_{i=1}^{n_j(0|j)} / n_{j,(0|j)} D_{ij}}{n_{j,(0|j)}} \). The variance for the expected duration, conditional on the number of times the process enters waiting state after trade in class \( j \) (\( n_{j,(0|j)} \)) would be

\[
Var(\overline{D}_j) \sim \frac{p(0|j),(0|j)}{(1-p(0|j),(0|j))^2 n_{j,(0|j)}} \tag{13}
\]

Another alternative is to use the maximum likelihood estimator of \( p(0|j),(0|j) \) and estimate the expected duration by

\[
\hat{\mu}_{D_j} = \frac{1}{1-\hat{p}(0|j),(0|j)} \tag{14}
\]

and then use a Taylor expansion of \( \hat{\mu}_{D_j} \) as a function of \( \hat{p}(0|j),(0|j) \) to obtain the asymptotic standard error. It turns out that the variance of \( \hat{\mu}_{D} \) computed this way can be shown to be the same as that of \( Var(\overline{D}_j) \).

A third, more direct approach, is to use the durations themselves and estimate the sample mean, \( \overline{D}_j \) (which is identical to \( \hat{\mu}_{D_j} \)) together with the variance

\[
\overline{Var}(\overline{D}_j) = \frac{S^2_{D_j}}{n_{j,(0|j)}} \tag{15}
\]

Equation (15) is based on the assumption of a constant hazard, an assumption which is not necessarily fulfilled, as will be illustrated in the application in Section 5.1.2. We will come back to this problem there. Another problem is the strong diurnal variation in
the durations. If we regard this as a fixed effect, it will imply that \( \hat{\text{Var}}(D_j) \) will tend to be biased upwards since this diurnal variation is of a low frequent character.

To make a further connection to survival analysis, an estimate of the discrete-time hazard function can be calculated as

\[
\hat{\mu}_j(x) = \frac{\#\{D_j = x\}}{\#\{D_j > x - 1\}}
\]

where \( \#\{D_j = x\} \) is the number of transactions that occurred \( x \) seconds after a previous transaction that was in class \( j \). \( \#\{D_j > x - 1\} \) is the number of durations larger than \( x - 1 \) seconds. Since the hazard is defined as a quantity conditional on the denominator, the standard errors can be calculated simply by a normal approximation to the binomial distribution.

### 3 Connection between the MC approach and some simple Markovian time series models

Assume that the ordinal time series \( \{z_t\} \) is derived from the time series \( \{x_t\} \) by categorizing it as

\[
z_t = j \quad \text{if} \quad x_t \in (c_{j-1}, c_j]
\]

where \( c_j, j = 0, 1, \ldots, k \). Usually \( c_0 = -\infty \) and \( c_k = \infty \). In order to study the connection between the data generating process of \( \{x_t\} \) and \( P \), the transition matrix for \( \{z_t\} \), we need to calculate the components of the formula

\[
p_{ij} = P(z_t = j | z_{t-1} = i) = \frac{P(c_{j-1} < x_t < c_j | c_{i-1} < x_{t-1} < c_i)}{P(c_{i-1} < x_{t-1} < c_i)}.
\]

Since both the numerator and denominator in (17) usually are unavailable in volatility models, we have to use simulation studies in these cases. First, however, we start by considering the AR(1) case, where an explicit solution is tractable. The models we study here are, in contrast to the ones in the previous section, usually employed for daily data. Nevertheless, since the properties they are meant to model carry over to the intraday case, we find it worthwhile to study them in this context.

#### 3.1 AR(1)

Assume that \( \{x_t\} \) are driven by the process

\[
x_t = \phi x_{t-1} + \varepsilon_t
\]

where \( \{\varepsilon_t\} \) is a normally distributed white noise process with standard deviation \( \sigma \). Then, given that the process starts in the stationary distribution

\[
\begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix} \sim N_2 \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \phi \sigma^2 \\ \phi \sigma^2 & \phi^2 \sigma^2 \end{bmatrix} \right)
\]

(19)
$p_{ij}$ can in this case be expressed in terms of the univariate and the bivariate normal distribution functions $F_1(\cdot)$ and $F_2(\cdot, \cdot)$.

$$p_{ij} = \frac{F_2(c_j, c_i) - F_2(c_j, c_{i-1}) - F_2(c_{j-1}, c_i) + F_2(c_{j-1}, c_{i-1})}{F_1(c_i) - F_1(c_{i-1})} \quad (20)$$

The three transition matrices below are examples of this for $\phi = -0.9$, $\phi = 0$ and $\phi = 0.9$ in an example where $k$ was set to 5. The thresholds were chosen by guidance of our application, presented in the next section. This lead to class 1 having 5% of the observations, class 2 13%, class 3 64%, class 4 13% and class 5 5%.

\[
\begin{array}{cccccc}
\phi = -0.9 & \phi = 0 & \phi = 0.9 \\
0.00 & 0.00 & 0.04 & 0.32 & 0.64 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.36 & 0.51 & 0.14 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.07 & 0.85 & 0.07 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.14 & 0.51 & 0.36 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.64 & 0.32 & 0.04 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
\end{array}
\]

Table 1: Transition matrices for discretised data from AR(1)-processes.

In the case of positive autocorrelation, which is the case in the right matrix, the conditional probabilities in the diagonal are large, meaning that a large return has a high probability to be followed by another large return with the same sign. The left matrix on the other hand, representing negative autocorrelation, illustrates the case where a large return is likely to be followed by a large return with the opposite sign. The matrix in the middle illustrates the case where there is no time dependence at all. However trivial, this example illustrates something important. The equality of the rows in the middle matrix does not only mean a lack of autocorrelation but a complete lack of first-order dependence, it indicates strict white noise. This is what is exploited in this paper to explore the data without too many parametric assumptions.

In the case above the transition matrix is constant over time, the time dynamics is only manifested in the differences between the rows. However, we can also consider a time varying conditional distribution. This can be illustrated by plotting the conditional probabilities against $\phi$. The graph below presents how the elements of the first column of $P$ varies when $\phi$ varies.

As can be seen in Figure 3.1, the probability of staying for at least one time unit in class 1, once getting there, increases monotonously with $\phi$ while the opposite is true for the probability of going to 1 if the previous state was 5. The explanation for this is simply given by the nature of autocorrelation. Leaving class 1 for class 2 is first getting more likely as $\phi$ increases. However, as $\phi$ becomes greater than 0.5 it starts to decrease, see row 1. This is connected to the increased probability of staying in class 1.

### 3.2 ARCH(1)

Here, $\{x_t\}$ is generated by the following scheme

$$(x_t|x_{t-1}) \sim N(0, \alpha + \beta x_{t-1}^2). \quad (21)$$

In this case we do not have a closed form expression for the bivariate distribution of $(x_t, x_{t-1})$. Because of this we instead simulated 100000 observations for three different, stationary, models. The first 100 observations were removed in order to deal with the effects of starting values.
Figure 1: The first column of the transition matrix for different values of $\phi$. The variance is held to one for all parameter combinations.

\[
P = \begin{bmatrix}
0.05 & 0.13 & 0.64 & 0.12 & 0.05 \\
0.06 & 0.12 & 0.64 & 0.12 & 0.06 \\
0.05 & 0.13 & 0.64 & 0.12 & 0.05 \\
0.05 & 0.12 & 0.65 & 0.12 & 0.06
\end{bmatrix} \quad P = \begin{bmatrix}
0.16 & 0.17 & 0.32 & 0.18 & 0.17 \\
0.05 & 0.13 & 0.63 & 0.14 & 0.05 \\
0.03 & 0.09 & 0.76 & 0.09 & 0.03 \\
0.17 & 0.17 & 0.32 & 0.18 & 0.17
\end{bmatrix} \quad P = \begin{bmatrix}
0.25 & 0.18 & 0.15 & 0.18 & 0.24 \\
0.05 & 0.13 & 0.66 & 0.12 & 0.05 \\
0.01 & 0.03 & 0.94 & 0.03 & 0.01 \\
0.25 & 0.18 & 0.15 & 0.19 & 0.24
\end{bmatrix}
\]

Table 2: Transition matrices for discretized data from ARCH(1)-processes.

The matrices for the values $\beta = 0.5$ and $\beta = 0.9$ above reflect the fact that large changes, of either sign, gives thicker tails of the distribution than small changes does. The deviation from symmetry of the columns are completely due to Monte Carlo uncertainty. With a sufficiently large dataset we are able to detect this phenomena without specifying it explicitly in terms of a parametric model. Since we specify the conditional distribution non parametrically we are also enabled to separate this kind of time dynamics from others, e.g. such as autocorrelation. The persistent volatility dynamics that is possible to generate with an ARCH model of this kind is manifested by the large variance in class one and five and small variance in class two, three and four of the conditional distribution defined by the matrix to the right of Table 3.2.

In Figure 2, the first column of transition matrices is illustrated in the form of four lines of the elements of the matrix against $\phi$. For $\beta = 0$, corresponding to no conditional heteroskedasticity, the probabilities are independent of what the previous state was. While $\beta$ increases the probabilities of going from (or remain in) state one and four at time $t - 1$ to state one at time $t$ decrease while the opposite is true for going to state two and three.
Figure 2: The first column of the transition matrix for different values of $\beta$. The variance is held to one for all parameter combinations. 100000 observations were generated for each parameter combination.

### 3.3 EGARCH(1,1)

In this model (Nelson, 1991), $\{x_t\}$ is generated

$$(x_t|x_{t-1}, \sigma_0) \sim N(0, \sigma_t^2)$$

where

$$\sigma_t^2 = \exp (\alpha + \delta \log (\sigma_{t-1}^2) + \beta (|v_{t-1}| - E [|v_{t-1}] + \gamma v_{t-1}))$$

and

$$v_t = \frac{x_t}{\sigma_t}$$

is the error term of the model, assumed to be a white noise. If $v_t$ is normally distributed with mean zero and variance one, $E [|v_{t-1}] = \sqrt{2/\pi}$. The feature of this model, distinguishing it from the ARCH model, is that the parameter $\gamma$, if non zero, allows positive shocks (positive $v_{t-1}$) to affect the volatility $\sigma_t$ differently than negative shocks. This is called the leverage effect since it means that a decrease in stock price implies higher leverage. An increased leverage is supposed to increase volatility more than a decreased leverage. The consequence of this on the transition matrix is that it will look less symmetric. Consider the examples below. In the first matrix, $\gamma$ is set to $-2$. In this case a positive chock will reduce volatility, while a negative chock will increase it. In the matrix to the right, $\gamma$ is $-0.5$. Here a positive chock increases volatility less than a negative chock.

Except for the statements made about the corresponding matrices for the ARCH(1) model in the previous subsection, another observation can be made from the matrices above. The row-wise symmetry is gone in the most extreme case of Figure 3.3, the most left matrix. The reason for this is that the variance in each row (conditional variance) is dependent not only of the size of the absolute value but also of the sign of the previous

---

$^1$The problem of starting the simulation of the series were dealt with in the same way as for the ARCH model in the previous section.
value. Because of the sign of $\gamma$ (negative), the conditional variance is larger in the first class than in the fifth since the variance is increased more by a large negative shock than by a large positive.

By varying the parameter $\gamma$ we can study the leverage effect in terms of transition probabilities. In Figure 3 the effect is shown for the first column of the transition matrix. For values close to $-2$ we can see that whether the previous return was moderate and positive (class 4) or large and positive (class 5) did basically not matter. However, a large and negative previous value result in a significantly larger probability of obtaining a large negative return (class 1) in the present time period than a moderate and negative return (class 2) does. As we move $\gamma$ closer to zero there are two points were $p_{41}$ crosses $p_{31}$ and $p_{21}$, respectively. This means that the $\gamma$-parameter regulates the degree of which the sign of the previous return affects the conditional distribution of the present.

### 3.4 AR(1)-ARCH(1)

Finally, two matrices of $\alpha$, as it turns out, more realistic situation are given. Here $\{x_t\}$ is driven by the data generating process given by

$$x_t = \phi x_{t-1} + \sigma_t \varepsilon_t$$

where $\varepsilon_t$ is a strict white noise with marginal distribution $N(0, 1)$ and
\[ \sigma_t^2 = \alpha_0 + \beta \varepsilon_t^2 \]

\[ \phi = -0.3, \beta = 0.9 \]

\[
P = \begin{bmatrix}
0.13 & 0.08 & 0.20 & 0.26 & 0.33 \\
0.03 & 0.07 & 0.68 & 0.16 & 0.06 \\
0.01 & 0.02 & 0.95 & 0.02 & 0.01 \\
0.16 & 0.17 & 0.68 & 0.06 & 0.04 \\
0.32 & 0.24 & 0.19 & 0.08 & 0.16 \\
\end{bmatrix}
\]

\[ \phi = 0.3, \beta = 0.9 \]

\[
P = \begin{bmatrix}
0.34 & 0.23 & 0.20 & 0.09 & 0.14 \\
0.06 & 0.16 & 0.65 & 0.08 & 0.04 \\
0.01 & 0.03 & 0.91 & 0.03 & 0.01 \\
0.04 & 0.07 & 0.66 & 0.17 & 0.06 \\
0.14 & 0.08 & 0.21 & 0.24 & 0.33 \\
\end{bmatrix}
\]

Table 4: Transition matrices for discretized data from AR(1)-ARCH(1)-processes.

The AR(1) parameter \( \phi \) is kept moderate since this is the situation one encounter in practice, see e.g. Campbell et al. (1997), Chapter 2, which concludes that “…the serial correlation is both statistically and economically insignificant”. By the same reason \( \beta \) is kept large. As will be seen later, the matrix obtained with negative \( \phi \) is the one most likely to occur if we consider trade-to-trade returns, ignoring durations. Combining the arguments for AR(1) and ARCH(1) processes in the previous examples, we conclude that the variation deduced for row 1 and 5 in the left matrix is larger than from the middle classes. Also, the negative autocorrelation is implied by the relatively large conditional probabilities on the up-right-to-down-left-diagonal.

All these observations would obviously have been possible to make by estimating the appropriate parametric models and study the relevant parameter estimates. However, the point is that then we would have had to know which model to study. With the, more intuitive and simple, Markov chain approach we can, with only the assumption of stationarity, obtain the same observation. The cost is that we have to subjectively choose the quantiles that define the classes and that the method is non-parametric and thereby needs more datapoints to do the same job as a parametric one.

4 The data

The dataset is the one of intraday prices of the IBM stock used in Engle (2000). Since the data set is collected during the period 1990-11-01 - 1991-01-31, i.e. during the first Gulf war, there are some particularly interesting dates to have in mind while studying the development of the stock price.

- November 29th, 1990 (day 19 in the dataset). The United Nations Security Council agrees resolution 678, a deadline of January 15th, 1991 for Iraq to withdraw from Kuwait and calls for the use of force to impose a withdrawal if necessary.
- January 15th, 1991 (day 48 in the dataset). The deadline expires.

The data has been prepared in the same way as in Engle (2000). Christmas Eve, New Years Eve and the day after thanksgiving have all been removed. Furthermore, all observations before 9.40 and after 16.00 have been removed as well as the overnight returns.
In order to use the exploratory method outlined in Section 2 we need to classify the data. We choose to do this in the following way. The first differences of the observed prices are classified so that the zeroes are defined as one class, class 3. These happen to constitute 64% of the observations. Of the remaining 36% of the observations 25% were allocated to the extreme classes 1 and 5, giving 5% to class 1, 13% to class 2, 13% to class 4 and 5% to class 5. The classes with the most extreme negative and positive values, respectively. The remaining observations are allocated to class 2 and 4 depending if they are negative or positive. The resulting time series is now a sequence of returns irregularly spaced in time. We now create, as explained in Section 2, a regularly spaced time series, one observation for each second, by introducing 5 new classes representing the events “no trade occurs given that the last event was in class $i$, $i = 1,2,3,4,5$. The resulting time series thus has 10 classes.

![IBM Stock Price Chart](image)

Figure 4: The IBM stock 1990-11-01.

To illustrate the nature of the data the price series for the first day of the data set is plotted against time in Figure 4. The observed series takes jumps when a trade occur. The trades occur irregularly in time. One can also note that, since the jumps occur in 1/8 of a dollar, the discretisation that we make does not summarize the information as much as would be the case if the data followed jumps drawn from a truly continuous distribution.

Furthermore, in order to evaluate the possible non-stability of the conditional distributions and the other quantities of interest, we will use the described technique on a sliding window of the length 9 days. The 60 days will thus yield a sequence of length 52 for each conditional probability and the other quantities, respectively. These will be plotted against time together with their confidence bands.
5 Results

5.1 Results for the entire sample

5.1.1 Estimated transition matrices

We first present the estimated one-step transition matrix for the entire sample. This matrix corresponds to (3) with \( k = 5 \) and 5\% of the sample allocated to classes 1 and 5, respectively.

With time measured in seconds we see that 94.9\% of trades in class 2 (moderately negative return) are followed by at least one second with no trading. This can be contrasted with class 5 where only 92.9\% of trades are followed by at least one second with no trade. In general, a second with no trade is more often followed by a second with no trade than is a second where a trade have occurred.

\[
\hat{P} = \begin{bmatrix}
0.003 & 0.943 & 0.000 & 0.000 & 0.029 & 0.000 & 0.003 & 0.000 & 0.022 & 0.000 \\
0.001 & 0.968 & 0.001 & 0.000 & 0.013 & 0.000 & 0.003 & 0.000 & 0.014 & 0.000 \\
0.000 & 0.000 & 0.001 & 0.949 & 0.026 & 0.000 & 0.022 & 0.000 & 0.002 & 0.000 \\
0.000 & 0.000 & 0.001 & 0.964 & 0.018 & 0.000 & 0.015 & 0.000 & 0.002 & 0.000 \\
0.001 & 0.000 & 0.004 & 0.000 & 0.937 & 0.004 & 0.000 & 0.002 & 0.000 & 0.000 \\
0.001 & 0.000 & 0.004 & 0.000 & 0.025 & 0.965 & 0.004 & 0.000 & 0.001 & 0.000 \\
0.002 & 0.000 & 0.018 & 0.000 & 0.041 & 0.000 & 0.002 & 0.936 & 0.000 & 0.000 \\
0.002 & 0.000 & 0.014 & 0.000 & 0.021 & 0.000 & 0.001 & 0.962 & 0.000 & 0.000 \\
0.020 & 0.000 & 0.004 & 0.000 & 0.043 & 0.000 & 0.002 & 0.000 & 0.002 & 0.929 \\
0.013 & 0.000 & 0.003 & 0.000 & 0.018 & 0.000 & 0.001 & 0.000 & 0.001 & 0.964 \\
\end{bmatrix}
\]

The number of observations in each class is given by

\[
n = (2212, 65665, 6646, 173597, 32004, 848989, 6686, 164255, 2227, 57657)'\]

We now give an example what can be deduced from the matrix. We can e.g. observe that the probability of a second with no trade one second after a trade in class 1 seems to be lower than the probability of no trade given that no trade occurred last second and the previous trade was in class 1. This can be seen by considering the approximate confidence intervals around the estimates \( \hat{p}_{1,(0|1)} \) and \( \hat{p}_{(0|1),(0|1)} \).

\[
\hat{p}_{1,(0|1)} \pm 1.96 \sqrt{\frac{\hat{p}_{1,(0|1)}(1 - \hat{p}_{1,(0|1)})}{n_{1,(0|1)}}} = 0.943 \pm 0.010
\]

and

\[
\hat{p}_{(0|1),(0|1)} \pm 1.96 \sqrt{\frac{\hat{p}_{(0|1),(0|1)}(1 - \hat{p}_{(0|1),(0|1)})}{n_{(0|1),(0|1)}}} = 0.968 \pm 0.001
\]

Pointing out the common misconception that two quantities whose 95\% confidence intervals fail to overlap are significantly different at the 5\%-level, Goldstein and Healy (1995) presents a procedure for calculating the probability that the confidence intervals of the means, \( \mu_i \) and \( \mu_j \), of two independent normally distributed samples are overlapping given
that they the means are equal. The formula for this probability, $\gamma_{ij}$, at the confidence level, $1 - 2\beta$, is

$$\gamma_{ij} = 2 \left[ 1 - \Phi \left( z_{\beta} \frac{\sigma_i + \sigma_j}{\sigma_{ij}} \right) \right]$$  \hspace{1cm} (25)$$

where $z_{\beta}$ is the $1 - \beta$ quantile of the standard normal distribution, $\Phi$ is the standard normal cumulative distribution function, $\sigma_i$ and $\sigma_j$ are the standard deviations of the two sample means and $\sigma_{ij} = \sqrt{\sigma_i^2 + \sigma_j^2}$ is the standard deviation of the sum of the two sample means.

Fitting the method to our case with two probabilities $p_i$ and $p_j$ (representing elements in the transition matrix) and sample sizes $n_i$ and $n_j$ (representing the number of observations in the rows) we can write

$$\sigma_k = \sqrt{p_k(1 - p_k)/n_k}, \quad k = i, j$$

and

$$\sigma_{ij} = \sqrt{p_i(1 - p_i)/n_i + p_j(1 - p_j)/n_j}$$

and (25) can, under the additional assumption that $p_i = p_j$, be rewritten

$$\gamma_{ij} = 2 \left[ 1 - \Phi \left( z_{\beta} \sqrt{\frac{1 + 2\sqrt{n_i n_j}}{n_i + n_j}} \right) \right]$$  \hspace{1cm} (26)$$

By using this method, the probability that the two intervals are not overlapping, given that the probabilities are equal, is calculated to 0.0225. It should be noted that, if considering probabilities in the same row of the transition matrix, there is a covariance between the two estimates. We would like to discourage from considering formal hypothesis tests in this framework since this might give the impression that one has not looked at the matrix prior to testing. The danger of mass significance is obvious.

We now consider the transition matrix from trade to trade, $\hat{P}^H$, i.e. ignoring the size of durations between trades.

$$\hat{P}^H = \begin{bmatrix}
0.038 & 0.016 & 0.419 & 0.087 & 0.440 \\
0.007 & 0.033 & 0.507 & 0.411 & 0.042 \\
0.033 & 0.117 & 0.715 & 0.108 & 0.027 \\
0.040 & 0.367 & 0.552 & 0.038 & 0.003 \\
0.347 & 0.084 & 0.511 & 0.020 & 0.038
\end{bmatrix}$$

with number of observations in each class given by

$$n^H = (2212, 6646, 32004, 6686, 2227)'$$

The negative dependence over time is obvious. Part of this is ascribed in the literature to the bid-ask bounce, i.e. that stocks are alternatively traded on the bid and ask prices inducing autocorrelation in trade-to-trade returns. This effects disappears if we go just one step further and consider the two-step-ahead transition matrix. Furthermore, the matrix indicates that the conditional variance is larger for row 1 and 5 than for row 3, in line with the stylized fact of volatility clustering. Also, a previous large negative return results in a larger volatility increase than does a previous positive return. This observation is consistent with the leverage effect but is difficult to distinguish as long as autocorrelation is present.
In connection with the results in Section 3 we note that among the models investigated there, the matrix $\hat{P}^H$ seem to correspond best to an AR(1) with negative $\phi$. However, instead of jumping to conclusions about the particular parametric specification of the time dynamics, we feel more comfortable by just saying that we see a clear indication of negative first-order autocorrelation, in line with the bid-ask bounce, see e.g. Campbell et al. (1997). We will not elaborate any further on the connection to parametric models here. Instead, the remainder of the paper will present the results of the windowed data analysis of the IBM data.

### 5.1.2 Hazard function

The estimated hazard functions in the second graph, eventhough rather blurry, show a peak at about 4 seconds for all states. This can be interpreted as that the momentaneous probability for a new transaction to occur is at its highest after about 4 seconds after the last one. After about 20 seconds the hazard function flattens out, implying that the probability of a trade to occur after, say, 30 seconds after the previous transaction is almost as likely as after 60 seconds. The hazard function is not calculated for each time window but only for the entire sample. The fact that durations are not constant over time naturally has its counterpart in that the hazard is neither. However if the probabilities of trade are not constant the hazard can be shown to be decreasing, see appendix A. This is also the case for our data with the exception of the mentioned peak around 4 seconds. Thus, we now have two hypotheses. The first is that there in fact is a “true” peak around 4 seconds, interpreted as a high probability of trades at that particular duration. On the other hand the time varying trade probabilities tell us that the (average) hazard should decrease under the assumption of geometrically distributed durations. One possible explanation from this deviation is that the data is imperfect in the sense that the smallest durations are not correctly logged. It might be difficult to distinguish (in time) between many trades all occurring in the timespace of a few seconds. A second possibility is that the effects are real and that the geometric assumption is invalid. It is however difficult to imagine how the latter would cause a peak in the hazard.

After seeing this one might wonder whether the transition matrices in Section 5.1.1 were seriously affected by the peculiar behaviour of the hazard function for small durations. Therefore the matrices were estimated once more, removing all transactions with duration smaller than 11 seconds. The resulting estimate, shown below, shows that there are only some slight differences.

$$\hat{P}^H = \begin{bmatrix}
0.037 & 0.014 & 0.411 & 0.078 & 0.460 \\
0.005 & 0.036 & 0.490 & 0.426 & 0.043 \\
0.038 & 0.138 & 0.671 & 0.124 & 0.030 \\
0.044 & 0.411 & 0.499 & 0.044 & 0.003 \\
0.403 & 0.083 & 0.453 & 0.020 & 0.041 
\end{bmatrix}$$

The unconditional proportions in the different classes are basically the same as for the entire sample. \(^2\) As for the conditional distribution, the removal of transactions with small duration does not affect the transition matrix much either.

\(^2\)A possible exception from this is that the fraction of visits in class 4 is lower than in the entire sample, 0.127 as opposed to 0.134 to the benefit of visits in classes 3 and 2.
5.2 Results for the windowed sample

5.2.1 Durations

The figures 6-10 show the observed durations and corresponding confidence intervals in the different states, i.e. the time it takes after a transaction in a particular state for the next transaction to occur. The duration after trades in the classes 2, 3 and 4 are rather similar except for $T = 5$ (the first time-window) and $T = 45 – 52$ where the waiting time for class 2 and 4 drops significantly faster than for class 3. The waiting time for classes 1 and 5 are similar. Compared to the waiting time for the other three classes there are, however, significant differences. For $T < 10$ class 1 and 5 show larger waiting time than the other classes do and around $T = 20$ the waiting time is shorter. Another interesting thing to observe here is that the duration after a trade in state 1 and 5, i.e. large negative and positive returns, respectively, peaks at day 39 (January 2nd, 1991), the first trading day of 1991. The duration is slightly larger for the other states at this day too but not as clear cut. A decrease in duration, and thereby an increase in trading activity, occurs during day 19, the day when a deadline was imposed on Iraq.

5.2.2 Unconditional distributions

To start with we look at Figure 13, the unconditional distribution and how it changes over time. As can be seen, it is obviously not stable over time. In the context of risk, we observe that volatility, measured as the probability of trades in the extreme classes 1 and 5, is high in the beginning of the time period. It then decreases to day 12 and is then stable up until day 39, the first day of the new trading year. After that, in the beginning of the year 1991, it increases until day 45, a few days before the deadline imposed by the
Figure 6: The average time of no transactions after a transaction in state 1, i.e. a trade with a large negative return.

Figure 7: The average time of no transactions after a transaction in state 2, i.e. a trade with a moderate negative return.
Figure 8: The average time of no transactions after a transaction in state 3, i.e. a trade with a zero return.

Figure 9: The average time of no transactions after a transaction in state 4, i.e. a trade with a moderate positive return.
Figure 10: The average time of no transactions after a transaction in state 5, i.e. a trade with a large positive return.

Figure 11: The time expected time, given the Markov model, of no transactions after a transaction in state 1, i.e. a trade with a very low return.
5.2.3 Transition probabilities

Another interesting observation is the evolving conditional distribution of returns\(^3\). Here we condition on trades so the transition matrix we consider now is the one based on (5). The general pattern here is that around day 49, the time of the start of the first Gulf war, remarkable things happen (because of the sliding window of length 9 it appears to be 4 days before in the graphs). The conditional probabilities of entering the extreme classes 1 and 5 increase while the conditional probabilities of making moderate changes, the classes 2 and 4, decreases dramatically. At the same time the probability of entering the “no change” class, class 3, decreases around this day. Also here we experimented with removing returns connected with durations less than 11 seconds. The effect was again marginal.

By comparing Figure 14 with Figure 15 we observe that, between day 30 and 40, the right tail of row 1 in the transition matrix is heavier than the left tail of row 5, implying recoils of large price changes to be more pronounced for large negative returns than for large positive. Furthermore, row 1 has less probability mass in the middle of the distribution than row 5 in this period, indicating that trades generating extreme negative returns are less often followed by trades of zero return than trades generating extreme positive returns are, see Figure 16 and Figure 17. As for the other conditional distributions they are fairly

\(^3\)the direct transitions \(P_{t,j}\) are too small for windowing so we consider condition on trades
symmetric with the exception that $P_{14}^T$, see Figure 18, is slightly larger than $P_{52}^T$, see Figure 19, for the first few time windows. These observations are not possible to make by just considering the unconditional distribution of the different states, see with Figure 13.

The results for transitions from class 3 are basically the same as those for the unconditional distribution in Section 5.2.2, see figures 20-24. As for the unconditional probabilities, there is a symmetry in the sense that transitions from class 3 to class 1 is approximately as likely as transitions from class 3 to class 5 etc during the entire time period.

5.2.4 Conditional distributions after waiting times

Conditional probabilities of trades taking waiting times into account, as opposed to ignoring them as was done in Section 5.2.3, are not stable over time either. The conditional probability of a trade taking place significantly increases after day 40, an indication that more information was arriving to the market. This seems reasonable given the expiring deadline in the Persian Gulf. However, compared to ignoring the waiting times, the result is remarkably different. As an example, compare Figure 25 and Figure 26, the increase after day 40 in the probability of a trade in class 3 following a trade in class 2 is reversed to a decrease if we do not account for the waiting time and just treat the series of irregularly spaced returns as if they were regularly spaced. The explanation to this can be seen by noting that since, from day 40 to day 50 the probability of a direct transition from class 2 to 3 is increasing from 1.5% to 2.75%, see equation (37). However, the probability of a trade taking place in any class given that the last trade was in class 2 and was followed by a waiting time is increasing at a faster rate from 2.75% to 6%. The consequence of this is that while $p_{(0|2),3}$ is increasing, $p_{2,3}^H$ is decreasing.
Figure 14: Conditional transition probabilities of a trade in class 5 after a trade in class 1, waiting time ignored.

Figure 15: Conditional transition probabilities of a trade in class 1 after a trade in class 5, waiting time ignored.
Figure 16: Conditional transition probabilities of a trade in class 3 after a trade in class 1, waiting time ignored.

Figure 17: Conditional transition probabilities of a trade in class 3 after a trade in class 5, waiting time ignored.
Figure 18: Conditional transition probabilities of a trade in class 4 after a trade in class 1, waiting time ignored.

Figure 19: Conditional transition probabilities of a trade in class 2 after a trade in class 5, waiting time ignored.
Figure 20: Conditional transition probabilities of a trade in class 1 after a trade in class 3, ignoring waiting times.

Figure 21: Conditional transition probabilities of a trade in class 2 after a trade in class 3, ignoring waiting times.
Figure 22: Conditional transition probabilities of a trade in class 3 after a trade in class 3, ignoring waiting times.

Figure 23: Conditional transition probabilities of a trade in class 4 after a trade in class 3, ignoring waiting times.
Figure 24: Conditional transition probabilities of a trade in class 5 after a trade in class 3, ignoring waiting times.

Figure 25: Conditional transition probabilities of a trade in class 3 after a trade in class 2, waiting time ignored.
Conclusions

We have presented a tool for exploratory analysis of the distributional and temporal properties of intradaily stock returns. The method is based on semi-Markov chains where the distribution of waiting times is supposed to be geometrically distributed with different probability parameter for different states. The products of the method include graphs of the time evolution of the unconditional as well as the conditional distributions of returns. Furthermore, the duration between trades, is implicitly present in the method through artificially constructed states representing waiting times after trades in the different trading states. Connections between the transition probabilities for this full Markov chain and the semi-Markov chain where we have conditioned on actual transaction times has been shown. Empirically, we have studied the intradaily returns and durations of the IBM stock during the period November 1990 until January 1991, a sample of 52145 observations. Also, a study of the results obtained by the method under different data generating processes was made in order to start the buildup of a catalogue of identifiable patterns, potentially useful for model selection. The natural way to extend the method to allow for a more extensive time-dependence would be to allow the elements of the transition matrix to have a time-dynamics of their own.

References


**Appendix A: Derivation of relationship between $P$ and $P^H$**

Consider a (population) contingency table $N$ as in (2) with transition probability matrix $P$ defined by

$$p_{i,j} = \frac{n_{i,j}}{n_{i,.}}, \quad (27)$$

$$p_{i,(0|i)} = \frac{n_{i,(0|i)}}{n_{i,.}}, \quad (28)$$

$$p_{(0|i),j} = \frac{n_{(0|i),j}}{n_{(0|i),.}} \quad (29)$$

and

$$p_{(0|i),(0|i)} = \frac{n_{(0|i),(0|i)}}{n_{(0|i),.}} \quad (30)$$

Since elements of the contingency table for the chain defined by the sequence of states ignoring the delays can be written

$$n^H_{i,j} = n_{i,j} + n_{(0|i),j} \quad (31)$$
the elements of the matrix $P^H$ can be written

$$p_{i,j}^H = \frac{n_{i,j} + n_{(0|i),j}}{\sum_{j=1}^{k}(n_{i,j} + n_{(0|i),j})},$$

(32)

Substituting the elements of (32) by

$$n_{i,j} = p_{i,j}n_{i,.},$$

(33)

$$n_{(0|i),j} = p_{(0|i),j}n_{(0|i),.},$$

(34)

$$n_{i,.} = n\pi_i,$$

(35)

and

$$n_{(0|i),.} = n\pi_{(0|i)}$$

(36)

where $n$ is the total number of observations and the $\pi$’s are the stationary probabilities for the different states. Equation (7) follows. An equivalent way of writing (7) is

$$p_{i,j}^H = p_{i,j} + p_{i,(0|i)}\frac{p_{(0|i),j}}{1 - p_{(0|i),(0|i)}}$$

(37)

Appendix B: The hazard function

Assume that the distribution of the duration, $D_i$, after the $i$’th trade in a particular state, is geometric with parameter $p_i$. The probability function is then

$$f_i(x) = (1 - p_i)x^{x-1}p_i, x = 1, 2, ...$$

Furthermore, we will also need $G_i(x) = P(D_i \geq x)$

$$G_i(x) = (1 - p_i)^{x-1}$$

in order to calculate the hazard at trade $i$. This will be

$$h_i(x) = \frac{f_i(x)}{G_i(x)}$$

Assume further that the durations from half of the trades originates from a geometric distribution with $p = p_1$ and the other half from a geometric distribution with $p = p_2$. Then the hazard for a randomly choosen trade can is

$$h(x) = \frac{f(x)}{G(x)}$$

where

$$f(x) = \frac{1}{2}f_1(x) + \frac{1}{2}f_2(x)$$

and

$$G(x) = \frac{1}{2}G_1(x) + \frac{1}{2}G_2(x)$$

Thus
\[ h(x) = \frac{(1 - p_1)^{x-1}p_1 + (1 - p_2)^{x-1}p_2}{(1 - p_1)^{x-1} + (1 - p_2)^{x-1}} = \frac{p_1}{1 + \left(\frac{1-p_2}{1-p_1}\right)^{x-1}} + \frac{p_2}{1 + \left(\frac{1-p_1}{1-p_2}\right)^{x-1}} \]

Now we see that the hazard starts at the averages of \( p_1 \) and \( p_2 \), i.e. yields

\[ h(1) = \frac{p_1 + p_2}{2} \quad (38) \]

Also, we observe that, when \( x \to \infty \),

\[ h(x) \to p_2 \text{ if } p_1 > p_2 \quad (39) \]

and

\[ h(x) \to p_1 \text{ if } p_1 < p_2 \quad (40) \]

(38),(39) and (40) yields that the (average) hazard is decreasing if \( p_1 \neq p_2 \).