Modeling the coupled return-spread high frequency dynamics of large tick assets

Gianbiagio Curato\textsuperscript{1} and Fabrizio Lillo\textsuperscript{1,2}

\textsuperscript{1} Scuola Normale Superiore, Pisa, Italy
\textsuperscript{2} Dipartimento di Fisica e Chimica, Università di Palermo, Italy
E-mail: gianbiagio.curato@sns.it, gianbiagio83@gmail.com and fabrizio.lillo@sns.it

Received 13 September 2014
Accepted for publication 18 December 2014
Published 21 January 2015

Online at stacks.iop.org/JSTAT/2015/P01028
doi:10.1088/1742-5468/2015/01/P01028

Abstract. Large tick assets, i.e. assets where one tick movement is a significant fraction of the price and bid-ask spread is almost always equal to one tick, display a dynamics in which price changes and spread are strongly coupled. We present an approach based on the hidden Markov model, also known in econometrics as the Markov switching model, for the dynamics of price changes, where the latent Markov process is described by the transitions between spreads. We then use a finite Markov mixture of logit regressions on past squared price changes to describe temporal dependencies in the dynamics of price changes. The model can thus be seen as a double chain Markov model. We show that the model describes the shape of the price change distribution at different time scales, volatility clustering, and the anomalous decrease of kurtosis. We calibrate our models based on Nasdaq stocks and we show that this model reproduces remarkably well the statistical properties of real data.

Keywords: models of financial markets, nonlinear dynamics, stochastic processes

ArXiv ePrint: 1310.4539
1. Introduction

In financial markets, the price of an order cannot assume arbitrary values but it can be placed on a grid of values fixed by the exchange [6,37,38]. The tick size is the smallest interval between two prices, i.e. the grid step, and it is measured in the currency of the asset. It is institutionally mandated and sets a limit on how finely prices may be specified. The grid is evenly spaced for a given asset, and the tick size depends on the price.

In recent years, there has been a growing interest in the role of tick size in determining the statistical properties of returns, spread, limit order book, etc [19,21,24,30,31,42,47,48,53]. The absolute tick size is not the best indicator for understanding and describing the high frequency dynamics of prices. Consider, for example, two highly liquid NASDAQ stocks, namely Apple (AAPL) and Microsoft (MSFT). For both stocks, the tick size is
one cent. However, in the period we investigated in this paper (July and August 2009), the average price of AAPL was $157 while the average price of MSFT was $24. Thus a one cent price movement for AAPL corresponds to 0.6 bp, while for MSFT it is 4.2 bp. MSFT must be regarded as a large tick asset, instead AAPL must be regarded as a small tick asset. Therefore we can expect that the high frequency dynamics of AAPL will be significantly different from that of MSFT. Recent literature has introduced the notion of a perceived tick size to account and quantify the different behavior of returns and spread processes of assets for a given value of tick size [20]. Qualitatively, we say that an asset has a large tick size when the price is averse to variations of the order of a single tick between two transactions and when the bid-ask spread is almost always equal to one tick [24]. Conversely, an asset is small tick size when the price is only weakly averse to variations of the order of a single tick between two transactions and the bid-ask spread can assume a wide range of values, e.g. from one to ten or more ticks [19,53]. In this work we follow the definition given by Dayri et al [20], defining the perceived tick size as the probability that the spread is equal to 1 tick. A probability near to one defines a large tick asset. Several papers in empirical and theoretical market microstructure have emphasized that large and small tick size assets belong to different ‘classes’ [24,46,61]. Order book models designed for small tick assets do not describe correctly the dynamics of large tick assets [30,46]. Moreover, the ultra high frequency statistical regularities of prices and of the order book are quite different in the two classes.

In this paper we are interested in modeling the dynamics of large tick assets at ultra high frequency [1] and taking explicitly into account the discreteness of prices. We propose zero-intelligence models [18] for prices and spread dynamics based on the hidden Markov and double chain Markov processes [3,8,55,63]. Hidden Markov models are widely used in the field of complex systems: condensed matter [39], DNA segmentation [57], econophysics [23], neuro-science [59] and biochemistry [40,44]. In the econometric literature hidden Markov processes [52] are also known as Markov switching models [36], in this work we use the two terms interchangeably. A modeling approach for the limit order book’s dynamics based on Markov or on matrix multiplicative processes was already used in the past [12,56]. However, in this model the Markov process describes the arrivals of market orders, limit orders and order cancellations [37] and not directly price and bid-ask spread dynamics. D’Amico and Petroni [15–17], instead, have proposed a model based on semi-Markov chains for studying directly high frequency financial returns. This model is able to reproduce the autocorrelation of squared returns observed at 1 min frequency. Here, we introduce a class of models describing the coupled dynamics of price changes and spread for large tick assets in transaction time\(^3\). In our models, returns are defined as mid-price changes\(^4\) and are measured in units of half tick, which is the minimum amount the mid-price can change. Therefore, these models are defined in a discrete state space [43,45] and the time evolution is described in discrete time. Our purpose is to model price dynamics in order to reproduce statistical properties of mid-price dynamics at different time scales and stylized facts like volatility clustering [13]. Notice that, rather than considering a non-observable efficient price [6] and describing the data as the effect of the round-off

\(^3\) Hereafter we define the transaction/trade time as an integer counter of events defined by the execution of a market order [37]. Note that if a market order is executed against several limit orders, our clock advances only by one unit.

\(^4\) With a little abuse of language we use returns and mid-price changes interchangeably.

doi:10.1088/1742-5468/2015/01/P01028
Modeling the coupled return-spread high frequency dynamics of large tick assets

Figure 1. Distribution of mid-price change between two transactions, \( r(t, \Delta t = 1) = p_m(t + 1) - p_m(t) \). The investigated stock is Microsoft.

error due to tick size [41], we directly model the observable quantities, such as spread and mid-price, by using a time series approach [34].

The motivation for our work comes from two interesting empirical observations. Let us consider first the unconditional distribution of mid-price changes at different time scales. In figure 1 we show the histogram of the mid-price change of MSFT at the finest transaction time scale, i.e. the change between two transactions. It is clear that most of the times the price does not change, while sometimes it changes by one or two half ticks. When we aggregate the returns on a longer time scale, for example 128 transactions (see figure 2), a non-trivial distribution emerges, namely a distribution where odd values of returns are systematically less populated than even values. It is important to notice that if we assume that returns of individual trades are independent and identically distributed, we would never be able to reproduce a histogram like the one shown in figure 2. In fact in this case the histogram would be, as expected, bell shaped. In a recent paper [62] Zaccaria et al have observed an asymmetry between even and odd values of the spread and, in the context of the order book model of [14], it has been related to the rate of limit orders. As we will show below, this spread asymmetry and the coupling between spread and returns can be used to explain the mid-price change distribution observed in figure 2. In fact, if odd spread values are much more probable than even values, than even price changes are much more probable than odd price changes (see figure 4 below for an intuition).

The second observation concerns the properties of volatility of the return process. Figure 3 shows the autocorrelation function of squared returns of MSFT in transaction time. Square returns can be seen here as a simple proxy of volatility. First of all notice that the autocorrelation is negative for small lags. It then reaches a maximum around 10 trades and then it decays very slowly to zero. We observe that between 10 and more than 500 trades, the decay of the autocorrelation function is well described by a power law function, \( \text{corr} (r^2(t), r^2(t+\tau)) \sim \tau^{-\gamma} \), and the estimated exponent \( \gamma \simeq 0.3 \) is similar to the one observed at lower frequency and by sampling returns in real time rather than transaction time.\(^6\) We conclude therefore that very persistent volatility clustering

\(^5\) For example, if we randomize our sample of transaction to transaction mid-price changes.

\(^6\) It is worth noticing that in general the round-off error severely reduces the correlation properties of a stochastic
and possibly long range volatility is observed also at the transaction to transaction time scale.

The purpose of this paper is to develop a discrete time series model that is able to explain and reproduce simultaneously these two empirical observations, namely the change of the distribution of price changes at different time scales and the shape of the volatility autocorrelation describing its time clustering. The key intuition behind our modeling approach is that for large tick assets the dynamics of mid-price and of spread are intimately related and that the process of returns is conditioned to the spread process, even if the Hurst exponent of a long memory process is preserved [41]. Therefore the autocorrelation function shown in figure 3 is a strong underestimation of the transaction to transaction volatility clustering of the unobservable efficient price.

Figure 2. Mid-price change distribution aggregated at 128 transactions, \( r(t, \Delta t = 128) = p_m(t + 128) - p_m(t) \). The investigated stock is Microsoft.

Figure 3. Sample autocorrelation function of transaction to transaction squared mid-price changes for Microsoft, reported in \( \log_{10} \text{--} \log_{10} \) scales. The plot is in log–log scale and the red dashed line is a best fit of the autocorrelation function in the considered region. The estimated exponent is \( \gamma = 0.301 \). The inset shows the behavior for small values of the lag.
The conditioning rule describes the connection between the stochastic motion of mid-price and spread on the price grid.

More specifically, for large tick assets the spread typically assumes only few values. For example, for MSFT spread size is observed to be 1 or 2 ticks almost always. The discreteness of mid-price dynamics can be connected to the spread dynamics if we observe that, when the spread is constant in time, returns can assume only even values in units of half tick. Instead when the spread changes, returns can display only odd values. Figure 4 shows the constrained relation between the two processes. The dynamics of returns is thus linked to dynamics of spread transitions. This relation leads us to design models in which the return process depends on the transition between two subsequent spread states, distinguishing the case in which the spread remains constant and the case when it changes. From a methodological point of view we obtain this by defining a variable of state that describes the spread transition. We use a hidden Markov, or Markov switching, model [36, 52] for returns, in which the spread transition is described by a Markov chain that defines different regimes for the return process.

The Markov switching approach is able to describe the change in shape of the distribution of mid-price change (figures 1 and 2), but not the persistence of volatility. To this end, we propose a more sophisticated model by allowing the return process to be an autoregressive process in which regressors are the past value of squared returns [3–5, 60]. We show how to calibrate the models on real data and we tested them on the large tick assets MSFT and CSCO, traded at NASDAQ market in the period July–August 2009. We show that the full model reproduces very well the empirical data.

The paper is organized as follows. In section 2 we review the main applications of Markov switching modeling in econometrics. In section 3 we present our modeling approach. In section 4 we present our data for the MSFT stock and we describe the

doi:10.1088/1742-5468/2015/01/P01028
observed stylized facts of price dynamics. In section 5 we describe the calibration of the models on real data and we discuss how well the different models reproduce the stylized facts. Finally, in section 6 we draw some conclusions and we discuss future works.

2. Review of Markov switching models in econometrics

Markov switching models (MS models) have become increasingly popular in econometric studies of industrial production, interest rates, stock prices and unemployment rates [36,55]. They are also known as hidden Markov models (HMM) [7,52,54], used for example in speech recognition and DNA analysis. In these models the distribution that generates an observation depends on the states of an underlying and unobserved Markov process. They are flexible general purpose models for univariate and multivariate time series, especially for discrete-valued series, including categorical variables and series of counts [63]. Markov switching models belong to the general class of finite mixture models [55]. Econometricians’ initial interest in this class of distributions was based on their ability to flexibly approximate general classes of density functions and generate a wider range of values for the skewness and kurtosis than is obtainable by using a single distribution. Along these lines Granger [32] and Clark [9] considered time-independent mixtures of normal distributions as a means of modeling non-normally distributed data. These models, however, did not capture the time dependence in the conditional variance found in many economic time series, as evidenced by the vast literature on ARCH models that started with Engle [25]. By allowing the mixing probabilities to display time dependence, Markov switching models can be seen as a natural generalization of the original time-independent mixture of normals model. Timmermann [58] has shown that the mixing property enables them to generate a wide range of coefficients of skewness, kurtosis and serial correlation even when based on a very small number of underlying states. Regime switches in economic time series can be parsimoniously represented by Markov switching models by letting the mean, variance, and possibly the dynamics of the series depend on the realization of a finite number of discrete states.

The basic MS model is:

\[ y(t) = \mu_{S(t)} + \sigma_{S(t)} \epsilon(t), \tag{1} \]

where \( S(t) = 1, 2, \ldots, k \) denotes the unobserved state indicator which follows an ergodic \( k \)-state Markov process and \( \epsilon(t) \) is a zero-mean random variable which is i.i.d. over time [26]. Another relevant model is the Markov switching autoregressive model (MSAR(\( q \))) of order \( q \) that allows for state-independent autoregressive dynamics:

\[ y(t) = \mu_{S(t)} + \sum_{j=1}^{q} \phi_j (y(t-j) - \mu_{S(t-j)}) + \sigma_{S(t)} \epsilon(t). \tag{2} \]

It became popular in econometrics for analyzing economic time series such as the GDP data through the work of Hamilton [35]. In its most general form the MSAR model allows that the autoregressive coefficients are also affected by \( S(t) \) [58]:

\[ y(t) = \mu_{S(t)} + \sum_{j=1}^{q} \phi_{j,S(t-j)} (y(t-j) - \mu_{S(t-j)}) + \sigma_{S(t)} \epsilon(t). \tag{3} \]
There is a key difference with respect to the ARCH models, which is another type of time-dependent mixture processes. While Markov switching models mix a finite number of states with different mean and volatility parameters based on an exogenous state process, ARCH models mix distributions with volatility parameters drawn from an infinite set of states driven by lagged innovations to the series.

The above described models can be used when the variable $y(t)$ under investigation is continuous. In our case the observed price difference is instead a discrete variable. Therefore the models for continuous variables presented above cannot be used in our problem. We propose to model the coupled dynamics of spreads and price differences in the setting defined by the double chain Markov models (DCMM) [3,4]. This is the natural extension of HMM models in order to allow the hidden Markov process to select one of a finite number of Markov chains to drive the observed process at each time point. If a time series can be decomposed into a finite mixture of Markov chains, then the DCMM can be applied to describe the switching process between these chains. DCMM belongs to the family of Markov chains in random environments [10,11].

In discrete time, DCMM describes the joint dynamics of two random variables: $x(t)$, whose state at time $t$ is unknown for an observer external to the process, and $y(t)$, which is observable. The model is described by the following elements:

- A set of hidden states, $S(x) = \{1, \cdots, N_x\}$.
- A set of possible outputs, $S(y) = \{1, \cdots, N_y\}$.
- The probability distribution of the first hidden state, $\pi_0 = \{\pi_{0,1}, \cdots, \pi_{0,N_x}\}$.
- A transition matrix between hidden states, $M = \{m_{ij}\}, i, j \in S(x)$.
- A set of transition matrices between successive outputs of $y(t)$ given a particular state of $x(t)$, $V_{x(t)=k,ij}, i, j \in S(y), k \in S(x)$.

There are three different estimation problems: the estimation of the probability of a sequence of observations $y(0), \cdots, y(T)$ given a model; the estimation of parameters $\pi_0, M, V_k$ given a sequence of observations; the estimation of the optimal sequence of hidden states given a model and a sequence of outputs.

Our limit order book data instead allow us to see directly the process that defines the hidden Markov process, i.e. the spread process. In this way we can estimate directly the matrices $M$ and $V_k$ by a simple maximum likelihood approach, without using the Expectation Maximization (EM) algorithm and the Viterbi algorithm, that are usually used when the hidden process is not observable [3,4]. We use the stationary probability distribution for the process $x(t)$ as initial probability distribution $\pi_0$ in order to perform our calculations and simulations. We use the DCMM model as a mathematical framework for spread and price differences processes without treating spread process as an hidden process.

Among the few financial applications of the DCMM model we mention [22, 28]. In the former paper, authors studied the credit rating dynamics of a portfolio of financial companies, where the unobserved hidden process is the state of the broader economy. In Eisenkopf [22] instead the author considered a problem in which a credit rating process is
influenced by unobserved hidden risk situations. To the best of our knowledge our paper is the first application of DCMM to the field of market microstructure and high frequency financial data.

3. Markov models for the coupled dynamics of spread and returns

In this section we present the two classes of models that describe the process of returns $r(t, \Delta t) = p_m(t + \Delta t) - p_m(t)$ at time scale $\Delta t$, where we define the mid-price as $p_m(t) = (p_{\text{ASK}}(t) + p_{\text{BID}}(t))/2$ and we choose to measure $r$ in units of half tick size. In our models, the return process follows different time series processes conditioned on the dynamics of transitions of the spread $s(t) = p_{\text{ASK}}(t) - p_{\text{BID}}(t)$. Hereafter we will use the notation $r(t) = r(t, \Delta t = 1)$. The spread variable $s$ is measured in units of 1 tick size, so we have $r(t, \Delta t) \in \mathbb{Z}$ and $s(t) \in \mathbb{N}$. The time variable $t \in \mathbb{N}$ is the transaction time. The first class of models is a Markov switching model and it is able to describe the empirically observed change of the distribution of price change, but it is unable to describe the volatility clustering. The intuition behind this approach relies on the empirical observation that the value of the spread is almost always one tick. This should imply a distribution of price change populated only by even values at each time scale. In order to model the presence of a small population of odd values at low frequencies, we introduce a simple Bernoulli or Markov dynamics for a two-state spread process. The second, more complex class of models, a double chain Markov model, uses the past price dynamics to determine the probabilities and it is able to describe all the investigated features of price.

3.1. Markov switching models

3.1.1. Spread process. It is well known that spread process is autocorrelated in time [6, 33, 50, 51]. We model the spread $s(t)$ as a stationary Markov(1) [49] process:

$$P(s(t) = j | s(t-1) = i, s(t-2) = k, \cdots) = P(s(t) = j | s(t-1) = i) = p_{ij},$$

where $i, j \in \mathbb{N}$ are spread values. As mentioned, we limit the set of spread values to $s \in \{1, 2\}$, because we want to describe the case of large tick assets. We also assume that the process $s(t)$ is not affected by the return process $r(t)$.

The spread process is described by the transition matrix $B \in M_{2,2}(\mathbb{R})$:

$$B = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix},$$

where the normalization is given by $\sum_{j=1}^{2} p_{ij} = 1$. The vector of stationary probabilities is the eigenvector $\pi$ of $B'$ relative to eigenvalue 1, which is:

$$B'\pi = \pi, \quad \pi = \begin{pmatrix} (1 - p_{22}) / (2 - p_{11} - p_{22}) \\ (1 - p_{11}) / (2 - p_{11} - p_{22}) \end{pmatrix},$$

We have tried other specifications of the spread process, such as, for example, a long memory process, but this does not change our results significantly.

This simplifying assumption does not imply that the switching models cannot describe implicitly a conditioning of the spread to returns. It can be shown that our models succeed to reproduce the empirical conditional distributions $P(s(t+1) | r(t))$.

doi:10.1088/1742-5468/2015/01/P01028
where \( B' \) denotes the transpose of the matrix \( B \). This vector represents the unconditional probabilities of \( s( t) \), so \( \pi_k = P( s( t) = k) \) with \( k = 1, 2 \).

Starting from the \( s( t) \) process, it is useful to define a new stationary Markov(1) process \( x( t) \) that describes the stochastic dynamics of transitions between states \( s( t) \) and \( s( t + 1) \) as:

\[
\begin{align*}
  x( t) = 1 & \quad \text{if} \quad s( t) = 1 \rightarrow s( t + 1) = 1, \\
  x( t) = 2 & \quad \text{if} \quad s( t) = 1 \rightarrow s( t + 1) = 2, \\
  x( t) = 3 & \quad \text{if} \quad s( t) = 2 \rightarrow s( t + 1) = 1, \\
  x( t) = 4 & \quad \text{if} \quad s( t) = 2 \rightarrow s( t + 1) = 2. \tag{6}
\end{align*}
\]

This Markov process is characterized by a new transition matrix \( M \in M_{4,4}(\mathbb{R}) \):

\[
M = \begin{pmatrix}
  p_{11} & p_{12} & 0 & 0 \\
  0 & 0 & p_{21} & p_{22} \\
  p_{11} & p_{12} & 0 & 0 \\
  0 & 0 & p_{21} & p_{22}
\end{pmatrix},
\]

where some transitions \( x( t) \rightarrow x( t + 1) \) are forbidden because do not correspond to any allowed spread sequence \( s( t) \rightarrow s( t + 1) \rightarrow s( t + 2) \). The stationary vector of \( M \) is given by:

\[
M' \lambda = \lambda, \quad \lambda = \begin{pmatrix}
  (p_{21}p_{11}) / (1 - p_{11} + p_{21}) \\
  p_{21} (1 - p_{11}) / (1 - p_{11} + p_{21}) \\
  p_{21} (1 - p_{11}) / (1 - p_{11} + p_{21}) \\
  (1 - p_{21})(1 - p_{11}) / (1 - p_{11} + p_{21})
\end{pmatrix}. \tag{7}
\]

A limiting case is when the spread process \( s( t) \) is described by a Bernoulli process. In this case we set \( P( s( t) = 1) = p \). Although \( s( t) \) is an i.i.d. process, the spread transition process \( x_B( t) \) is a Markov process defined by:

\[
M_B = \begin{pmatrix}
  p & (1 - p) & 0 & 0 \\
  0 & 0 & p & (1 - p) \\
  p & (1 - p) & 0 & 0 \\
  0 & 0 & p & (1 - p)
\end{pmatrix}, \quad \lambda_B = \begin{pmatrix}
  p^2 \\
  p (1 - p) \\
  p (1 - p) \\
  (1 - p)^2
\end{pmatrix}.
\]

In the general case, the process \( x( t) \) is defined by two parameters \( p_{11}, p_{21} \) (which are reduced to \( p \) in the Bernoulli case) that we can estimate from spread data.

3.1.2. Mid-price process. We can now define a Markov switching process for returns \( r( t) \) which is conditioned to the process \( x( t) \), i.e. to the spread transitions. Returns are measured in half ticks and we limit the set of possible values to \( r( t) \in \{-2, -1, 0, 1, 2\} \), as observed in our sample (see figure 1). The discreteness of the price grid imposes the mechanical constraints:

\[
\begin{align*}
  x( t) = 1 & \quad \rightarrow r( t) \in \{-2, 0, 2\}, \\
  x( t) = 2 & \quad \rightarrow r( t) \in \{-1, 1\}, \\
  x( t) = 3 & \quad \rightarrow r( t) \in \{-1, 1\}, \\
  x( t) = 4 & \quad \rightarrow r( t) \in \{-2, 0, 2\}. \tag{8}
\end{align*}
\]

The mapping between transitions \( x( t) \) and allowed values of the mid-price changes \( r( t) \) has been done by using the cases shown in figure 4. This assumption is grounded on the
empirical observation that mid-price changes $|r(t)| > 2$ are extremely rare for large tick assets (see section 4).

In the simplest model, we assume that the probability distribution of returns between two transactions depends only on the spread transition between them. We can therefore define the following conditional probabilities defining the process of returns:

\begin{align*}
P(r(t) = ±2|x(t) = 1; θ) &= θ_1, \\
P(r(t) = 0|x(t) = 1; θ) &= 1 - 2θ_1, \\
P(r(t) = ±1|x(t) = 2; θ) &= 1/2, \\
P(r(t) = ±1|x(t) = 3; θ) &= 1/2, \\
P(r(t) = ±2|x(t) = 4; θ) &= θ_4, \\
P(r(t) = 0|x(t) = 4; θ) &= 1 - 2θ_4.
\end{align*}

Notice that we have assumed symmetric distributions for returns between positive and negative values and $θ = (θ_1, θ_4)'$ is the parameter vector that we can estimate from data. The parameter $θ_1 (θ_4)$ describes the probability that mid-price changes when the spread remains constant at one (two) ticks.

The coupled model of spread and return described here will be termed the MS model. When we consider the special case of spread described by a Bernoulli process we will refer to it as the MSB model.

### 3.1.3. Properties of price returns

Here we derive the moments and the autocorrelation functions $ζ(τ) \equiv \text{corr}(r(t), r(t + τ))$ and $ρ(τ) \equiv \text{corr}(r^2(t), r^2(t + τ))$ under the MS model. The quantity $ζ(τ)$ is useful to study the statistical efficiency of prices \[6,13\], while $ρ(τ)$ describes volatility clustering \[13\] in transaction time.

We compute first the vectors of conditional first, second and fourth moments:

\begin{align*}
E[r(t) | x(t) = k] &= m_{1,k}, \\
E[r^2(t) | x(t) = k] &= m_{2,k}, \\
E[r^4(t) | x(t) = k] &= m_{4,k},
\end{align*}

where $m_{j,k}$ indicates the $k$-th component of the vector $m_j$. We have $m_1 = 0$, $m_2 = (8θ_1, 1, 1, 8θ_4)'$ and $m_4 = (32θ_1, 1, 1, 32θ_4)'$. Then we compute unconditional moments by using the stationary vector $λ$ as:

\begin{align*}
E[r(t)] &= \sum_{k=1}^{4} E[r(t) | x(t) = k] P[x(t) = k] = m_1'λ, \\
E[r^2(t)] &= \sum_{k=1}^{4} E[r^2(t) | x(t) = k] P[x(t) = k] = m_2'λ, \\
E[r^4(t)] &= \sum_{k=1}^{4} E[r^4(t) | x(t) = k] P[x(t) = k] = m_4'λ, \\
\text{Var}[r(t)] &= m_2'λ - (m_1'λ)^2, \\
\text{Var}[r^2(t)] &= m_4'λ - (m_2'λ)^2
\end{align*}

In order to compute the linear autocorrelation function $ζ(τ)$ we need to compute $E[r(t) r(t + τ)]$, by using the conditional independence of $r(t)$ with respect to $x(t)$.

\textit{doi:10.1088/1742-5468/2015/01/P01028}
We obtain:

$$E [ r(t) r(t + \tau) ] = \sum_{i=1}^{4} \sum_{j=1}^{4} \left( E [ r(t) r(t + \tau) | x(t) = i, x(t + \tau) = j \right)$$

$$P [ x(t) = i, x(t + \tau) = j ]$$

$$= \sum_{i=1}^{4} \sum_{j=1}^{4} \left( E [ r(t) | x(t) = i ] E [ r(t + \tau) | x(t + \tau) = j ] \right)$$

$$= \sum_{i=1}^{4} \sum_{j=1}^{4} m_{i,j} m_{i,j} \lambda_i M_{ij}^T = X \Lambda M^T m_1,$$ (12)

where the matrix $\Lambda = \text{diag}(m_{1,1}, m_{1,2}, m_{1,3}, m_{1,4})$. The autocorrelation function of returns is given by:

$$\zeta(\tau) = \frac{X \Lambda M^T m_1 - (m_1^T \lambda)^2}{m_2^T \lambda - (m_1^T \lambda)^2},$$ (13)

in our specific case, $\zeta(\tau) = 0$, because symmetry leads to $m_1 = 0$.

Instead, the autocorrelation function of squared returns $\rho(\tau)$ is:

$$\rho(\tau) = \frac{X \Sigma M^T m_2 - (m_2^T \lambda)^2}{m_2^T \lambda - (m_1^T \lambda)^2},$$ (14)

where the matrix $\Sigma = \text{diag}(m_{2,1}, m_{2,2}, m_{2,3}, m_{2,4})$. It is direct to show that $\rho(\tau)$ is an exponential function, $\exp(-a\tau)$, with $a = -\ln(p_{11} - p_{21})$.

As expected, both correlation functions depend on powers of the transition probability matrix $M$. For a Markov process, $M$ is diagonalizable and we can write $M^T = C M_B^T C^{-1}$, where:

$$M_B^T = \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & (p_{11} - p_{21})^T
\end{pmatrix},$$

$$C = \begin{pmatrix}
1 & 0 & 1 & 1 \\
\frac{p_{11}}{(p_{11} - 1)} & 0 & 1 & \frac{p_{21}}{(p_{11} - 1)} \\
0 & 1 & 1 & \frac{p_{21}}{(p_{11} - 1)} \\
0 & \frac{p_{21}}{p_{21} - 1} & \frac{p_{21}}{p_{11} - 1}
\end{pmatrix}.$$  

In the limit case in which the spread is described by a Bernoulli process, the matrix $M_B$ is not diagonalizable but has all eigenvalues in $\mathbb{R}$, i.e. $\text{sp}(M_B) = (0, 0, 0, 1)$, and we can compute its Jordan canonical form $J_B$. Thus we can rewrite the lag dependence as $M_B^T = E J_B E^{-1}$, where:

$$J_B = \begin{pmatrix}
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}, \quad E = \begin{pmatrix}
(p - p^2) & (1 - p^2) & p^2 & 0 \\
-p^2 & -p^2 & p^2 & 0 \\
(p - p^2) & -p^2 & p^2 & \frac{p-1}{p} \\
-p^2 & -p^2 & p^2 & 1
\end{pmatrix}.$$  

The structure of the block diagonal matrix $J_B$ implies that $J_B^T = J_B^T = 0$, $\forall \tau \geq 2$ and that $\rho(\tau)$ is a constant function for $\tau \geq 2$.  

\[\text{doi:10.1088/1742-5468/2015/01/P01028}\]
Modeling the coupled return-spread high frequency dynamics of large tick assets

3.1.4. Discussion. The qualitative comparison of real data and model shows that the MS model is able to reproduce the distribution of returns quite well. This can be seen by comparing figures 1 and 2 with figures 5 and 6. It is worth noting that, at least qualitatively, the Bernoulli model MS$_B$ is also able to reproduce the underestimation of odd values of returns with respect to the even values, as observed in real data. Therefore it is the coupling of spread and return described by equation (9), rather than the memory properties of spread, which is responsible of the behavior of the aggregated return distribution of figure 2. The fact that returns are uncorrelated is a consequence of our choice of symmetric conditional returns, i.e. $m_1 = 0$. Note, however, that the symmetry of unconditional returns is not enough to guarantee lack of linear correlations. It is in fact possible to generalize the model of equation (9) in such a way that conditional returns are not symmetric, but unconditional returns are symmetric, with exponentially...
decaying autocorrelation function of returns. However the model fails to describe the volatility clustering. In fact, we have seen that the model has an exponentially decaying $\rho(\tau)$. Moreover, as the data calibration shows (see figure 7), the predicted $\rho(\tau)$ under the MS model is much smaller than the one observed in real data. Therefore this model is unable to reproduce the volatility clustering as well as any long memory property. This observation motivates us to develop a model that, preserving the structure of the coupling between spread and returns discussed so far, is able to describe non-exponential volatility clustering. This model is developed in the next section.

### 3.2. A double chain Markov model with logit regression

The Markov switching model is not able to explain the empirically observed correlation of squared returns shown in figure 3. Therefore in the second class of models we consider an autoregressive switching model for returns [27, 55] in order to study the correlation of squared returns. The idea is to use logit regressions on past values of returns and squared returns. The model is thus defined by the following conditional probabilities:

$$P(r(t) | x(t) = k, \Omega(t - 1); \theta_k), \ k \in \{1, 2, 3, 4\};$$

$$\Omega'(t - 1) = \left(r^2(t - 1), ..., r^2(t - p), \right)$$

$$r(t - 1), ..., r(t - e) = (\Omega'_{r}, \Omega'_{e});$$

$$\theta'_k = (\alpha_k, \beta'_k, \gamma'_k), \quad (15)$$

where we define an informative $(p + e)$-dimensional vector of regressors $\Omega$, made of the past $e$ returns and $p$ squared returns. Each parameter vector $\theta_k$ is composed by the scalar $\alpha_k$, the $p$-dimensional vector $\beta_k$ which describes the regression on past values of squared returns, and the $e$-dimensional vector $\gamma_k$ which describes the regression on past returns.

In order to handle the discreteness of returns we make use of a logit regression.

**Figure 7.** Autocorrelation function of squared returns $\rho(\tau)$, reported in log$_{10}$-linear scales. The black circles are the real data of MSFT asset. The red squares are the result of the MS$_B$ model, the green diamonds refer to the MS model, the blu up triangles refer to the DCMM(1) model and the pink down triangles refer to the DCMM(3) model, all calibrated on the MSFT asset.
To this end, we first convert the returns series in a binary series $b(t) \in \{0, 1\}$. When the spread remains constant between $t$ and $t+1$ (i.e. $x(t) = 1$ or $x(t) = 4$), we set:

\[ r(t) = \pm 2 \rightarrow b(t) = 1, \]
\[ r(t) = 0 \rightarrow b(t) = 0, \]

while when the spread changes, (i.e. $x(t) = 2$ or $x(t) = 3$) we set:

\[ r(t) = 1 \rightarrow b(t) = 1, \]
\[ r(t) = -1 \rightarrow b(t) = 0. \]

Then, denoting by $\eta_k(t)$ the probability of having $b(t) = 1$ conditional on $x(t) = k$, the logit regression postulates that

\[ \eta_k(t) = \frac{\exp(\alpha_k + \Omega_{x,2} (t - 1) \beta_k + \Omega_{x} (t - 1) \gamma_k)}{1 + \exp(\alpha_k + \Omega_{x,2} (t - 1) \beta_k + \Omega_{x} (t - 1) \gamma_k)}. \]

Therefore we obtain for the process $^9 r(t)$

\[
\begin{align*}
P (r(t) = \pm 2 | x(t) = 1, \Omega (t - 1) ; \theta_1) &= \eta_1(t) / 2, \\
P (r(t) = 0 | x(t) = 1, \Omega (t - 1) ; \theta_1) &= 1 - \eta_1(t) ; \\
P (r(t) = 1 | x(t) = 2, \Omega (t - 1) ; \theta_3) &= \eta_2(t), \\
P (r(t) = -1 | x(t) = 2, \Omega (t - 1) ; \theta_3) &= 1 - \eta_2(t) ; \\
P (r(t) = 1 | x(t) = 3, \Omega (t - 1) ; \theta_3) &= \eta_3(t), \\
P (r(t) = -1 | x(t) = 3, \Omega (t - 1) ; \theta_3) &= 1 - \eta_3(t) ; \\
P (r(t) = \pm 2 | x(t) = 4, \Omega (t - 1) ; \theta_3) &= \eta_4(t) / 2, \\
P (r(t) = 0 | x(t) = 4, \Omega (t - 1) ; \theta_3) &= 1 - \eta_4(t). 
\end{align*}
\]

The intuition of the first two equations is the following. When between $t$ and $t+1$ the spread remains equal to one tick ($x(t) = 1$, see left panel of figure 4), the probability that the midprice moves of two half ticks or remains the same depends on $\eta_1(t)$. This is a logistic function of a linear combination of the past $e$ returns and, more importantly, of past $p$ squared returns. Thus high volatile periods generate high volatility and for this reason the model is able to generate clustered volatility. Similar considerations hold for the other equations. Thus our model translates in a discrete logistic framework the self-excitation mechanism of volatility.

These equations define the general DCMM($e, p$) model, because we can define a Markov process for $r(t)$. In the rest of the paper we will consider the case $e = 0$ and for the sake of simplicity we will denote DCMM($p$) = DCMM(0, $p$). In our case the independent latent Markov process is represented by the transition process $x(t)$ and the dependent Markov process is represented by the $r(t)$ processes. For the sake of clarity, here we consider the case $p = 1$, while its extension to a general value for $p$ is considered in appendix B.

We define conditional Markov processes for $r(t)$ and $r^2(t)$ by means of the logit probabilities $\eta_k(t)$. The definition of the process for $r(t) \in \{-2, -1, 0, 1, 2\}$, and $i, j \in \{1, 2, 3, 4, 5\}$, in the case of $p = 1$ (DCMM($1$)) is the following:

\[ P (r(t) = 3 - j | x(t) = k, r(t - 1) = 3 - i ; \theta_k) = A_{k,ij}. \]

We have four possible transition matrices $A_{x(t)=k}$ for $k \in \{1, 2, 3, 4\}$, determined by the latent process $x(t)$. Their form is given in appendix A.

---

^9 Note that for the case $x(t) = 1, 4$ we have assumed equal probability for $r(t) = \pm 2$. 

doi:10.1088/1742-5468/2015/01/P01028
Modeling the coupled return-spread high frequency dynamics of large tick assets

Assuming that the latent process has reached the stationary distribution defined by equation (7), we can define an overall Markov chain by the transition matrix \( N \) that describes the \( r(t) \) process:

\[
N = \sum_{k=1}^{4} \lambda_k A_k. \tag{20}
\]

The matrix \( N \) is defined by 6 + 4\( p \) parameters: \( p_{11}, p_{21}, \alpha_k, \beta_k' \). Analogously the probabilities for the process \( r^2(t) \in \{0, 1, 4\}, \) and \( i, j \in \{1, 2, 3\} \), in the case of \( p = 1 \) (DCMM(1)) is

\[
P \left( r^2(t) = (3 - j)^2 \mid x(t) = k, r^2(t - 1) = (3 - i)^2 ; \theta_k \right) = V_{k,ij}.
\]

which can be calculated from the matrix \( A \) (see appendix A).

We can again define an overall Markov process for \( r^2(t) \) described by a transition matrix \( S \), assuming that the transition process \( x(t) \) has reached the stationary distribution:

\[
S = \sum_{k=1}^{4} \lambda_k V_k. \tag{21}
\]

The matrix \( S \) is defined by 4 + 2\( p \) parameters: \( p_{11}, p_{21}, \alpha_k, \beta_k' \), where \( k \in \{1, 4\} \), because \( V_{k=2,3} \) do not depend by the past values of \( r^2(t) \). In fact, the squared return is a constant when \( x(t) = 2, 3 \), i.e. from equation (19) we have \( r^2(t) = 1, \forall t \).

The function \( \text{corr} \left( r^2(t), r^2(t + \tau) \right) = \rho(\tau) \) for the DCMM(1) process is the correlation of the Markov(1) process defined by \( S \). We solve the eigenvalue equation for \( S \) relative to the eigenvalue 1 in order to determine the stationary probability vector \( \psi \):

\[
S' \psi = \psi. \tag{22}
\]

If we define the vectors \( \delta, \delta_2 \) and \( \xi \), where \( \delta_i = (3 - i)^2, \delta_{2,i} = (3 - i)^4 \) and \( \xi = \delta \odot \psi \), the moments are given by:

\[
\begin{align*}
E \left[ r^2(t) \right] &= \delta' \psi, \\
E \left[ r^4(t) \right] &= \delta_2' \psi, \\
E \left[ r^2(t) r^2(t + \tau) \right] &= \xi' S^\tau \delta. \tag{23}
\end{align*}
\]

Finally, we have the expression for \( \rho(\tau) \) in the case \( p = 1 \):

\[
\rho(\tau) = \frac{\xi' S^\tau \delta - (\delta' \psi)^2}{\delta_2' \psi - (\delta' \psi)^2}. \tag{24}
\]

The generalization of the calculation of \( \rho(\tau) \) to any value of the order \( p \) is reported in the appendix B.

3.2.1. Estimation. In order to estimate the parameter vector \( \theta' = (\theta'_1, \theta'_2, \theta'_3, \theta'_4) \) we maximize the following partial-loglikelihood:

\[
\mathcal{L}(\theta) = \sum_{t=p+1}^{T} \log \left[ \sum_{k=1}^{4} P \left( x(t) = k \mid \Omega(t - 1) ; \theta_k \right) P \left( b(t) \mid x(t) = k, \Omega(t - 1) ; \theta_k \right) \right],
\]

\[
\text{doi:10.1088/1742-5468/2015/01/P01028}
\]

16
where $T$ is the length of the sample, and we assume that parameters $p_{11}$ and $p_{21}$ are known. Since the dynamics of spread transitions is independent from the past informative set, i.e. $P(x(t) = k | \Omega(t-1); \theta_k) = P(x(t) = k)$, we have:

$$\mathcal{L}(\theta) = \sum_{t=p+1}^{T} \log \left[ \sum_{k=1}^{4} P(x(t) = k) P(b(t) | x(t) = k, \Omega(t-1); \theta_k) \right].$$

(26)

In the case of large tick assets, it is $\lambda_1 \approx 1$ and we can use the approximation

$$\mathcal{L}(\theta) \approx \sum_{t=p+1}^{T} \log \left( P(b(t) | x(t) = 1, \Omega(t-1); \theta_1) \right).$$

(27)

For example for MSFT we have $\lambda_1 \approx 0.9$. With this approximation we estimate only the vector $\theta_1$ and the parameter $\theta_4$ of equation (9), that are enough in order to define matrices $V_k$. Moreover, we have the following approximation:

$$V_{x(t)=4} \approx \left( \begin{array}{ccc} 2\theta_4 & 0 & 1 - 2\theta_4 \\ 2\theta_4 & 0 & 1 - 2\theta_4 \\ 2\theta_4 & 0 & 1 - 2\theta_4 \end{array} \right).$$

In this way we neglect the contribution of regressors $\Omega(t-1)$ (weighted by $\beta_4$) and make use of the simpler expression in equation (9) when $x(t) = 4$. As before, this approximation holds if the weight of $V_{x(t)=4}$ is negligible, i.e. $\lambda_4 \approx 0$, i.e. when there is a small number of spread transitions $s(t) = 2 \rightarrow s(t+1) = 2$. This is the case when we have large tick assets, where we have almost always $s(t) = 1$. In the case of MSFT asset for example we have $\lambda_4 \approx 0.04$.

We have performed the calculation of the autocorrelation $\rho(\tau)$ of the squared returns for $p = 1, 3$ and the result is reported in figure 7. We have calibrated the parameters on the MSFT asset (see next sections for details). We note that the MS and MSB models underestimate very strongly $\rho(\tau)$. Note that for the MS model, $\rho(\tau)$ calibrated on real data is very small but not zero as predicted by the theory. The DCMM(p) model, on the other hand, is able to fit very well $\rho(\tau)$ up to lag $\tau = p$. Remarkably the model captures very well also the negative correlation for very short lags. However this observation indicates that an higher order DCMM(p) model might be able to fit better the real data. In the next sections we will show that this is indeed the case.

4. Data

We have investigated two stocks, namely Microsoft (MSFT) and Cisco (CSCO), both traded at NASDAQ market in the period July–August 2009, corresponding to 42 trading days. Data contain time stamps corresponding to order executions, prices, size of trading volume and direction of trading. The time resolution is 1 ms. In this article we report mostly the results for MSFT asset, which are very similar to those for CSCO.

Non-stationarities can be very important when investigating intraday financial data. For this reason and in order to restrict our empirical analysis to roughly stationary time intervals, we first compute the intensity of trading activity at time $t$ conditional to a specific value $k$ of mid-price change, i.e. $p(t|r(t) = k)$. As we can see from figure 8, the unconditional trading intensity $p(t)$ is not stationary during the day [2]. Trading activity
Modeling the coupled return-spread high frequency dynamics of large tick assets

Figure 8. Unconditional and conditional probability distributions describing the number of transaction events per unit time, reported in linear-log scales. We bin recorded time events for MSFT into 6 min intervals.

Table 1. Summary statistics for assets MSFT and CSCO for the two subsamples of high and low trading activity. $\sigma$ is the standard deviation, ex. ku. is the excess kurtosis of transaction to transaction returns and $\hat{\pi}_1$ is the fraction of time the spread is equal to one tick. The values of the mean are not significantly different from zero. A one sample t-test for the mean of returns, performed on 42 trading days, does not reject the null-hypothesis of a zero value for the mean.

| Asset | Activity | # Trades | Mean (ticks/2) | $\sigma$ | ex. ku. | $\hat{\pi}_1$ |
|-------|----------|----------|----------------|----------|---------|--------------|
| MSFT  | High     | 184542   | $-2.82 \times 10^{-4}$ | 0.652    | 5.13    | 0.92         |
|       | Low      | 348253   | $8.96 \times 10^{-4}$   | 0.514    | 9.89    | 0.95         |
| CSCO  | High     | 145084   | $-1.32 \times 10^{-3}$ | 0.673    | 4.73    | 0.92         |
|       | Low      | 275879   | $1.44 \times 10^{-3}$   | 0.551    | 8.46    | 0.95         |

is very high at the beginning and at the end of the day and, as usually done to remove the effect of opening and closing auction, we discard transaction data in the first and last six minutes of trading day. Moreover the figure shows that the relative frequencies of the three values of returns change during the day, except for returns larger than two ticks that are very rare throughout the day. Most important, in the first part of the day, one tick or two tick returns are more frequent than zero returns, while after approximately 10:30 the opposite is true. For this reason we split our times series in two subsamples. The first sample, corresponding to a period of high trading intensity, covers the time sets $t \in (9:36, 10:30) \cup (15:45, 15:54)$, where time is measured in hours. The second sample, corresponding to low trading intensity, covers the time set $t \in [10:30, 15:45]$. Table 1 reports a summary statistics of the two subsamples.

We then analyze the empirical autocorrelation function of squared returns $\text{corr}(r^2(t), r^2(t+\tau)) = \rho(\tau)$ for these two series. As we can see from figure 9, for $\tau > 5$ transactions both time series display a significant positive and slowly decaying autocorrelation, which is a quantitative manifestation of volatility clustering. The series
corresponding to low trading activity displays smaller, yet very persistent, volatility clustering.

5. Estimation details and comparison with real data

We have estimated the models described in sections 3.1 and 3.2 and we have used Monte Carlo simulations to generate artificial time series calibrated on real data. The statistical properties of these time series have been compared with those from real data.

More specifically, we have considered three models: (i) the MS$_B$ model, where spread is described by a Bernoulli process and there are no logit regressors; (ii) the MS model, where spread is a Markov(1) process and there are no logit regressors; (iii) the DCMM($p$) model, where spread is a Markov(1) process and the set of logit regressors includes only the past $p$ values of squared returns. Notice therefore, that in this last model we set $e = 0$. Finally, we have estimated the model separately for high and low activity regime.

5.1. Estimation of the models

From spread and returns data we computed the MLE estimators $\hat{\pi}_1$, $\hat{p}_{11}$, $\hat{p}_{21}$, $\hat{\theta}_1$, $\hat{\theta}_4$ of the parameters defined in section 3.1. They are given by

$$\hat{\pi}_1 = \frac{n_1}{N_s},$$
$$\hat{p}_{ij} = \frac{n_{ij}}{\sum_{j=1}^2 n_{ij}},$$
$$\hat{\theta}_k = \frac{1}{2} \left( 1 - \frac{n_{0k}}{N_k} \right),$$

(28)

where $n_1$ is the number of times $s(t) = 1$, $N_s$ is the length of the spread time series, $n_{ij}$ is the number of times the value of spread $i$ is followed by the value $j$, $n_{0k}$ is the number...
of times returns are zero in the regime $x(t) = k$, and $N_k$ is the length of the subseries of returns in the same regime. For the last estimator $\hat{\theta}_k$ we count only zero returns because we assumed that the returns are distributed symmetrically in the set $(-2, 0, 2)$. We have checked that this assumption represents a good approximation for our data sets. The estimated parameters for the MSFT asset are shown in Table 2.

In order to estimate the DCMM($p$) model we need to estimate the vector $\theta$. For both regimes we use the approximated log-likelihood of equation (27) because we have for low volatility series $P(x(t) = 1) \approx 0.92$ and for high volatility $P(x(t) = 1) \approx 0.87$. Thus we need to estimate only the vector $\theta_1 = (\alpha_1, \beta_1')$ by a standard generalized linear regression and we use an iterative reweighted least squares technique [29]. In this way, we generate the returns series in regime $x(t) = 1$, but for the other regimes the generator follows the rules in equation (9), i.e. we use the estimator $\hat{\theta}_4$. The order of the model is fixed to $p = 50$ in order to investigate the impact of past squared returns on the returns process. For simplicity, we report here only the results from high activity time series.

We find $\alpha_1 = -2.921(0.019)$ and we report the first 25 values of $\beta_{1i}$ in Table 3. The estimates of $\beta_{1i}$ are significantly positive for $i > 2$ up to $i = 50$, with the exception of $i = 36, 37$. Moreover, they display a maximum for $i = 6$. We perform a power law fit on these parameters, $\beta_{1i} \propto i^{-\alpha}$, and we find a significant exponent $\alpha = 0.626(0.068)$. We hypothesize that this functional dependence of $\beta_{1i}$ from $i$ could be connected to the slow decay of the autocorrelation function of squared returns, but we have not investigated further this aspect.

### 5.2. Comparison with real data

After having estimated the three models on the real data, we have generated for each model 25 data samples of length $10^6$ observations. In this way, we are able to determine an empirical statistical error on quantities that we measure on these artificial samples. We have considered three quantities to be compared with real data. Beside the autocorrelation of squared returns, in order to analyze the return distribution at different transaction time scales $\Delta t$, we have measured the empirical standard deviation and excess kurtosis:

$$\sigma(\Delta t) = \left( E \left[ \left( p_m(t + \Delta t) - p_m(t) \right) - E[p_m(t + \Delta t) - p_m(t)] \right]^2 \right)^{1/2},$$

(29)

$$\kappa(\Delta t) = \frac{E \left[ \left( p_m(t + \Delta t) - p_m(t) \right) - E[p_m(t + \Delta t) - p_m(t)] \right]^4}{\sigma^4(\Delta t)} - 3.$$  

(30)

doi:10.1088/1742-5468/2015/01/P01028

---

**Table 2.** Estimated parameters for the two subsamples of high and low trading activity for the MSFT asset.

| Par. | High     | Low      |
|------|----------|----------|
| $\hat{\pi}_1$ | $9.17 \times 10^{-1}$ | $9.52 \times 10^{-1}$ |
| $\hat{p}_{11}$ | $9.53 \times 10^{-1}$ | $9.72 \times 10^{-1}$ |
| $\hat{p}_{21}$ | $5.22 \times 10^{-1}$ | $5.50 \times 10^{-1}$ |
| $\hat{\theta}_1$ | $4.81 \times 10^{-2}$ | $2.85 \times 10^{-2}$ |
| $\hat{\theta}_4$ | $1.51 \times 10^{-3}$ | $2.65 \times 10^{-4}$ |
The normalized standard deviation $\sigma_N(\Delta t) = \sigma(\Delta t)/\sqrt{\Delta t}$ gives information on the diffusive character of the price process, because $\sigma_N(\Delta t)$ is constant for diffusion. The behavior of $\kappa(\Delta t)$ as a function of $\Delta t$, instead, gives information about the presence or not of nonlinear long range correlations between returns [6]. We first investigate the autocorrelation properties of squared returns $\rho(\tau)$. This function is compatible with zero for MSB and MS models except for the first lag where we have measured a significant positive value, e.g. $\rho(\tau = 1) \approx 0.01$ in the case of MSB for MSFT. The model with regressors DCMM(50), instead, is able to reproduce remarkably well the values of $\rho(\tau)$ up to $\tau = 50$, as we can see from figures 10 and 11, both for MSFT and for CSCO. The behavior of $\rho(\tau)$ around $\tau \approx 0$ is also very well reproduced by the model. The model underestimates the values of the autocorrelation of the real process for $\tau > 50$ but it generates values that are still significantly positive. We have performed a power law fit, i.e. $\rho(\tau) \propto \tau^{-\alpha}$, on real and DCMM(50) simulated data for values of lags corresponding to $\tau \in [6,50]$. For real data we found $\alpha = 0.298(0.023)$ and for simulated data $\alpha = 0.300(0.028)$. Since $\alpha < 1$ this model is able to reproduce long
Figure 10. Empirical autocorrelation functions $\text{corr} \left( r^2(t), r^2(t + \tau) \right)$ for real (black circles) and simulated (blue diamonds) data according to DCMM(50) model. The red line is a power law fit on the real data. The panel refers to the model fitted on MSFT data for high volatility series.

Figure 11. Empirical autocorrelation functions $\text{corr} \left( r^2(t), r^2(t + \tau) \right)$ for real (black circles) and simulated (blue diamonds) data according to the DCMM(50) model. The red line is a power law fit on the real data. The panel refers to the model fitted on CSCO data for high volatility series.

memory shape of correlation $\rho(\tau)$ for a number of values of lags $\tau$ equal to the order of model $p$.

We then analyzed the distributional properties, i.e. normalized standard deviation $\sigma_N(\Delta t)$ and excess kurtosis $\kappa(\Delta t)$. For each value of $\Delta t$ and for each model we calculate the average and standard deviation of the 25 simulations and we compare the simulation results with real data (see figures 12 and 13).

The three models are clearly diffusive. This is a consequence of our modeling assumption, i.e. uncorrelatedness of returns. The most important thing to observe is that MS and DCMM(50) models reproduce the empirical values of $\sigma_N$ better than the MS$_B$ model. The difference between MS and DCMM(50) models are appreciable only for $\Delta t > 128$, i.e. this parameter is almost the same for these two models. The behavior of
Figure 12. Rescaled volatility, reported in log\textsubscript{10}-linear scales, \(\sigma_N(\Delta t)\) of aggregated returns on time scale \(\Delta t\) for MS\textsubscript{B} (red line), MS (green line), and DCMM\((p = 50)\) (blue line), compared with the same quantity for MSFT data for high volatility series (black line). Error bars are the standard deviation obtained from 25 Monte Carlo simulations of the corresponding models.

Figure 13. Excess kurtosis, reported in log\textsubscript{10}-log\textsubscript{10} scales, \(\kappa(\Delta t)\) of aggregated returns on time scale \(\Delta t\) for MS\textsubscript{B} (red line), MS (green line), DCMM\((p = 50)\) (blue line), compared with the same quantity for MSFT data for high volatility series (black line). Error bars are the standard deviation obtained from 25 Monte Carlo simulations of the corresponding models.

Excess kurtosis, instead, is different between the models (see figure 13). The excess kurtosis for MS\textsubscript{B} and MS models is well fit by a power law \(\kappa(\Delta t) \sim \Delta^{-\alpha}\) with \(\alpha = 0.901(0.027)\) (MS\textsubscript{B}) and \(\alpha = 0.997(0.052)\) (MS). These values are consistent with a short range correlation of volatility. In fact, it can be shown [6] that stochastic volatility models with short range autocorrelated volatility are characterized by \(\alpha = 1\). On the contrary, stochastic volatility models with long range autocorrelated volatility display a slower decay. This is exactly what is observed for real data and for the DCMM\((50)\) model.
both cases we observe an anomalous scaling of kurtosis that is more compatible with a stochastic volatility model in which volatility is a long memory process.

6. Conclusions

We have developed Markov switching models for describing the coupled dynamics of spread and returns of large tick assets in the transaction time. The underlying Markov process is the process of transitions between consecutive spread values. In this way returns are described by different processes depending on whether the spread is constant or not in time. We have shown that this mechanism is needed in order to model the different shapes of the distribution of mid-price changes at different aggregations in the number of trades, i.e. different time scales. In order to be able to model the persistent volatility clustering, we have introduced a Markov model with logit regressors represented by past values of returns and squared returns.

We have calibrated the model on NASDAQ stocks and we have found that the model reproduces remarkably well and in a quantitative way the empirical stylized facts. In particular we are able to reproduce the shape of the distribution at different time scales, uncorrelated returns, diffusivity, slowly decaying autocorrelation function of squared returns, and anomalous decay of kurtosis on different time scales.

As a possible extension, we observe that, if we want to reproduce more precisely the autocorrelation function of squared returns up to a certain number of lags, we need to estimate a number of parameters, i.e. order of model, at least equal to this value. We find that these parameters scale with a power law function of the parameter’s index, i.e. it is a function of the number of past lags at which regressors are defined. A possible improvement of this model could be to develop a model in which we estimate directly a parametric function with a small number of parameters (for example a power law function) that can describe how these parameters scale when we consider a certain order for the model.

A modeling approach based on an independent Markov process for the spread dynamics is only a simplified framework. Future research will be needed in order to investigate much sophisticated models that can allow long memory spread processes and a direct impact of returns dynamics on the spread dynamics. This could be useful in order to extend our modeling approach to a broader class of assets like small tick assets.

Acknowledgments

We would like to thank A Profeti and A Carlo Giuseppe Mennucci for the development and support of the computer facility HAF922.sns used to perform data analysis and Monte Carlo simulations, written in R language, reported in this article. Authors acknowledge partial support by the grant SNS11LILLB ‘Price formation, agents het- erogeneity, and market efficiency’. We thank the two anonymous referees for the recommendations to improve the paper.

doi:10.1088/1742-5468/2015/01/P01028
Appendix A. Matrices of the DCMM(1) model

There are four possible transition matrices \( A_{x(t)=k} \) for \( k \in \{1, 2, 3, 4\} \), determined by the latent process \( x(t) \):

\[
A_{x(t)=1} = \begin{pmatrix}
\eta_1 (r^2(t - 1) = 4) / 2 & 0 & 1 - \eta_1 (r^2(t - 1) = 4) & 0 & \eta_1 (r^2(t - 1) = 4) / 2 \\
\eta_1 (r^2(t - 1) = 1) / 2 & 0 & 1 - \eta_1 (r^2(t - 1) = 1) & 0 & \eta_1 (r^2(t - 1) = 1) / 2 \\
\eta_1 (r^2(t - 1) = 0) / 2 & 0 & 1 - \eta_1 (r^2(t - 1) = 0) & 0 & \eta_1 (r^2(t - 1) = 0) / 2 \\
\eta_1 (r^2(t - 1) = 1) / 2 & 0 & 1 - \eta_1 (r^2(t - 1) = 1) & 0 & \eta_1 (r^2(t - 1) = 1) / 2 \\
\eta_1 (r^2(t - 1) = 4) / 2 & 0 & 1 - \eta_1 (r^2(t - 1) = 4) & 0 & \eta_1 (r^2(t - 1) = 4) / 2 \\
\end{pmatrix},
\]

where the temporal dependence is given only by the values of past squared returns, i.e. \( r^2(t - 1) \), 1-step before the present time \( t \). The others two matrices have the same definitions: \( A_{i} = A_{1} (\eta_{i} \rightarrow \eta_{4}) \) and \( A_{3} = A_{2} (\eta_{2} \rightarrow \eta_{3}) \). Here we are simply using equation (19) remembering that, when \( k \) is chosen, some states are forbidden.

In the case of \( r^2(t) \), there are four possible transition matrices \( V_{x(t)=k} \) for \( k \in \{1, 2, 3, 4\} \), determined by the latent process \( x(t) \):

\[
V_{x(t)=1} = \begin{pmatrix}
\eta_1 (r^2(t - 1) = 4) & 0 & 1 - \eta_1 (r^2(t - 1) = 4) & 0 \\
\eta_1 (r^2(t - 1) = 1) & 0 & 1 - \eta_1 (r^2(t - 1) = 1) & 0 \\
\eta_1 (r^2(t - 1) = 0) & 0 & 1 - \eta_1 (r^2(t - 1) = 0) & 0 \\
\eta_1 (r^2(t - 1) = 1) & 0 & 1 - \eta_1 (r^2(t - 1) = 1) & 0 \\
\eta_1 (r^2(t - 1) = 4) & 0 & 1 - \eta_1 (r^2(t - 1) = 4) & 0 \\
\end{pmatrix},
\]

The others two matrices have the same definitions: \( V_{4} = V_{1} (\eta_{1} \rightarrow \eta_{4}) \) and \( V_{3} = V_{2} (\eta_{2} \rightarrow \eta_{3}) \).

Appendix B. Correlation of squared returns for the DCMM(p) model

The definition of the Markov process for \( r^2(t) \in \{0, 1, 4\} \) in the case of a general value of \( p \) for the DCMM model follows from equation (19). This stochastic process is a stationary Markov process of order \( p \) for each value of \( k \) [63]:

\[
P \left( r^2(t) = (3 - i_{p+1})^2 \mid x(t) = k; r^2(t - 1) = (3 - i_{p})^2, \ldots, r^2(t - p) = (3 - i_{1})^2; \theta_{k} \right) = V_{x(t):i_{1}i_{2}\ldots i_{p+1}}(t),
\]

where we have \( k \in \{1, 2, 3, 4\} \) and a \( p+1 \)-dimensional vector of indices \( \hat{i} = (i_{1}, i_{2}, \ldots, i_{p+1}) \), where each index can assume values \( i_{l} \in \{1, 2, 3\} \) for each \( l \in \{1, 2, \ldots, p + 1\} \). We stress the concept that the index \( i_{p+1} \) defines the present value of the squared return \( r^2(t) \), instead the indices \( i_{1}, \ldots, i_{p} \) define the past history of the process of squared returns, i.e. \( i_{1} \) defines the oldest value of \( r^2 = r^2(t - p) \). The transition probabilities are given by:

\[
V_{x(t)=k|i_{1}i_{2}\ldots i_{p+1}=1} = \eta_{k} (i_{1}, \ldots, i_{p}) = \frac{\exp \left[ \alpha_{k} + \sum_{l=1}^{p} \beta_{k,l} \left(3 - i_{p-l+1}\right)^2 \right]}{1 + \exp \left[ \alpha_{k} + \sum_{l=1}^{p} \beta_{k,l} \left(3 - i_{p-l+1}\right)^2 \right]},
\]

\[
V_{x(t)=k|i_{1}i_{2}\ldots i_{p+1}=2} = 0,
\]

doi:10.1088/1742-5468/2015/01/P01028 25
We can obtain this by the following mappings:

\[
V_{x(t)=k(1,4):i_1i_2...i_{p+1}=3} = \frac{1}{1 + \exp \left[ \alpha_k + \sum_{l=1}^{p} \beta_{k,l} \left( 3 - i_{p-l+1} \right)^2 \right]},
\]

\[
V_{x(t)=k(2,3):i_1i_2...i_{p+1}=1} = 0, \\
V_{x(t)=k(2,3):i_1i_2...i_{p+1}=2} = 1, \\
V_{x(t)=k(2,3):i_1i_2...i_{p+1}=3} = 0, \quad (B2)
\]

for each value of the \( p \)-dimensional vector \( i = (i_1, \ldots, i_p) \). We have \( 3^{p+1} \) values for the transition probabilities with normalization:

\[
\forall k; \forall i_1, \ldots, i_p : \sum_{i_{p+1}=1}^{3} V_{x(t)=k(i_1i_2...i_{p+1})} = 1. \quad (B3)
\]

We can recover an equivalent Markov(1) process defined on vector-states \( Y(t) \). We define a \( p \)-dimensional vector of squared returns:

\[
Y(t)[i] = \left( r^2(t-p+1) = (3-i_1)^2, \ldots, r^2(t) = (3-i_p)^2 \right). \quad (B4)
\]

In this case the index \( i_p \) defines the present state of the squared return \( r^2(t) \). The vector-process \( Y(t) \) is a first order Markov chain on the state space \( \{0, 1, 4\}^p \), i.e. \( Y(t) \) can assume \( 3^p \) different values. We define four transition matrices \( U_{x(t)=k} \in M_{3^p,3^p}(\mathbb{R}) \) in order to represent the equivalent Markov process for each possible value of \( x(t) \). These matrices describe the transition \( Y(t) \rightarrow Y(t+1) \), that we could represent also by the transition between vectors of indices: \( (i_1, \ldots, i_p) \rightarrow (i_2, \ldots, i_{p+1}) \). We have to map the transition probabilities \( V_{x(t)=k(i_1i_2...i_{p+1})} \) to the elements of matrix \( U_{k;m,n} \), where \( m, n \in \{1, \ldots, 3^p\} \).

We can obtain this by the following mappings:

\[
(i_1, \ldots, i_{p+1}) \rightarrow (m, n), \\
m(i_1, \ldots, i_p) = \left[ \sum_{l=1}^{p-1} 3^{p-l}(3-i_l) \right] + 4 - i_p, \\
n(i_2, \ldots, i_{p+1}) = \left[ \sum_{l=1}^{p-1} 3^{p-l}(3-i_{l+1}) \right] + 4 - i_{p+1}, \\
U_{x(t)=k;m,n} = V_{x(t)=k(i_1i_2...i_{p+1})}. \quad (B5)
\]

These rules are unable to fill the entire matrix \( U_{k;m,n} \), because when we study the Markov process for \( Y(t) \) we have many forbidden transitions, so the elements of the matrix that aren’t captured by the above rules have 0 values. For the case \( p = 2 \) the shape of \( U_k \) is:

\[
U_1 = \begin{pmatrix}
[1 - \eta_1 (0, 0)] & 0 & \eta_1 (0, 0) & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & [1 - \eta_1 (0, 1)] & 0 & \eta_1 (0, 1) & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & [1 - \eta_1 (0, 4)] & 0 & \eta_1 (0, 4) \\
[1 - \eta_1 (1, 0)] & 0 & \eta_1 (1, 0) & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & [1 - \eta_1 (1, 1)] & 0 & \eta_1 (1, 1) & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & [1 - \eta_1 (1, 4)] & 0 & \eta_1 (1, 4) \\
[1 - \eta_1 (4, 0)] & 0 & \eta_1 (4, 0) & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & [1 - \eta_1 (4, 1)] & 0 & \eta_1 (4, 1) \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix},
\]

doi:10.1088/1742-5468/2015/01/P01028
From the 3

where

process in term of the matrix

In \( U_1 \) we have 

where \( \lambda_k \) are given by equation (7).

Now our goal is to calculate the moments for the variable \( r^2 (t) \) from the process defined by equation (B6). First of all, we have to solve the eigenvalue equation for \( S \) relative to the eigenvalue 1 in order to determine the stationary probability vector for \( Y(t) \):

\[
S \Psi = \Psi. \tag{B7}
\]

The 3\( ^p \)-dimensional vector \( \Psi \) represents all possible values of the stationary distribution of the variable \( Y(t) \):

\[
P (Y(t) [i_1, \cdots, i_p]) = \Psi_{m(i_1, \cdots, i_p)}. \tag{B8}
\]

From the 3\( ^p \)-dimensional vector \( \Psi \) we compute the stationary 3D probability vector \( \psi' = (\psi_1, \psi_2, \psi_3) \) for the process \( r^2 (t) \), i.e. we have for each index \( i_p \in \{1, 2, 3\} \):

\[
\psi_{i_p} = P \left[ r^2 (t) = (3 - i_p)^2 \right] = \sum_{i_1=1}^3 \cdots \sum_{i_{p-1}=1}^3 \Psi_{m(i_1, \cdots, i_p)}, \tag{B9}
\]

where \( i_p \) defines the present value of \( r^2 (t) \) and we use mappings defined in equation (B5). The stationary probability to have a fixed value of \( r^2 \) at time \( t \) depends on all possible values of \( r^2 \) during the past \( p - 1 \) lags. In order to determine the present probabilities we have to sum probabilities corresponding to all possible past trajectories defined by the past \( p - 1 \) lags.

We compute \( \text{corr}(r^2(t), r^2(t + \tau)) = \rho(\tau) \) by means of the transition probabilities 

\[
P (r^2 (t) = (3 - a)^2, r^2 (t + \tau) = (3 - b)^2), \tag{B10}
\]

where \( i(a) = (i_1, \cdots, i_p = a) \) and \( j(b) = (j_1, \cdots, j_p = b) \) are the \( p \)-dimensional vectors of indices describing the past \( p - 1 \) lags respect to times \( t \) and \( t + \tau \). We have to perform the sum of probabilities corresponding to each of the possible values of \( i_1, \cdots, i_{p-1} \) and \( j_1, \cdots, j_{p-1} \), i.e. on \( i_t, j_t \in \{1, 2, 3\}, \forall t \in \{1, \cdots, p - 1\} \). We use mappings defined in

doi:10.1088/1742-5468/2015/01/P01028
that leads us to equation (27). In this way we have found the results reported in figure 7 for DCMM($p = 3$).

The moments of our interest are:

\[ E [r^2 (t)] = \sum_{i=1}^{3} (3 - i)^2 \psi_i = 4\psi_1 + \psi_2, \]

\[ E [r^4 (t)] = \sum_{i=1}^{3} (3 - i)^4 \psi_i = 16\psi_1 + \psi_2, \]

\[ E [r^2 (t) r^2 (t + \tau)] = \sum_{a=1}^{3} \sum_{b=1}^{3} \left[ (3 - a)^2 (3 - b)^2 P (\mathbf{i} (a) \mid \mathbf{j} (b)) \right], \]

from which we can determine the function $\rho (\tau)$. We have determined the function $\rho (\tau)$ for $p = 3$ making the following approximation for $V_4$:

\[ V_{x(t)=k=4, i_2, \ldots, i_{p+1}=1} = 2\theta_4, \]

\[ V_{x(t)=k=4, i_2, \ldots, i_{p+1}=2} = 0, \]

\[ V_{x(t)=k=4, i_2, \ldots, i_{p+1}=3} = 1 - 2\theta_4, \]

this approximation is justified only in the case $\lambda_1 \approx 1$, i.e. we have the same approximation that leads us to equation (27). In this way we have found the results reported in figure 7 for DCMM($p = 3$).

References

[1] Abergel F, Bouchaud J-P, Foucault T, Lehalle C-A and Rosenbaum M 2012 Market Microstructure: Confronting Many Viewpoints (Cornwall: Wiley)
[2] Andersen T G and Bollerslev T 1997 J. Empirical Finance 4 115
[3] Berchtold A 1999 Commun. Stat. Theory Methods 28 2569
[4] Berchtold A 2002 Stochastic Models 18 193
[5] Berchtold A and Raftery A E 2002 Stat. Sci. 17 328
[6] Bouchaud J-P and Potters M 2003 Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management (Cambridge: Cambridge University Press)
[7] Bulla J and Bulla I 2006 Stat. Sci. 21 192
[8] Cappe O, Moulines E and Ryden T 2005 Inference in Hidden Markov Models (New York: Springer)
[9] Clark P C 1973 Econometrica 41 135
[10] Cogburn R 1990 Ann. Probab. 18 642
[11] Cogburn R 1984 Z. Wahrscheinlichkeitstheorie verw. Gebiete 66 109
[12] Cont R and de Larrard A 2013 SIAM J. Financial Math. 4 1–25
[13] Cont R 2001 Quant. Finance 1 223
[14] Cristelli M, Alfi V, Pietronero L and Zaccaria A 2010 Eur. Phys. J. B 73 41–9
[15] D’Amico G and Petroni F 2012 Physica A 391 4867–76
[16] D’Amico G and Petroni F 2011 J. Stat. Mech. P12009
[17] D’Amico G and Petroni F 2012 J. Stat. Mech. P07015
[18] Daniels M G, Farmer J D, Gillemot L, Iori G and Smith E 2003 Phys. Rev. Lett. 90 108102

doi:10.1088/1742-5468/2015/01/P01028
Modeling the coupled return-spread high frequency dynamics of large tick assets

[19] Dayri K and Rosenbaum M 2012 arXiv:1207.6325
[20] Dayri K, Bacry E and Muzy J-F 2011 The nature of price returns during periods of high market activity Econophysics of Order-driven Markets ed F Abergel et al (New York: Springer) p 155
[21] Dayri K 2011 These de doctorat, Ecole Polytechnique
[22] Eisenkopf A 2008 http://ssrn.com/abstract=968311
[23] Eisler Z, Perelló J and Masoliver J 2007 Phys. Rev. E 76 056105
[24] Eisler Z, Bouchaud J-P and Kockelkoren J 2012 Quant. Finance 12 1395
[25] Engl e R F 1982 Econometrica 50 987
[26] Engle C and Hamilton J D 1990 Am. Economic Rev. 89 689
[27] Ferland R, Latour A and Oraichi D 2004 J. Time Ser. Anal. 27 923
[28] Fitzpatrick M and Marchev D 2013 Stat. Comput. 23 467
[29] Fokianos K and Kedem B 2002 Regression Models for Time Series Analysis (Hoboken, NJ: Wiley-Interscience)
[30] Garèche A, Disdier G, Kockelkoren J and Bouchaud J-P 2013 Phys. Rev. E 88 032809
[31] Gillemot L, Farmer J D and Lillo F 2006 Quant. Finance 6 371
[32] Granger C W J 1972 J. Am. Stat. Assoc. 67 275
[33] Groß-Klußmann A and Hautsch N 2013 J. Forecasting 32 724–42
[34] Hamilton J D 1994 Time Series Analysis (Princeton, NJ: Princeton University Press)
[35] Hamilton J D 1989 Econometrica 57 357
[36] Hamilton J D 2008 Regime switching models The New Palgrave Dictionary of Economics 2nd edn, ed S N Durlauf and L E Blume (London: Macmillan)
[37] Harris L 2003 Trading and Exchanges (Oxford: Oxford University Press)
[38] Hasbrouck J 2007 Empirical Market Microstructure (Oxford: Oxford University Press)
[39] House M G, Jiang H W and Zhang X C 2009 Phys. Rev. B 80 113308
[40] La Spada G, Farmer J D and Lillo F 2013 Stud. Nonlinear Dyn. Econometrics 18 445–82
[41] La Spada G, Farmer J D and Lillo F 2011 Tick size and price diffusion Econophysics of Order-Driven Markets ed F Abergel et al (New York: Springer) p 173
[42] Liesenfeld R, Nolte I and Pohlmeier W 2006 Empirical Economics 30 795
[43] McKenzie E 2003 Discrete variate time series Handbook of Statistics vol 21, ed D N Shanbhag and C R Rao (Amsterdam: Elsevier) p 573
[44] Mike S and Farmer J D 2008 J. Economic Dyn. Control 32 200
[45] Munnix M C, Schafer R and Gihr T 2010 Physica A 389 4828
[46] Onnela J-P, Toivonen H, Gauvin L, Barthelemy M and Arenas A 2009 Proc. Natl Acad. Sci. U.S.A. 106 6655
[47] Plerou V, Gopikrishnan P and Stanley H E 2005 Phys. Rev. E 71 046131
[48] Ponzi A, Lillo F and Mantegna R N 2009 Phys. Rev. E 80 016112
[49] Rabiner L R 1989 Proc. IEEE 77 257
[50] Robert C Y and Rosenbaum M 2011 J. Financial Econometrics 9 344
[51] Ryden T, Terasvirta T and Asbrink S 1998 J. Appl. Econometrics 13 217
[52] Slanina F 2001 Phys. Rev. E 64 056136
[53] Thakur V, Azad R K and Ramaswamy R 2007 Phys. Rev. E 75 011915
[54] Urdapilleta E 2011 Phys. Rev. E 84 041904
[55] Wallach H M 2004 http://repository.upenn.edu/cis_reports/22
[56] Wallach H M 2004 http://repository.upenn.edu/cis_reports/22
[57] Wyart M, Bouchaud J-P, Kockelkoren J, Potters M and Vettorazzo M 2008 Quant. Finance 8 41
[58] Zucchini W and MacDonald I L 2009 Hidden Markov Models for Time Series: an Introduction Using R (Boca Raton, FL: CRC Press)

doi:10.1088/1742-5468/2015/01/P01028