A new approach to the modeling of financial volumes

GUGLIELMO D’AMICO
Department of Pharmacy, University "G. d’Annunzio" of Chieti-Pescara
e-mail: g.damico@unich.it

FILIPPO PETRONI
Department of Economy and Business, University of Cagliari, Cagliari

Abstract
In this paper we study the high frequency dynamic of financial volumes of traded stocks by using a semi-Markov approach. More precisely we assume that the intraday logarithmic change of volume is described by a weighted-indexed semi-Markov chain model. Based on this assumptions we show that this model is able to reproduce several empirical facts about volume evolution like time series dependence, intra-daily periodicity and volume asymmetry. Results have been obtained from a real data application to high frequency data from the Italian stock market from first of January 2007 until end of December 2010.

Keywords semi-Markov; high frequency data; financial volume

1 Introduction
Studies on market microstructure have acquired a crucial importance in order to explain the price formation process, see e.g. De Jong and Rindi (2009). The main variables are (logarithmic) price returns, volumes and duration. Sometimes they are modeled jointly and the main approach is the so-called econometric analysis, see e.g. Manganelli (2005) and Podobnik et al. (2009) and the bibliography therein.

The volume variable is very important not only because it interacts directly with duration and returns but also because a correct specification of forecasted volumes can be used for Volume Weighted Average Price trading, see e.g. Brownlees et al. (2011). This variable has been investigated for long time and several statistical regularities have been highlighted, see e.g. Jain and Joh (1988).

In this paper we propose an alternative approach to the modeling of
financial volumes which is based on a generalization of semi-Markov processes called Weighted-Indexed Semi-Markov Chain (WISMC) model. This choice is motivated by recent results in the modeling of price returns in high-frequency financial data where the WISMC approach was demonstrated to be particular efficient in reproducing the statistical properties of financial returns, see D’Amico and Petroni (2011, 2012, 2014). The WISMC model is very flexible and for this reason we decided to test its appropriateness also for financial volumes.

It should be remarked that WISMC models generalize semi-Markov processes and also non-Markovian models based on continuous time random walks that were used extensively in the econophysics community, see e.g. Mainardi et al. (2000) and Raberto et al. (2002).

The model is applied to a database of high frequency volume data from all the stocks in the Italian Stock Market from first of January 2007 until end of December 2010. In the empirical analysis we find that WISMC model is a good choice for modeling financial volumes which is able to correctly reproduce stylized facts documented in literature of volumes such as the autocorrelation function.

The paper is divided as follows: First, Weighted-Indexed-Semi-Markov chains are shortly described in Section 2. Next, we introduce our model of financial volumes and an application to real high frequency data illustrates the results. Section 4 concludes and suggests new directions for future developments.

2 Weighted-Indexed Semi-Markov Chains

The general formulation of the WISMC as developed in D’Amico and Petroni (2012, 2014) is here only discussed informally.

WISMC models share similar ideas as those that generate Markov processes and semi-Markov processes. These processes are all described by a set of finite states \( J_n \) whose transitions are ruled by a transition probability matrix. The semi-Markov process differs from the Markov process because the transition times \( T_n \) where the states change values, are generated according to random variables. Indeed, the time between transitions \( T_{n+1} - T_n \) is random and may be modeled by means of any type of distribution functions. In order to better represent the statistical characteristics of high-frequency financial data, in a recent article, the idea of a WISMC model was introduced.
in the field of price returns, see D’Amico and Petroni (2012). The novelty, with respect to the semi-Markov case, consists in the introduction of a third random variable defined as follows:

$$I_n(\lambda) = \sum_{k=0}^{n-1} \sum_{a=T_{n-1-k}}^{T_{n-k}-1} f^\lambda(J_{n-1-k}, T_n, a),$$  

(1)

where $f$ is any real value bounded function and $I^0_\lambda$ is known and non-random. The variable $I_n^\lambda$ is designated to summarize the information contained in the past trajectory of the $\{J_n\}$ process that is relevant for future predictions. Indeed, at each past state $J_{n-1-k}$ occurred at time $a \in \mathbb{N}$ is associated the value $f^\lambda(J_{n-1-k}, T_n, a)$, which depends also on the current time $T_n$. The quantity $\lambda$ denotes a parameter that represents a weight and should be calibrated to data. In the applicative section we will describe the calibration of $\lambda$ as well as the choice of the function $f$.

The WISMC model is specified once a dependence structure between the variables is considered. Toward this end, the following assumption is formulated:

$$\mathbb{P}[J_{n+1} = j, T_{n+1} - T_n \leq t | J_n, I_n \lambda, J_{n-1}, T_{n-1}, I_{n-1} \lambda, \ldots] = \mathbb{P}[J_{n+1} = j, T_{n+1} - T_n \leq t | J_n, I_n] := Q^\lambda_{n,j}(I_n^\lambda; t).$$  

(2)

Relation (2) asserts that the knowledge of the values of the variables $J_n, I_n^\lambda$ is sufficient to give the conditional distribution of the couple $J_{n+1}, T_{n+1} - T_n$ whatever the values of the past variables might be. Therefore to make probabilistic forecasting we need the knowledge of the last state of the system and the last value of the index process. If $Q^\lambda(x; t)$ is constant in $x$ then the WISMC kernel degenerates in an ordinary semi-Markov kernel and the WISMC model becomes equivalent to classical semi-Markov chain models, see e.g. D’Amico and Petroni (2012a) and Fodra and Pham (2015).

The probabilities $Q^\lambda_{ij}(x; t))_{i,j \in E}$ can be estimated directly using real data. In D’Amico and Petroni (2017) it is shown that the estimator

$$\hat{Q}_{ij}^\lambda(x; t) := \frac{N_{ij}(x; t)}{N_i(x)},$$

is the approached maximum likelihood estimator of the corresponding transition probabilities. The quantity $N_{ij}(x; t)$ expresses the number of transitions from state $i$, with an index value $x$, to state $j$ with a sojourn time in state $i$ less or equal to $t$. The quantity $N_i(x)$ is the number of visits to state $i$ with an index value $x$. 

3
Let us assume that the trading volume of the asset under study is described by the time varying process $V(t)$, $t \in \mathbb{N}$.

The (logarithmic) change of volume at time $t$ over the unitary time interval is defined by

$$Z^V(t) = \log \frac{V(t+1)}{V(t)}.$$  

(3)

On a short time scale, $Z^V(t)$ changes value in correspondence of an increasing sequence of random times, $\{T^V_n\}_{n \in \mathbb{N}}$. According to the notation adopted in the previous section, we denote the values assumed at time $T^V_n$ by $J^V_n$ and the corresponding values of the index process by

$$I^V_n(\lambda) = \sum_{k=0}^{n-1} \sum_{a=T^V_{n-1-k}}^{T^V_n-1} f^\lambda(J^V_{n-1-k}, T^V_n, a).$$  

(4)

If we assume that the variables $(J^V_n, T^V_n, I^V_n(\lambda))$ satisfy relationship (2) then the volume process can be described by a WISMC model and if $N^V(t) = \sup\{n \in \mathbb{N} : T^V_n \leq t\}$ is the number of transition of the volume process, then $Z^V(t) = J^V_{N^V(t)}$ is the WISMC process that describes the volume values at any time $t$.

We assume also that the set $E$ is finite and is obtained by an opportune discretization of the values of the financial volumes. A description of the adopted discretization and of the state space model is described in next section. The choice of a finite state space can be formalized by defining the state space by

$$E = \{-z_{min}\Delta, \ldots, -2\Delta, -\Delta, 0, \Delta, 2\Delta, \ldots, z_{max}\Delta\}.$$  

Our objective is to demonstrate that a WISMC model for financial volume is able to reproduce some important facts of volume dynamics. One of the most important feature is the persistence of the volume process. As found in Manganelli (2005), the volume process is strongly persistent for frequently traded stocks. A possible explanations is that the volume is pushed up by the arrival of new information to the market participants and some times is necessary in order to delete the effects of the information arrival. In order to investigate the goodness of the WISMC model we define the autocorrelation
of the modulus of volumes as

\[ \Sigma(t, t + \tau) = C\text{ov}(|Z^V(t + \tau)|, |Z^V(t)|). \tag{5} \]

Another important statistic is the first passage time distribution of the volume process. First we need to define the accumulation factor of the volume process from the generic time \( t \) to time \( t + \tau \):

\[ M^V_t(\tau) = \frac{V(t + \tau)}{V(t)} = e^{\sum_{r=1}^{\tau-1} z^V(t+r)}. \tag{6} \]

We will denote the fpt by

\[ \Gamma_\sigma = \min\{\tau \geq 0; M^V_0(\tau) \geq \sigma\}, \]

where \( \sigma \) is a given threshold. We are interested in finding the distributional properties of the fpt, that is to compute

\[ \mathbb{P}[\Gamma_\sigma > t|(J^V, T^V, I^V(\lambda))^{-m}_0 = (i, t)^{-m}_0], \]

where

\[ (J^V)_0^{-m} = (J^{-m}_0, J^{-m+1}_0, \ldots, J^V_0), \]
\[ (T^V)_0^{-m} = (T^{-m}_0, T^{-m+1}_0, \ldots, T^V_0), \]
\[ (I^V(\lambda))_0^{-m} = (I^{-m}_0(\lambda), I^{-m+1}_0(\lambda), \ldots, I^V_0(\lambda)). \]

4 Application to real high frequency data

The data used in this work are tick-by-tick quotes of indexes and stocks downloaded from www.borsaitaliana.it for the period January 2007-December 2010 (4 full years). The data have been re-sampled to have 1 minute frequency. Every minutes the last price and the cumulated volume (number of transaction) is recorded. For each stock the database is composed of about \( 5 \times 10^5 \) volumes and prices. The list of stocks analyzed and their symbols are reported in Table 1.

In Figures 1 and 2 are shown the trading volume and the logarithmic change of the trading volume for the 4 stocks in the analyzed period.

To model \( Z^V(t) \) as a WISM model the first step is to discretize the variable. We choose 5 states of discretization with the following edges:
Table 1: Stocks used in the application and their symbols

| Symbol | Company       |
|--------|---------------|
| F      | Fiat          |
| ISP    | Intesa San Paolo |
| TIT    | Telecom       |
| TEN    | Tenaris       |

Figure 1: Trading volume of the analyzed stocks.
Figure 2: The (logarithmic) change of trading volume of the analyzed stocks.
Figure 3: Number of occupancy of each discretized state of the logarithm of volume change.
\{-\infty, -4, -1, 1, 4, \infty\}. The number of times that \(Z^V(t)\) fall in each state is shown in Figure 3.

Following D’Amico & Petroni (2012b) we use as definition of the function \(f^\lambda\) in (4) an exponentially weighted moving average (EWMA) of the squares of \(Z^V(t)\) which has the following expression:

\[
f^\lambda(J_{n-1-k}, T_n, a) = \frac{\lambda^{T_n-a} J_{n-1-k}^2}{\sum_{k=0}^{n-1} \sum_{a=T_n-1-k}^{T_n} \lambda^{T_n-a}} = \frac{\lambda^{T_n-a} J_{n-1-k}^2}{\sum_{a=1}^{T_n} \lambda^a}
\]

and consequently the index process becomes

\[
I_n^V(\lambda) = \sum_{k=0}^{n-1} \sum_{a=T_n-1-k}^{T_n} \left( \frac{\lambda^{T_n-a} J_{n-1-k}^2}{\sum_{a=1}^{T_n} \lambda^a} \right).
\]

The index \(I_n^V(\lambda)\) was also discretized into 5 states of low, medium low, medium, medium high and high volume variation. Using these definitions and discretizations we estimated, for each stock, the probabilities defined in the previous section by using their estimators directly from real data. By means of Monte Carlo simulations we were able to produce, for each of the 4 stocks, a synthetic time series.

Each time series is a realization of the stochastic process described in the previous section with the same time length as real data. Statistical features of these synthetic time series are then compared with the statistical features of real data. In particular, we tested our model for the ability to reproduce the autocorrelation functions of the absolute value of \(Z^V(t)\) and the first passage time distribution.

We estimated \(\Sigma(\tau)\) (see equation 5) for real data and for synthetic data and show in Figure 4 a comparison between them for all stocks. For each stock we estimated the percentage root mean square error (RMSE) the results are reported in Table 2.

It is possible to note that our model is able to reproduce almost perfectly the autocorrelation of the absolute value of the logarithmic volume change stocks.

For each of the stocks in our database we estimate the first passage time distribution directly from the data (real data) and from the synthetic time series generated as described above.

From Figure 5 it is obvious that the WISMC model is also able to reproduce the first passage time distribution of real data.
Figure 4: Autocorrelation functions of the absolute value of $Z^V(t)$ for real data (solid line) and synthetic (dashed line) time series.

| Stock | Error |
|-------|-------|
| F     | 3.6%  |
| ISP   | 3.2%  |
| TIT   | 3.0%  |
| TEN   | 3.9%  |

Table 2: Percentage square root mean error between real and synthetic autocorrelation function reported in Figure 4.
5 Conclusions

In this paper we advanced the use of Weighted-Indexed Semi-Markov Chain models for modeling high frequency financial volumes. We applied the model on real financial data and we shown that the model is able to reproduce important statistical fact of financial volumes as the autocorrelation function of the absolute values of volumes and the first passage time distribution. Further developments will be a more extensive application to other financial stocks and indexes and the proposal of a complete model where returns, volumes and durations are jointly described.

References

[1] C.T. Brownlees, F. Cipollini, G.M. Gallo. Journal of Financial Econometrics 9(3) (2011) 489-518.

[2] G. D’Amico, F. Petroni, ”A semi-Markov model with memory for price changes”, Journal of Statistical Mechanics: Theory and Experiment, P12009 (2011).

[3] G. D’Amico, F. Petroni, ”Weighted-indexed semi-Markov models for modeling financial returns”, Journal of Statistical Mechanics: Theory and Experiment, P07015 (2012).
[4] G. D’Amico, F. Petroni, "A semi-Markov model for price returns", Physica A 391 (2012) 4867-4876.

[5] G. D’Amico, F. Petroni, "Multivariate high-frequency financial data via semi-Markov processes", Markov Processes and Related Fields, 20 (2014) 415-434.

[6] G. D’Amico, F. Petroni, "Copula Based Multivariate Semi-Markov Models with Applications in High-Frequency Finance", Submitted.

[7] F. De Jong, B. Rindi, "The Microstructure of Financial Markets", Cambridge University Press (2009), Cambridge, New York.

[8] Jain, P., Joh, G., 1988. "The dependence between hourly prices and trading volume", Journal of Financial and Quantitative Analysis 23, 269-283.

[9] F. Mainardi, M. Raberto, R. Gorenflo, E. Scalas. Physica A 287 (2000) 468.

[10] S. Manganelli, "Duration, Volume and Volatility impact of trades", Journal of Financial Markets, 8, (2005) 377-399.

[11] B. Podobnik, D. Horvatic, A.M. Petersen, H.E. Stanley, "Cross-Correlations between Volume Change and price Change", PNAS, December 29(52) (2009) 22079-22084.

[12] M. Raberto, E. Scalas, F. Mainardi. Physica A 314 (2002) 749.