Intra-daily Volume Modeling and Prediction for Algorithmic Trading

Christian T. Brownlees* Fabrizio Cipollini† Giampiero M. Gallo‡

March 2009

Version prepared for the 1st FBF - IDEI-R Conference on
Investment Banking and Financial Markets,
Toulouse March 26-27, 2009.

Abstract

The explosion of algorithmic trading has been one the most recent prominent trends in the financial industry. Algorithmic trading consists of automated trading strategies that attempt to minimize transaction costs by optimally placing transactions orders. The key ingredient of many of these strategies are intra-daily volume predictions. This work proposes a dynamic model for intra-daily volume forecasting that captures salient features of the series such as intra-daily periodicity and volume asymmetry. Results show that the proposed methodology is able to significantly outperform common volume forecasting methods and delivers significantly more precise predictions in a VWAP tracking trading exercise.

*Department of Finance, Stern School of Business, NYU and Dipartimento di Statistica "G. Parenti", Università di Firenze, Italy. e-mail: `ctb@stern.nyu.edu`

†Dipartimento di Statistica "G. Parenti", Università di Firenze, Italy. e-mail: `cipollini@ds.unifi.it`

‡Corresponding author: Dipartimento di Statistica "G. Parenti", Università di Firenze, Viale G.B. Morgagni, 59 - 50134 Firenze – Italy. e-mail: `gallog@ds.unifi.it`

We would like to thank Robert Almgren, Rob Engle, Eric Ghysels, conference partecipants at the Chicago London Conference. All mistake are ours.

1 Introduction

Portfolio management and asset allocation require the acquisition or liquidation of positions: when the related volume is sizeable according to prevailing market conditions, placing an order is potentially able to change the price of that asset. This is particularly true for actions taken by institutional investors (e.g. pension funds or insurance companies managing large capitals) and for illiquid assets. The interaction between market participants may determine the creation of positions with the hope to profit from being on the other side of the large order. By the same token, large orders may need so–called price concessions in order to attract an adequate counterparty. The decision to buy or sell an asset in large quantities, in other words, must be informed as of the potential price impact which that particular trade may have (an effect known as *slippage*). This may result in lower profits or higher losses if the order is executed (transaction risk) or in the order not being executed at all.

In recent years, and increasingly so, services are being offered by specialized firms which provide *program trading* to institutional investors under the premise that their expertise will translate into a more efficient management of the transactions, minimizing slippage, or even into the assumption of some transaction risk. To be clear, there is no easy solution to transaction risk: the uncertainty around the actual execution price relative to one's own expected price (or of the execution of the order itself) must be weighed against the unavoidable uncertain market movements to face, should one decide to wait to place the order (*market risk*). To this extent, the relevant strategy is to plan how to place the orders relative to the characteristics of the financial market (rules and regulations, e.g. opening price formation), of the particular asset (e.g. liquidity, volatility, etc.), and, at a more advanced level, of that asset relative to other assets in a portfolio (e.g. correlation, common features, etc.).

Algorithmic trading (a.k.a. algo–trading) is widely used by investors who want to manage the market impact of exchanging large amounts of assets. It is favored by the development and diffusion of computer–based pattern recognition, so that information is processed instantaneously and action is taken accordingly with limited (if any) human judgement and intervention. The size of orders generated and executed by algo–trading is quite large and is increasing. In October 2006, the NYSE has boosted a mixed system of electronic and face–to–face auction which brings automated trades to about 50% of total trades, and similar trends are valid for other financial markets (smaller proportions when assets are more complex, e.g. options). It is generally recognized that algorithmic trading has reduced the average trade size (smaller liquidity) in the markets and hence has pushed institutional investors to split their orders in order to seek better price execution (cf. Chordia *et al.* (2008)).

The daily Volume Weighted Average Price (VWAP) was introduced by Berkowitz *et al.* (1988)) as a weighted average (calculated at the end of the day) of intra-daily transaction prices with weights equal to the relative size of the corresponding traded volume to the total volume traded during the day (defined as full VWAP in Madhavan (2002)). In the original paper, the difference between the price of a trade and the recorded VWAP was used to measure the market impact of that trade. The goal of institutional investors is to minimize such impact. VWAP is a very transparent measure, easily calculated at the end of the day with tick–by–tick data: it allows to evaluate how average traded prices were favorable to the trader. A VWAP replicating strategy is thus defined as a procedure for splitting a certain number of shares into smaller size orders during the day, which will be executed at different prices netting an average price that is close to the VWAP. Whether the VWAP benchmark is proposed on an agency base or on a guaranteed base (in exchange for a fee) is a technical aspect which does not have any bearings in what we discuss.

As we will see, in order to implement a VWAP replicating strategy we need to be able to forecast volumes in their intra-daily pattern, while it is less important to be able to predict prices. Reproducing the VWAP is based on estimating and forecasting the relative weight that volumes have during the day. As we will show in what follows, there are different components in the dynamics of traded volumes recorded at intra-daily intervals (relative to outstanding shares). We concentrate on single assets and we record intra-daily behavior at regular intervals. From an initial descriptive analysis of the series we derive some indications as of what features the model should reproduce. Beside the well documented U-shaped pattern of intra-daily trading activity which translates into a periodic component, we find that there are two other components which relate to a daily evolution of the volumes and to intra-daily non–periodic dynamics, respectively. We use these findings as a guideline to specify a *Component Multiplicative Error Model* where each element has its own dynamic specification. The model is specified in a semiparametric fashion, thus avoiding the choice of a specific distribution of the error term. We estimate all the parameters at once by Generalized Method of Moments. The estimated model can then be used to dynamically forecast the relative intra-daily volumes.

To be sure, our approach is just the first step into the implementation of an actual VWAP based strategy. Microstructure considerations put institutional investors in a different position from those traders who exploit intra-daily volatility and are not constrained by specific choices of assets. In the interaction between the two types, the latter will scan the books to detect whether some peculiar activity may reveal the presence of a large order placed by the former. At any rate, some orders may still be too large (relative to daily volume) to be filled in one day, so that the market impact is possibly unavoidable.

Our model shares the same logic as the component GARCH model suggested by Engle *et al.* (2006b), to model intra-daily volatility. The main difference lies in the evolution of the daily and intra-daily components. Exploiting the scheme proposed here, all parameters of the model can be estimated simultaneously, instead of recurring to a multi-step procedure. Engle *et al.* (2006a) propose econometric techniques for transaction cost analysis. Some connections can be found also with P-GARCH models introduced by Bollerslev and Ghysels (1996); relative to their suggestion, we achieve a simplification of the specification by imposing the same periodic pattern to the model coefficients (but see also Martens *et al.* (2002)). The literature on econometric models for intra-daily patterns of financial time series is quite substantial: from the initial contributions on price volume relationship (cf. the survey by Karpoff (1987)), the idea of relating intra-daily volatility and trading volumes as a function of an underlying latent information flow is contained in Andersen (1996). More recently, attention was specifically devoted to measuring the amount of liquidity of an asset based on the relationship between volume traded and price changes: Gouriéroux *et al.* (1999) concentrate on modelling weighted durations, that is the time needed to trade a given level (in quantity or value) of an asset. Dufour and Engle (2000) look at the time between trades and how that has an impact on price movements. Białkowski *et al.* (2008) concentrate on volume dynamics and take a factor analysis approach in a multivariate framework in which there is a common volume component to all stocks in an index and idiosyncratic components related to each stock which evolves according to a SETAR model. At any rate, the approach proposed here is quite general, given that some features of volumes are common to other non-negative intra-daily financial time series, such as realized volatilities, number of trades and average durations.

In this paper, we start from stylized facts (Section 2) to motivate the Component MEM (Section 3). Section 3.3 contains the details on the estimation procedure. The empirical application is divided up between model estimation and diagnostics 3.4 and volume forecasting and VWAP forecast comparisons 4. Concluding remarks follow (Section 5).

2 The Empirical Regularities of Intra-daily Volumes

We chose to analyze Exchange Traded Funds (ETFs), innovative financial products which allow straightforward trading in market averages as if they were stocks, while avoiding the possible idiosyncracies of single stocks. In the present framework, we count on a dataset consisting of regularly spaced intra-daily turnover and transaction price data for three popular equity index ETFs: DIA (Dow Jones ETF), QQQQ (Nasdaq ETF) and SPY (S&P 500 ETF). The corresponding turnover series are defined as the ratio of intra-daily transaction volume over the number of daily shares outstanding multiplied by 100. The frequency of the intra-daily data is 30 minutes, leading to 13 intra-daily bins. Volumes are computed as the sum of all transaction volumes occurred within each intra-daily bin. Prices are derived as the last recorded transaction price before the end of each bin. The sample period used in the analysis spans from January 2002 to December 2006 and we only consider days in which there were no empty bins, which corresponds to 1248 trading days and 16224 observations. The ultra high–frequency data used in the analysis are extracted from the TAQ while shares outstanding are taken from the CRSP. Details on the series handling and management are documented in Brownlees and Gallo (2006).

We first focus on the empirical regularities of the SPY turnover series which later will be used as a guideline for the suggested model. Similar evidence also holds for the other tickers for which we report summary descriptive statistics only.

Let us start with a graphic appraisal of the overall turnover (top panel of Figure 1): as with many financial time series, it clearly exhibits clustering of trading activity. Not surprisingly, turnover clustering is closely connected to clustering in realized volatility, but it measures a different dimension of market trading activity. Figure 2 plots the turnover and realized volatility time series (defined as the square root of realized variance based on 1-minute frequency intervals) together with their scatter plot. Interestingly, the clustering is retained if we take daily averages (cf. the series in the second panel of Figure

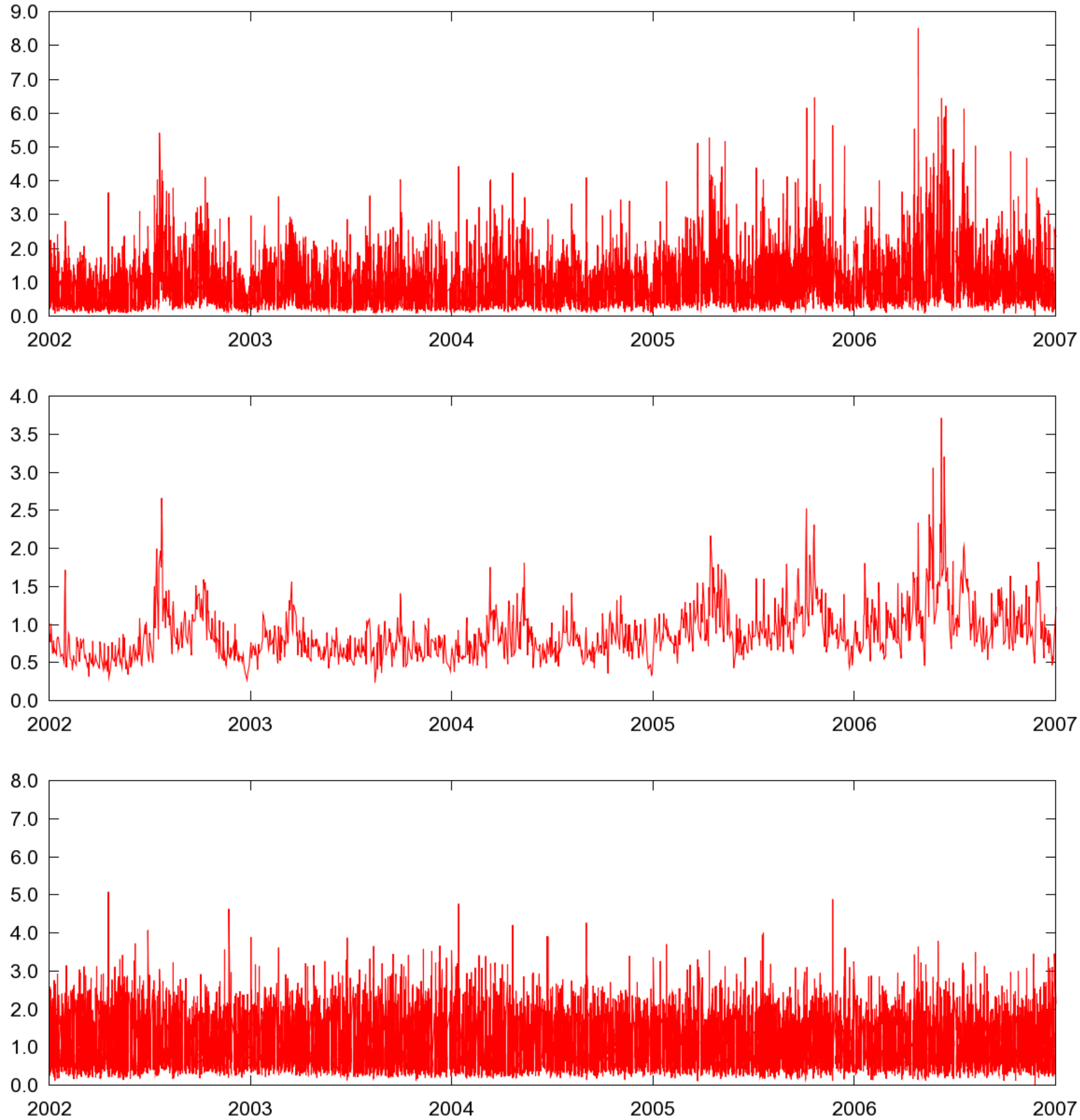

Figure 1: SPY Turnover Data: Original Turnover Data (top); Daily Averages (center), Intra–daily Component (bottom).

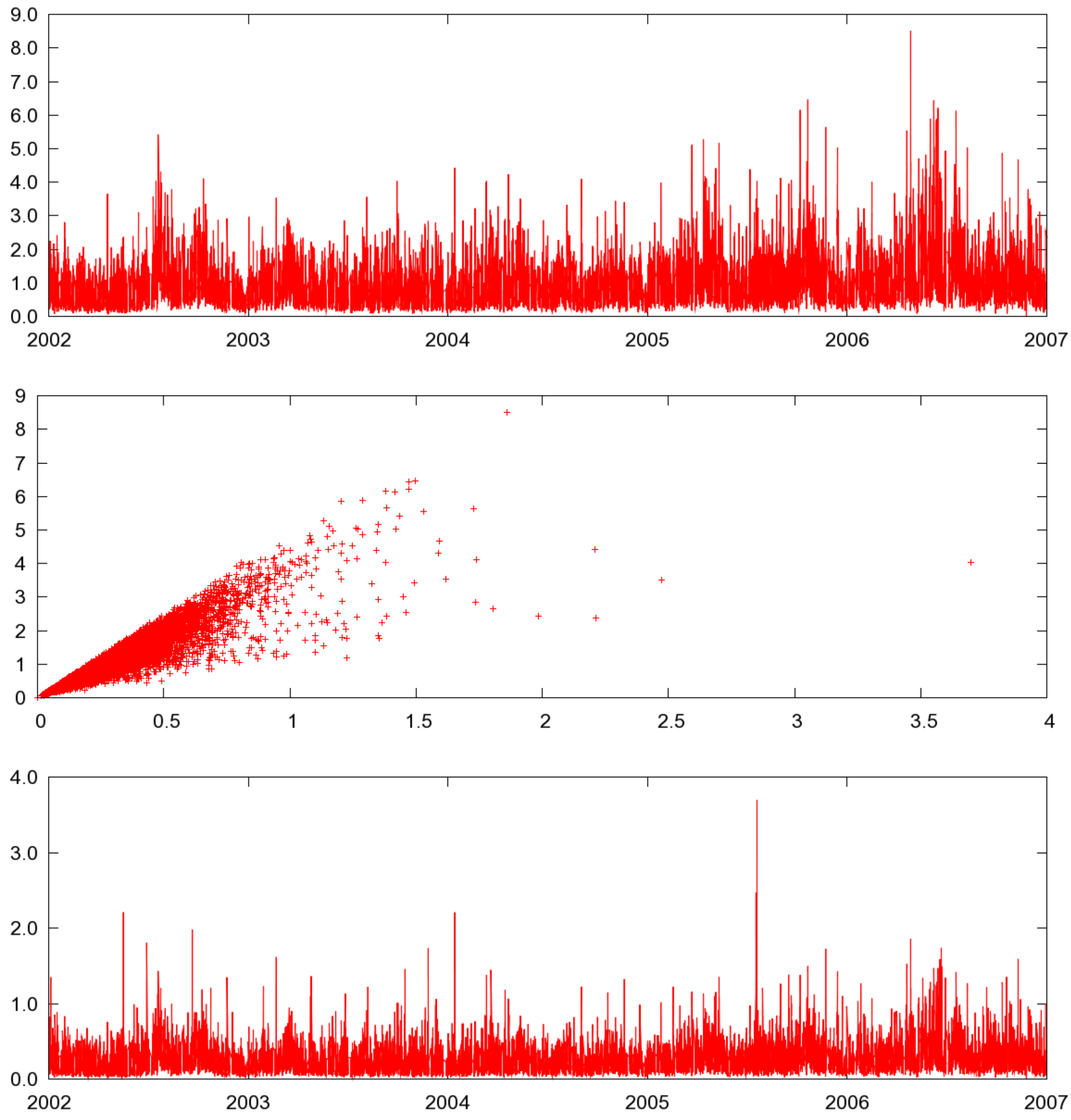

Figure 2: SPY Turnover Data: Turnover Data (top); (Realized Volatility, Turnover) Scatter Plot (center); Realized Volatility (bottom).

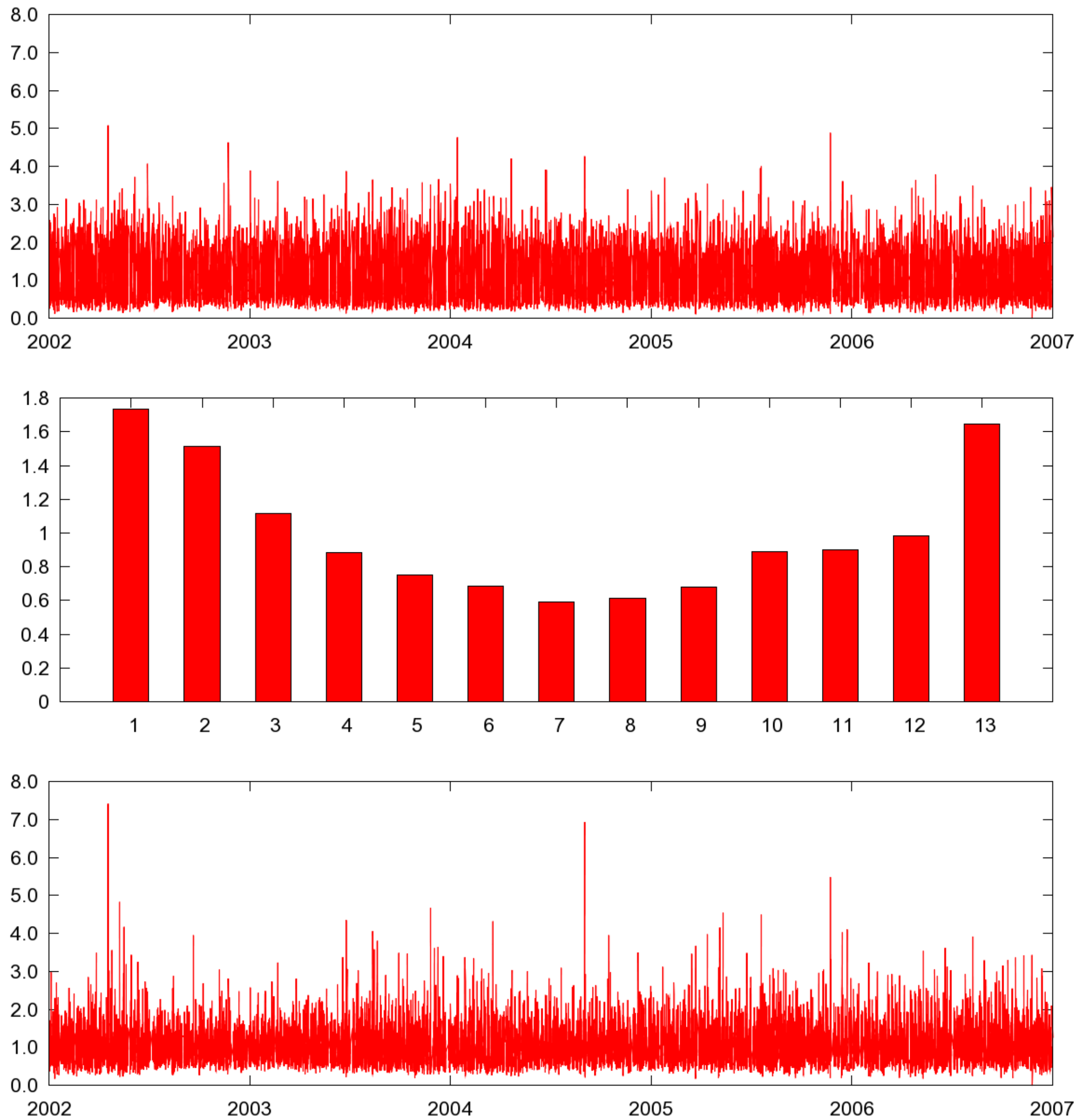

Figure 3: SPY Turnover Data: Intra–daily Component (top); Intra–daily Periodic Component (center); Intra–daily Non-periodic Component (bottom).

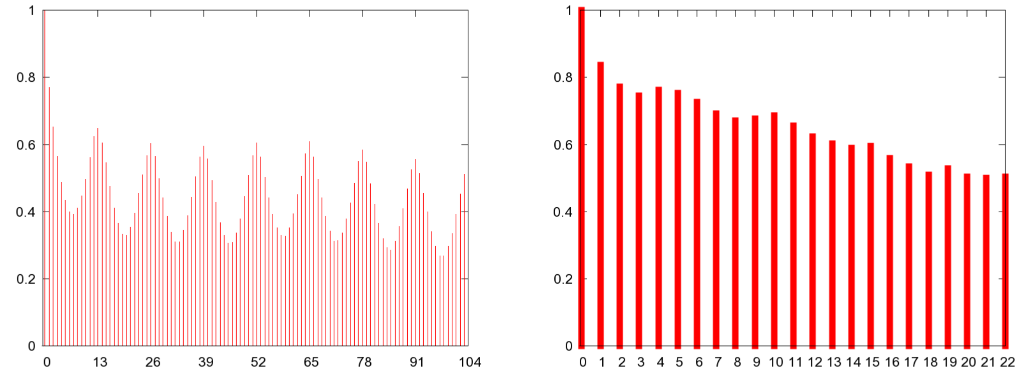

Figure 4: Autocorrelation Function of the Overall Turnover and of the Daily Averages.

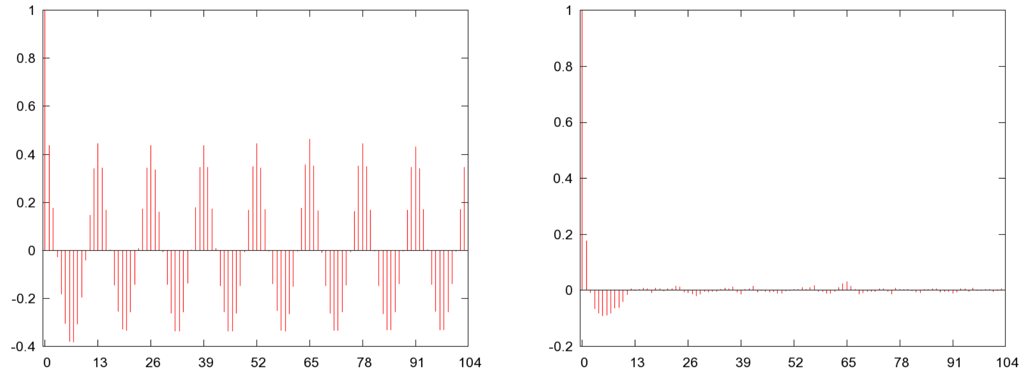

Figure 5: Autocorrelation Function of the Intra-daily (Periodic and Non–periodic) Components.

1). Dividing each observation by the corresponding daily average we obtain what we call an intra-daily pattern (bottom panel of Figure 1); following other stylized facts about intra-daily data, we suppose to have a periodic component and a non periodic component. Such descriptive analysis is performed in Figure 3 where we reproduce again, for ease of reference, the overall intra-daily pattern (top panel). The periodic component can be evaluated from the above series (i.e. that obtained by dividing data by the daily averages), as the average of 13 intra-daily bins (center panel). This average turnover by time of day exhibits the familiar U-shape of other intra-daily financial time series (e.g. average durations) which is consistent with the notion that trading activity is higher at the opening and closing of the markets and is low around mid-day. The ratio between these two series gives a non–periodic component which is shown in the bottom panel of Figure 3.

We believe that these three series (daily, intra-daily periodic, and intra-daily non periodic) have interesting dynamic features, which are confirmed by the analysis (Figure 4) of the correlograms of the original series (left panel) and of the daily averages (right panel). The use of unconditional intra-daily periodicity to adjust the original series results in a time series with a correlogram where periodicity is removed but some short–lived dependence is retained (Figure 5).

The inspection of the autocorrelations on the components (table 1) of the three tickers raw series substantially confirms the graphical analysis. The overall time series display relatively high levels of persistence which are also slowly decaying. The autocorrelations do not decrease by daily averaging. By dividing the overall turnover by its daily average (intra-daily component), a substantial part of dependence in the series is removed. Finally, once the intra-daily periodic component is removed, the resulting series show significant low order correlations only. Interestingly, the magnitudes of the various autocorrelations of the series are remarkably similar across the assets.

	overall		daily		intra-daily		intra-daily non-periodic	
	$\hat{\rho}_1$	$\hat{\rho}_{1\text{ day}}$	$\hat{\rho}_1$	$\hat{\rho}_{1\text{ day}}$	$\hat{\rho}_1$	$\hat{\rho}_{1\text{ day}}$	$\hat{\rho}_1$	$\hat{\rho}_{1\text{ week}}$
DIA	0.65	0.46	0.72	0.59	0.35	0.27	0.13	0.01
QQQQ	0.70	0.52	0.77	0.66	0.48	0.40	0.20	0.00
SPY	0.77	0.60	0.84	0.75	0.44	0.34	0.18	0.00

Table 1: Autocorrelations at selected lags of the turnover time series components.

3 A Multiplicative Error Model for Intra-daily Volumes

The empirical regularities discussed in Section 2 suggest a specification for intra-daily volumes which decomposes the series in three components: one daily and two intra-daily (one periodic and one dynamic). Let us first establish the notation used throughout the paper. Days are denoted with $t \in \{1, \ldots, T\}$; each day is divided into J equally spaced intervals (referred to as bins) indexed by $j \in \{1, \ldots, J\}$. In what follows, in order to simplify the notation we may label observations indexed by the double subscript $t\,j$ with a single progressive subscript $\tau = J \times (t-1) + j$. Correspondingly, we denote the total number of observations by N (equal to $T \times J$ if all J bins of data are available for all T days).

The non-negative quantity under analysis relative to bin j of day t is denoted as $x_{t\,j}$ or, alternatively, as x_τ. $\mathcal{F}_{t\,j-1}$ indicates the information about $x_{t\,j}$ available before forecasting it. Usually, we will assume $\mathcal{F}_{t\,0} = \mathcal{F}_{t-1\,J}$ but, if needed, it is possible to include additional pieces of information into $\mathcal{F}_{t\,0}$, specifically related to market opening structure.

In what follows we will adopt the following convention: if $\mathbf{x}_1, \ldots, \mathbf{x}_K$ are (m, n) matrices then $(\mathbf{x}_1; \ldots; \mathbf{x}_K)$ means the (mK, n) matrix obtained stacking the $\mathbf{x}_t$ matrices columnwise.

3.1 Model Definition

Being the $x_{t\,j}$'s non-negative, a model for their daily-intra-daily dynamic can be specified by extending the logic of Multiplicative Error Models (MEM) proposed by Engle (2002). Moreover, by relying on the stylized facts showed in Section 2, we structure the model by combining different components, each one able to capture a different feature of the dynamic of the time series. We will provide further remarks about the link of the model with the empirical regularities in Section 3.2.

We then assume a Component MEM (CMEM)

$$x_{t\,j} = \eta_t \, s_j \, \mu_{t\,j} \, \varepsilon_{t\,j}. \tag{1}$$

The multiplicative innovation term $\varepsilon_{t\,j}$ is assumed i.i.d., non-negative, with mean 1 and constant variance σ^2:

$$\varepsilon_{t\,j}|\mathcal{F}_{t\,j-1} \sim (1, \sigma^2). \tag{2}$$

The conditional expectation of $x_{t\,j}$ is the product of three multiplicative elements:

- η_t, a daily component;
- s_j, an intra-daily periodic component aimed to reproduce the time–of–day pattern;
- $\mu_{t\,j}$, an intra-daily dynamic (non-periodic) component.

In order to simplify the exposition, we assume a relatively simple specification for the components. If needed, the formulation proposed can be trivially generalized, for instance by including other predetermined variables or more lags (see the empirical application in Section 3.4).

The reference formulation for the daily component is

$$\eta_t = \omega^{(\eta)} + \beta_1^{(\eta)} \eta_{t-1} + \alpha_1^{(\eta)} x_{t-1}^{(\eta)} + \gamma_1^{(\eta)} x_{t-1}^{-\,(\eta)} \tag{3}$$

where $x^{(\eta)}$ is a quantity related to the innovations and $x^{-\,(\eta)}$ is an 'asymmetric version' of $x^{(\eta)}$, aimed to capture possible asymmetric effects related to daily returns (more on this below).

The intra-daily dynamic component is specified as

$$\mu_{t\,j} = \omega^{(\mu)} + \beta_1^{(\mu)} \mu_{t\,j-1} + \alpha_1^{(\mu)} x_{t\,j-1}^{(\mu)} + \gamma_1^{(\mu)} x_{t\,j-1}^{-\,(\mu)} \tag{4}$$

where, again, $x^{(\mu)}$ is a quantity related to the innovations and $x^{-\,(\mu)}$ is an 'asymmetric version' of $x^{(\mu)}$, useful for modeling possible asymmetric effects related to bin returns.

More precisely, the quantities $x_{t\,j}^{(\mu)}$, $x_t^{(\eta)}$ and the corresponding asymmetric effects $x_{t\,j}^{-\,(\mu)}$, $x_t^{-\,(\eta)}$ are defined as follows:

$$x_{t\,j}^{(\mu)} = \frac{x_{t\,j}}{\eta_t s_j}, \tag{5}$$

$$x_t^{(\eta)} = \frac{1}{J} \sum_{j=1}^{J} \frac{x_{t\,j}}{s_j \mu_{t\,j}} \tag{6}$$

$$x_{t\,j}^{-\,(\mu)} = x_{t\,j}^{(\mu)} \, \mathrm{I}(r_{t\,j} < 0) \tag{7}$$

$$x_t^{-\,(\eta)} = x_t^{(\eta)}\,\mathrm{I}(r_{t.} < 0), \tag{8}$$

where $r_{t\,j}$ indicates the return at bin j of day t and $r_{t\,.}$ denotes the daily return at day t.

The intra-daily non-periodic component can be initialized with the latest quantities available, namely those computed on the previous day, i.e.

$$\mu_{t\,0} = \mu_{t-1\,J} \qquad x_{t\,0}^{(\mu)} = x_{t-1\,J}^{(\mu)} \qquad x_{t\,0}^{-\,(\mu)} = x_{t-1\,J}^{-\,(\mu)}. \tag{9}$$

Both η_t and $\mu_{t\,j}$ are assumed to be mean-stationary. Furthermore, $\mu_{t\,j}$ is constrained to have unconditional expectation equal to 1 in order to make the model identifiable. This allow us to interpret it as a pure intra-daily component and implies $\omega^{(\mu)} = 1 - \beta_1^{(\mu)} - \alpha_1^{(\mu)} - \gamma_1^{(\mu)}/2$. From these assumptions we obtain also that reasonable starting conditions for the system can be $\eta_0 = x_0^{(\eta)} = \overline{x}$, $x_0^{-\,(\eta)} = \overline{x}/2$, $\mu_{1,0} = x_{1,0}^{(\mu)} = 1$ and $x_{1,0}^{-\,(\mu)} = 1/2$, where $\overline{x}$ indicates the sample average of the modeled variable x.

In synthesis, the system nests the daily and the intra-daily dynamic components by alternating the update of the former (from η_{t-1} to η_t) and of the latter (from $\mu_{t\,0} = \mu_{t-1\,J}$ to $\mu_{t\,J}$). η_t aims to adjust the mean level of the series, whereas the intra-daily component $s_j\mu_{t\,j}$ captures bin–specific departures from such an average level.

Note that defining $x_{t\,j}^{(\mu)}$ as in (5) implies $x_{t\,j}^{(\mu)} = \mu_{t\,j}\varepsilon_{t\,j}$. Combining this with (2) one obtains

$$E(x_{t\,j}^{(\mu)}|\mathcal{F}_{t\,j-1}) = \mu_{t\,j} \qquad V(x_{t\,j}^{(\mu)}|\mathcal{F}_{t\,j-1}) = \mu_{t\,j}^2\sigma^2 \tag{10}$$

that coincide with the properties of the usual MEM (Engle (2002)). Interestingly, a similar consideration can be made for $x_t^{(\eta)}$. In fact, definition (6) implies $x_t^{(\eta)} = \eta_t\overline{\varepsilon}_t$, where $\overline{\varepsilon}_t = J^{-1}\sum_{j=1}^{J}\varepsilon_{t\,j}$, and thus

$$E(x_t^{(\eta)}|\mathcal{F}_{t-1\,J}) = \eta_t \qquad V(x_t^{(\eta)}|\mathcal{F}_{t-1\,J}) = \eta_t^2\sigma^2/J. \tag{11}$$

The intra-daily periodic component s_j can be specified in different manners. A conceptually simple specification resorts to dummy variables:

$$\ln s_j = \sum_{k=1}^{J}\delta_k\,\mathrm{I}(k = j), \tag{12}$$

where $\mathrm{I}(.)$ denotes the indicator function and the δ_k's are coefficients, constrained to sum to zero. However, such a scheme is not efficient in practice, since it implies a lot of parameters when small bins are used (for instance, by considering 10-minutes data in a a market trading day between 9:30AM and 4:00PM, expression (12) would have 38 free parameters) and does not exploit the fact that the periodic intra-daily pattern has a substantially smooth shape. To this aim, a more parsimonious parameterization can be obtained by structuring s_j via Fourier (sine/cosine) representation:

$$\ln s_j = \sum_{k=1}^{K} \left[\delta_{1\,k} \cos\left(fjk\right) + \delta_{2\,k} \sin\left(fjk\right)\right] \tag{13}$$

where $f = 2\pi/J$, $K = \left[\frac{J+1}{2}\right]$, $\delta_{1\,K} = 0$ if J is odd, $\delta_{2\,K} = 0$. However, the number of terms into (13) can be considerably reduced if the periodic intra-daily pattern is sufficiently smooth, since few low frequencies harmonics may be enough.

3.2 Discussion

Let us first show that our model responds to the previous motivation based on the descriptive analysis in Section 2.

The daily average $\overline{x}_{t\,.} = J^{-1}\sum_{j=1}^{J} x_{t\,j}$ represents a proxy of the daily component η_t. In fact, by taking its expectation conditionally on the previous day, we have

$$E(\overline{x}_{t\,.}|\mathcal{F}_{t-1\,J}) = \eta_t \frac{1}{J}\sum_{j=1}^{J} s_j E(\mu_{t\,j}|\mathcal{F}_{t-1\,J}) \simeq \eta_t \frac{1}{J}\sum_{j=1}^{J} s_j = \eta_t \overline{s}, \tag{14}$$

where the approximate equality can be justified by noting that the non-periodic intra-daily component $\mu_{t\,j}$ has unit unconditional expectation, so that we can reasonably guess that it moves around this value.[1]

Once the daily average is computed, the ratio $x_{t\,j}^{(I)} = x_{t\,j}/\overline{x}_{t\,.}$ can be used as a proxy of the whole intra-daily component $s_j \mu_{t\,j}$, since

$$x_{t\,j}^{(I)} = \frac{x_{t\,j}}{\overline{x}_{t\,.}} \simeq \frac{\eta_t s_j \mu_{t\,j} \varepsilon_{t\,j}}{\eta_t \overline{s}} = \frac{s_j \mu_{t\,j} \varepsilon_{t\,j}}{\overline{s}}. \tag{15}$$

The bin average of the quantities into (15), namely $\overline{x}_{.\,j}^{(I)} = T^{-1}\sum_{t=1}^{T} x_{t\,j}^{(I)}$, represents a proxy of the intra-daily periodic component s_j. In fact,

$$\overline{x}_{.\,j}^{(I)} = \frac{1}{T}\sum_{t=1}^{T} x_{t\,j}^{(I)} \simeq \frac{s_j}{\overline{s}} \frac{1}{T}\sum_{t=1}^{T} \mu_{t\,j} \varepsilon_{t\,j}. \tag{16}$$

By taking its expectation conditionally on the starting information, we have

$$E(\overline{x}_{.\,j}^{(I)}|\mathcal{F}_{0\,J}) \simeq \frac{s_j}{\overline{s}} \frac{1}{T}\sum_{t=1}^{T} E(\mu_{t\,j}|\mathcal{F}_{0\,J}) \simeq \frac{s_j}{\overline{s}}. \tag{17}$$

The last approximation can be motivated by considering that the average of the $\mu_{t\,j}$'s for

[1] We remark as the log formulation of the intra-daily periodic component guarantees $\prod_{j=1}^{J} s_j = 1$ but not $\overline{s} = 1$. However, for the applications considered $\overline{s}$ is quite close to one.

bin j converges, in some sense, to the unconditional average 1.

Finally, the residual quantity $x_{tj}^{(I)}/\overline{x}_{.j}^{(I)} = x_{tj}/\overline{x}_{.j}$ can be justified as proxy of the intradaily non-periodic component, since

$$\frac{x_{tj}^{(I)}}{\overline{x}_{.j}^{(I)}} \simeq \frac{s_j \mu_{tj} \varepsilon_{tj}/\overline{s}}{s_j/\overline{s}} = \mu_{tj}\varepsilon_{tj}. \tag{18}$$

The CMEM of Section 3.1 has some relationships with the component GARCH model suggested by Engle *et al.* (2006b), for modeling intra-daily volatility. Our proposal differs however in many points. In particular, the main difference lies in the evolution of the daily and intra-daily components. Exploiting the scheme proposed, all parameters of the model can be estimated jointly, instead to recurring to a multi-step procedure.

The structure of the CMEM shares some features with the P-GARCH model (Bollerslev and Ghysels (1996)) as well. By grouping intra-daily components s_j and μ_{tj} and referring to equation (4) for the latter, the combined component can be written as

$$s_j \mu_{tj} = \omega_j^{(\mu)} + \beta_{1j}^{(\mu)} \mu_{tj-1} + \alpha_{1j}^{(\mu)} x_{tj-1}^{(\mu)} + \gamma_{1j}^{(\mu)} x_{tj-1}^{-(\mu)}, \tag{19}$$

where

$$\omega_j^{(\mu)} = \omega^{(\mu)} s_j \qquad \beta_{1j}^{(\mu)} = \beta_1^{(\mu)} s_j \qquad \alpha_{1j}^{(\mu)} = \alpha_1^{(\mu)} s_j \qquad \gamma_{1j}^{(\mu)} = \gamma_1^{(\mu)} s_j. \tag{20}$$

In practice, those defined in (20) are periodic coefficients: their pattern is ruled by s_j but each of them is rescaled by a (possibly) different value. The main difference relative to the P-GARCH formulation lies in the considerable simplification obtained by imposing the same periodic pattern to all coefficients. In this respect, we are inspired by the results in Martens *et al.* (2002) that a relatively parsimonious formulation, based on an intra-daily periodic component scaling the dynamical (GARCH-like) component of the variance, provides forecasts of the intra-daily volatility that are only marginally worse of a more computationally expensive P-GARCH. Martens *et al.* (2002) provide also empirical evidence in favor of the exponential formulation of the periodic intra-daily component and support its representation in a Fourier form (even if they consider only to the first 4 harmonics in their application). This notwithstanding, we depart from their approach in at least two substantial points: we include an explicit dynamic structure for the daily component, interpreting the intra-daily component as a corresponding scale factor; all parameters of the CMEM are estimated jointly.

3.3 Inference

Let us now illustrate how to obtain inferences on the model specified in Section 3.1. We group the main parameters of interest into the p-dimensional vector $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(\eta)}; \boldsymbol{\theta}^{(s)}; \boldsymbol{\theta}^{(\mu)})$, where the three subvectors refer to the corresponding components of the model. Relative to these, the variance of the error term, σ^2, represents a nuisance parameter.

Since the model is specified in a semiparametric way (see (2)), we focus our attention on the Generalized Method of Moments (GMM – Newey and McFadden (1994) and Wooldridge (1994)) as an estimation strategy not needing the specification of a density function for the innovation term. Links with other estimation methods leading to comparable inferences are illustrated in Section 3.3.3.

3.3.1 Efficient GMM inference

Let

$$u_\tau = \frac{x_\tau}{\eta_t s_j \mu_\tau} - 1, \tag{21}$$

where we simplified the notation by suppressing the reference to the dependency of u_τ on the parameters $\boldsymbol{\theta}$, on the information $\mathcal{F}_{\tau-1}$ and on the current value of the dependent variable x_τ.

From (2) one obtain immediately that u_τ is a conditionally homoskedastic martingale difference, that is

$$E(u_\tau|\mathcal{F}_{\tau-1}) = 0 \tag{22}$$

$$V(u_\tau|\mathcal{F}_{\tau-1}) = \sigma^2. \tag{23}$$

Following Wooldridge (1994, sect. 7) a conditional moment restriction like (22) can be used as a key ingredient for estimation. Any $(M, 1)$ vector $\mathbf{G}_\tau$ depending deterministically on the information $\mathcal{F}_{\tau-1}$ gives

$$E(\mathbf{G}_\tau u_\tau|\mathcal{F}_{\tau-1}) = \mathbf{0}, \tag{24}$$

and then, by the law of iterated expectations,

$$E(\mathbf{G}_\tau u_\tau) = \mathbf{0}, \tag{25}$$

so that $\mathbf{G}_\tau$ is uncorrelated with u_τ,[2] and can be taken as an instrument (it is dependent on $\mathcal{F}_{\tau-1}$ and uncorrelated with u_τ). $\mathbf{G}_\tau$ may depend on nuisance parameters, maybe including $\boldsymbol{\theta}$ also. We collect them into the vector $\boldsymbol{\psi}$ and, in order for us to concentrate on estimating $\boldsymbol{\theta}$, we assume for the moment that $\boldsymbol{\psi}$ is a known constant, postponing any further discussion about its role and how to inference it to the end of this section and to Section 3.3.2.

In order to discuss some general aspects about GMM inference, we denote

$$\mathbf{g}_\tau = \mathbf{G}_\tau u_\tau \tag{26}$$

[2] As remarked by Wooldridge (1994, p. 2693), equation (25) requires that the absolute values of u_τ and $\mathbf{G}_\tau u_\tau$ have finite expectations.

so that (24) and (25) can be rewritten as

$$E(\mathbf{g}_\tau) = E(\mathbf{g}_\tau|\mathcal{F}_{\tau-1}) = \mathbf{0}. \tag{27}$$

Equation (27) provides M moment conditions. If $M = p$, we have as many equations as the dimension of $\boldsymbol{\theta}$, thus leading to the MM criterion

$$\overline{\mathbf{g}} = \mathbf{0}, \tag{28}$$

where

$$\overline{\mathbf{g}} = \frac{1}{N}\sum_{\tau=1}^{N} \mathbf{g}_\tau. \tag{29}$$

Otherwise, if $M > p$ or (28) does not have a solution, then a GMM criterion

$$\min_{\boldsymbol{\theta}} \frac{1}{2}\overline{\mathbf{g}}'\widehat{\boldsymbol{\Lambda}}_N\overline{\mathbf{g}} \tag{30}$$

can be employed, where $\widehat{\boldsymbol{\Lambda}}_N$ is an (M, M) weighting matrix assumed non-negative definite and converging in probability to a non-stochastic limiting matrix $\boldsymbol{\Lambda}_0$.

Under correct specification of the equations arising (27) (the η_t, s_j, and μ_{tj} equations in our case) and some regularity conditions (among which that $\widehat{\boldsymbol{\Lambda}}_N$ converges in probability as specified above), the GMM estimator $\widehat{\boldsymbol{\theta}}_N$, obtained solving (28) or (30) for $\boldsymbol{\theta}$, is consistent (Wooldridge (1994, th. 7.1)). Furthermore, under some additional regularity conditions we have asymptotic normality of $\widehat{\boldsymbol{\theta}}_N$, with asymptotic variance matrix

$$\mathrm{Avar}(\widehat{\boldsymbol{\theta}}_N) = \frac{1}{N}(\mathbf{S}'\boldsymbol{\Lambda}_0\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Lambda}_0\mathbf{V}\boldsymbol{\Lambda}_0\mathbf{S}(\mathbf{S}'\boldsymbol{\Lambda}_0\mathbf{S})^{-1} \tag{31}$$

where

$$\mathbf{S} = \lim_{N\to\infty}\frac{1}{N}\sum_{\tau=1}^{N} E\left(\nabla_{\boldsymbol{\theta}'}\mathbf{g}_\tau\right) \tag{32}$$

$$\mathbf{V} = \lim_{N\to\infty}\frac{1}{N}V\left(\sum_{\tau=1}^{N}\mathbf{g}_\tau\right) \tag{33}$$

(Wooldridge (1994, th. 7.2)).

The asymptotic variance (31) can be simplified: either by choosing $\widehat{\boldsymbol{\Lambda}}_N$ to be a consistent estimator of $\mathbf{V}^{-1}$ (in such a case, in fact, $\boldsymbol{\Lambda}_0 = \mathbf{V}^{-1}$) or when $M = p$ (in such a circumstance $\mathbf{S}$ is non-singular, because $\mathrm{rank}(\mathbf{S}) = p$ is one of the cited additional assumptions, and we can take $\widehat{\boldsymbol{\Lambda}}_N$ as the identity matrix). In both these cases

$$\mathrm{Avar}(\widehat{\boldsymbol{\theta}}_N) = \frac{1}{N}(\mathbf{S}'\mathbf{V}^{-1}\mathbf{S})^{-1}. \tag{34}$$

Given the structure of the model under analysis we can adopt some considerable simplifications, stemming in particular from the fact that u_τ is a martingale difference (see Wooldridge (1994)). In fact, such a characteristic implies that $\mathbf{g}_\tau = \mathbf{G}_\tau u_\tau$ also is a martingale difference (equation (27)): this leads to simplifications in the assumptions needed for the asymptotic normality, by virtue of the martingale CLT, and is a sufficient condition for the $\mathbf{g}_\tau$ terms in (33) to be serially uncorrelated. We thus have

$$\mathbf{V} = \lim_{N\to\infty}\left[\frac{1}{N}\sum_{\tau=1}^{N} E\left(\mathbf{g}_\tau \mathbf{g}_\tau'\right)\right]. \tag{35}$$

The martingale difference structure of u_τ gives also a simple formulation for the *efficient* choice of the instrument $\mathbf{G}_\tau$, where efficient is meant producing the 'smallest' asymptotic variance among the GMM estimators arisen by $\overline{\mathbf{g}}$ functions structured as in (29), with $\mathbf{g}_\tau = \mathbf{G}_\tau u_\tau$ a and $\mathbf{G}_\tau$ being an instrument. Such efficient choice is

$$\mathbf{G}_\tau^* = -E(\nabla_{\boldsymbol{\theta}} u_\tau|\mathcal{F}_{\tau-1})V(u_\tau|\mathcal{F}_{\tau-1})^{-1}. \tag{36}$$

Computing $E\left(\mathbf{g}_\tau \mathbf{g}_\tau'\right)$ into (35) and $E\left(\nabla_{\boldsymbol{\theta}'}\mathbf{g}_\tau\right)$ into (32) we obtain

$$E\left(\mathbf{g}_\tau \mathbf{g}_\tau'\right) = -E\left(\nabla_{\boldsymbol{\theta}'}\mathbf{g}_\tau\right) = \sigma^2 E\left(\mathbf{G}_\tau^* \mathbf{G}_\tau^{*\prime}\right),$$

so that

$$\mathbf{V} = -\mathbf{S} = \sigma^2 \lim_{N\to\infty}\frac{1}{N}\sum_{\tau=1}^{N} E\left(\mathbf{G}_\tau^* \mathbf{G}_\tau^{*\prime}\right)$$

and (34) specializes as

$$\mathrm{Avar}(\widehat{\boldsymbol{\theta}}_N) = \frac{1}{N}(\mathbf{S}'\mathbf{V}^{-1}\mathbf{S})^{-1} = -\frac{1}{N}\mathbf{S}^{-1} = \frac{1}{N}\mathbf{V}^{-1}. \tag{37}$$

Considering the analytical structure of u_τ in the model (equation (21)), we have

$$\nabla_{\boldsymbol{\theta}} u_\tau = -\mathbf{a}_\tau(u_\tau + 1),$$

where

$$\mathbf{a}_\tau = \eta_t^{-1}\nabla_{\boldsymbol{\theta}}\eta_t + \mu_\tau^{-1}\nabla_{\boldsymbol{\theta}}\mu_\tau + s_j^{-1}\nabla_{\boldsymbol{\theta}} s_j \tag{38}$$

so that (36) becomes

$$\mathbf{G}_\tau^* = \mathbf{a}_\tau \sigma^{-2}.$$

Replacing it into (26) and this, in turn, into (29), we obtain that the GMM estimator of $\boldsymbol{\theta}$ in the CMEM solves the MM equation

$$\frac{1}{N}\sum_{\tau=1}^{N}\mathbf{a}_\tau u_\tau = \mathbf{0}. \tag{39}$$

An important characteristic of (39) lies in the fact that it does not depend on the nuisance parameter σ^2 (that can be removed from the equation). This implies that its estimation does not have any bearing on the estimation and inference relative to the main parameter $\boldsymbol{\theta}$.

The asymptotic variance matrix of $\widehat{\boldsymbol{\theta}}_N$ is

$$\mathrm{Avar}(\widehat{\boldsymbol{\theta}}_N) = \frac{\sigma^2}{N}\left[\lim_{N\to\infty}\frac{1}{N}\sum_{\tau=1}^{N} E(\mathbf{a}_\tau \mathbf{a}_\tau')\right]^{-1} \tag{40}$$

that can be consistently estimated by

$$\widehat{\mathrm{Avar}}(\widehat{\boldsymbol{\theta}}_N) = \widehat{\sigma}_N^2 \left[\sum_{\tau=1}^{N} \mathbf{a}_\tau \mathbf{a}_\tau'\right]^{-1} \tag{41}$$

where $\widehat{\sigma}_N^2$ is a consistent estimator of σ^2 (Section 3.3.2) and $\mathbf{a}_\tau$ is here evaluated at $\widehat{\boldsymbol{\theta}}_N$.

3.3.2 Inference on σ^2

Equation (23) suggests that a natural estimator for the nuisance parameter σ^2 can be

$$\widehat{\sigma}_N^2 = \frac{1}{N}\sum_{\tau=1}^{N} u_\tau^2 \tag{42}$$

where u_τ denotes here the working residual (21) computed by using current values of $\widehat{\boldsymbol{\theta}}_N$. An interesting characteristic of such estimator, is that it is not compromised by zeros in the data.

3.3.3 Relationships with other inferential methods

We can check that the estimator $\widehat{\boldsymbol{\theta}}_N$, obtained solving the moment equation (39), can be justified also as a Maximum Likelihood (ML) estimator of $\boldsymbol{\theta}$ when the conditional distribution of the error term $\varepsilon_{t\,j}$ is assumed $Gamma(\sigma^{-2}, \sigma^{-2})$. In fact, denoting

$$\varepsilon_\tau = x_\tau/(\eta_t \; s_j \; \mu_\tau) \tag{43}$$

we have $f_x(x_\tau|\mathcal{F}_{\tau-1}) = f_\varepsilon(\varepsilon_\tau|\mathcal{F}_{\tau-1})\varepsilon_\tau/x_\tau$, so that the portion of the score function relative to $\boldsymbol{\theta}$ is given by

$$\nabla_{\boldsymbol{\theta}}\left[\frac{1}{N}\sum_{\tau=1}^{N}\ln f_x(x_\tau|\mathcal{F}_{\tau-1})\right] = \frac{1}{N}\sum_{\tau=1}^{N}\nabla_{\varepsilon_\tau}[\ln f_\varepsilon(\varepsilon_\tau|\mathcal{F}_{\tau-1}) + \ln\varepsilon_\tau - \ln x_\tau]\nabla_{\boldsymbol{\theta}}\varepsilon_\tau. \tag{44}$$

Replacing the formulations of $\ln f_\varepsilon(\varepsilon_\tau|\mathcal{F}_{\tau-1})$ and $\nabla_{\boldsymbol{\theta}}\varepsilon_\tau$ into (44), we obtain that the ML estimator of $\boldsymbol{\theta}$ solves (39).

The ML estimator of σ^2, still under the $Gamma$ assumption, is however different from

(42). In fact, the portion of the score relative to σ^2 is

$$\nabla_{\sigma^2}\left[\frac{1}{N}\sum_{\tau=1}^{N}\ln f_x(x_\tau|\mathcal{F}_{\tau-1})\right] = \frac{1}{N}\sum_{\tau=1}^{N}\sigma^{-4}\left[\ln\sigma^2+\psi(\sigma^{-2})+\ln\varepsilon_\tau-\varepsilon_\tau+1\right],$$

leading to the score equation

$$\ln\sigma^2+\psi(\sigma^{-2}) = 1+\frac{1}{N}\sum_{\tau=1}^{N}\left[\ln\varepsilon_\tau-\varepsilon_\tau\right], \qquad (45)$$

where ε_τ is computed here using (43) evaluated at $\widehat{\boldsymbol{\theta}}_N$. We note that this equation has to be solved numerically and that, more importantly, cannot be employed with zeros in the data since, in this case, the corresponding values of ε_τ are zero.

3.4 Empirical Application: In Sample Volume Analysis

The empirical application focuses on the analysis of the tickers DIA, QQQQ and SPY in 2002–2006. We consider four variants of the CMEM introduced in Section 3.1 (Equations (3) and (4)):

base CMEM with lag-1 dependence and no asymmetric effects;

asym base CMEM with lag-1 asymmetric effects on both the daily and the intra-daily dynamic components;

intra2 base CMEM with intra-daily autoregressive components of order 2;

asym-intra2 intra2 CMEM with lag-1 asymmetric effects (daily and intra-daily).

The parameter estimates of the daily and intra-daily components are reported in Table 2, together with residual diagnostics. The periodic component, omitted from the table, is expressed in Fourier form (Equation (13)). Also, $\omega^{(\mu)}$ lacks a t-statistic because estimated via expectation targeting by imposing $E(\mu_\tau) = 1$.

Some comments are in order. The parameter estimates of each model are similar across assets, suggesting common behavior in the volume dynamics. We have a high (close to 1) level of daily persistence (measured as $\alpha^{(\eta)}+\beta^{(\eta)}$ in the symmetric, respectively, $\alpha^{(\eta)}+\gamma^{(\eta)}/2+\beta^{(\eta)}$ asymmetric specifications). Contrary to customary values in a typical GARCH(1,1) estimates on daily returns, in the present context $\alpha^{(\eta)}$ is much larger. Intra-daily asymmetric effects are always strongly significant, while daily asymmetric effects are significant for the DIA and SPY tickers only. Their signs are always positive, coherently with the notion that negative past returns have a greater impact on the level of market activity in comparison to the positive ones. The second order intra-daily lag is negative and with a relatively large magnitude, but it is such that the Nelson and Cao (1992) non-negativity condition for the corresponding component is satisfied in all cases, and has the effect of increasing the level of the intra-daily persistence, as can be observed from the column labeled $pers(\mu)$ in table 2. Correspondingly, the less–than–satisfactory performance of serial correlation residual diagnostics (reported in the last columns of table 2) – even with asymmetric effects – is improved when the second order term is included in the dynamic intra-daily component.

4 Intra-daily Volume Forecasting for VWAP Trading

A VWAP replicating strategy is defined as a procedure for splitting a certain number of shares into smaller size orders during the day in the attempt to obtain an average execution price that is close to the daily VWAP. Let the VWAP for day t be defined as

$$\mathsf{VWAP}_t = \frac{\sum_{i=1}^{I_t} v_t(i)\, p_t(i)}{\sum_{i=1}^{I_t} v_t(i)}.$$

where $p_t(i)$ and $v_t(i)$ denote respectively the price and volume of the i-th transaction of day t and I_t is the total number of trades of day t. For a given partition of the trading day into J bins, it is possible to express the numerator of the VWAP as

$$\begin{aligned}\sum_{i=1}^{I_t} v_t(i) p_t(i) &= \sum_{j=1}^{J}\left(\sum_{i\in\mathcal{I}_j} v_t(i)\right)\bar{p}_{t\,j} \\ &= \sum_{j=1}^{J} x_{t\,j}\;\bar{p}_{t\,j},\end{aligned}$$

where $\bar{p}_{t\,j}$ is the VWAP of the j-th bin and $\mathcal{I}_j$ denotes the set of indices of the trades belonging to the j-th bin. Hence,

$$\mathsf{VWAP}_t = \frac{\sum_{j=1}^{J} x_{t\,j}\;\bar{p}_{t\,j}}{\sum_{j=1}^{J} x_{t\,j}} = \sum_{j=1}^{J} \tilde{x}_{t\,j}\;\bar{p}_{t\,j} = \tilde{\mathbf{x}}_{\mathbf{t}}'\bar{\mathbf{p}}_{\mathbf{t}} \tag{46}$$

where $\tilde{x}_{t\,j}$ is the daily proportion of volumes of day t traded in bin j, that is $\tilde{x}_{t\,j} = x_{t\,j}/\sum_{i=1}^{J} x_{t\,i}$. Let $\mathbf{y} = (y_1, \ldots, y_J)$, an order slicing strategy over day t with the same bin intervals. We can define the Average Execution Price as the quantity

$$\mathsf{AEP}_t = \sum_{j=1}^{J} y_j \bar{p}_{t\,j} = \mathbf{y}'\bar{\mathbf{p}}_{\mathbf{t}}$$

where the assumption is made that the traders execute at or *close* to the average price (more on this later). The choice variable being the vector $\mathbf{y}$, we can solve the problem of minimizing the distance between the two outcomes in a mean square error sense, namely

$$\min_{\mathbf{y}} \delta_t = (\tilde{\mathbf{x}}_{\mathbf{t}}'\bar{\mathbf{p}}_{\mathbf{t}} - \mathbf{y}'\bar{\mathbf{p}}_{\mathbf{t}})^2,$$

where, solving the minimization problem leads to the first order conditions

$$\frac{d\delta_t}{d\mathbf{y}} = \mathbf{0} \Rightarrow -2\bar{\mathbf{p}}_\mathbf{t}(\tilde{\mathbf{x}}_\mathbf{t} - \mathbf{y})'\bar{\mathbf{p}}_\mathbf{t} = \mathbf{0}$$

which has a meaningful solution for $\mathbf{y} = \tilde{\mathbf{x}}_\mathbf{t}$, that is when the order slicing sequence for each sub period in the day reproduces exactly the overall relative volume for that sub period. It follows that the VWAP replication problem can be cast as an intra-daily volume proportion forecasting problem: the better we can predict the intra-daily volumes proportions, the better we can track VWAP.

4.1 VWAP Replication Strategies

Following Białkowski *et al.* (2008), we consider two types of VWAP replication strategies: Static and Dynamic. The Static VWAP replication strategy assumes that the order slicing is set before the market opening and it is not revised during the trading day. In the Dynamic VWAP replication strategy scenario on the other hand, order slicing is revised at each new sub period as new intra-daily volumes are observed. Hence, Static VWAP trading is based on volume forecasts conditionally on the previous day while Dynamic VWAP trading is based on volume forecasts conditionally on the previous intra-daily bin.

Let $\hat{x}_{t\,j|t-1}$ be shorthand notation for the prediction of $x_{t\,j}$ conditionally on the previous day full information set $\mathcal{F}_{t-1\,J}$. The *static VWAP* replication strategy is implemented using slices with weights given by

$$\widehat{\widehat{x}}_{t\,j|t-1} = \frac{\hat{x}_{t\,j|t-1}}{\sum_{i=1}^{J} \hat{x}_{t\,i|t-1}} \quad j = 1, .., J\ ,$$

that is the proportion of volumes for bin j is given by predicted volume in bin j divided by the sum of the volume predictions.

Let $\hat{x}_{t\,j|j-1}$ be shorthand notation to denote the prediction of $x_{t\,j}$ conditionally on $\mathcal{F}_{t\,j-1}$. The *Dynamic VWAP* replication strategy is implemented using slices with weights given by

$$\widehat{\widehat{x}}_{t\,j|j-1} = \begin{cases} \frac{\hat{x}_{t\,j|j-1}}{\sum_{i=j}^{J} \hat{x}_{t\,i|j-1}} \left(1 - \sum_{i=1}^{j-1} \widehat{\widehat{x}}_{t\,i|i-1}\right) & j = 1, ..., J-1 \\ \left(1 - \sum_{i=1}^{J-1} \widehat{\widehat{x}}_{t\,i|i-1}\right) & j = J \end{cases}$$

that is, for each intra-daily bin from 1 to $J-1$ the predicted proportion is given by the proportion of 1–step ahead volumes with respect to the sum of the remaining predicted volumes multiplied by the slice proportion left to be traded. On the last period of the day J, the predicted proportion is equal to the remaining part of the slice that needs to be traded.

The evaluation of the performance of the VWAP replication strategies is based upon the intra-daily volume/weights errors and the daily VWAP tracking errors. We consider *vol-*

ume and *slicing* errors defined as

$$e^{\mathsf{v}}_{t\,j} = x_{t\,j} - \hat{x}_{t\,j|\cdot} \quad , \quad e^{\mathsf{s}}_{t\,j} = \tilde{x}_{t\,j} - \widehat{\tilde{x}}_{t\,j|\cdot} \quad ,$$

where $\widehat{x}_{t\,j|\cdot}$ and $\widehat{\tilde{x}}_{t\,j|\cdot}$ denote the volume and slice prediction, respectively, from some VWAP replication and volume forecasting strategy. In order to assess directly the efficiency in replicating the VWAP, we also consider *VWAP tracking errors* defined as

$$e^{\mathsf{VWAP}}_{t} = \left(\frac{\mathsf{VWAP}_t - \widehat{\mathsf{VWAP}}_t}{\mathsf{VWAP}_t} \right) 100,$$

where VWAP_t is the VWAP of day t and $\widehat{\mathsf{VWAP}}_t$ is the realized average execution price obtained using some VWAP replication strategy and volume forecasting method. Both VWAP_t and $\widehat{\mathsf{VWAP}}_t$ are computed using the last recorded price of the j–th bin as a proxy of the average price of the same interval. The VWAP tracking error for day t can be seen as an average of slicing errors within each bin weighted by the relative deviation of the price associated to that bin with respect to the VWAP:

$$e^{\mathsf{VWAP}}_{t} = \left(\sum_{j=1}^{J} (\tilde{x}_{t\,j} - \widehat{\tilde{x}}_{t\,j}) \, \frac{\bar{p}_{t\,j}}{\mathsf{VWAP}_t} \right) 100 = \left(\sum_{j=1}^{J} e^{\mathsf{s}}_{t\,j} \, \frac{\bar{p}_{t\,j}}{\mathsf{VWAP}_t} \right) 100.$$

Note that the deviations of the prices from the daily VWAP add an extra source of noise which can spoil slicing forecasts. In light of this, we recommend evaluating the precision of the forecasts by means of the slicing errors.

4.2 Empirical Application: Out-of-Sample VWAP Prediction

Our empirical application consists of VWAP tracking exercise of the tickers DIA, QQQQ and SPY between January 2005 and December 2006 (502 days, 6526 observations). We track VWAP using turnover predictions from our CMEM specification using both Static and Dynamic VWAP replication strategies based on parameter estimates over the 2002–2004 data. In order to assess the usefulness of the proposed approach with respect of a simple benchmark we also track VWAP using (periodic) Rolling Means (RM), that is the predicted volume for the j–th bin is obtained as the mean over the last 40 days at the same bin. The Rolling Means are used to track VWAP using the Static VWAP replication approach.

Table 3 reports the RMSE of the volume and slicing errors together with asterisks denoting the significance of a Diebold-Mariano test of equal predictive ability with respect to RM using the corresponding loss function. The CMEM Dynamic VWAP Tracking performs best and significantly outperforms the benchmark, followed by the CMEM Static VWAP which generally outperforms the benchmark as well.

Table 3 reports also the results on the RMSE of the VWAP tracking exercise. In this context, the evidence is more mixed: while it is true that the CMEM Dynamic VWAP replication strategy achieves the best out–of–sample performance, statistical significance is less clear cut. It is strong for DIA (and extends to the Static model), but it is less so for QQQ and SPY.

5 Conclusions

In this paper we have suggested a dynamic model with different components which captures the behavior of traded volumes (relative to outstanding shares) viewed from daily and (periodic and non–periodic) intra–daily time perspectives. The ensuing Component Multiplicative Error Model is well suited to be simultaneously estimated by Generalized Method of Moments. The application to three major ETFs shows that both the static and the dynamic VWAP replication strategies generally outperform a naïve method of rolling means for intra-daily volumes.

We would need to extend the analysis to a wider group of tickers to check whether the stylized facts are shared by other classes of assets (e.g. single stocks) and to investigate whether overall market capitalization or the percentage of holdings by institutional investors have a bearing on the characteristics of the estimated dynamics.

The CMEM can be used in other contexts in which intra–daily bins are informative of some periodic features (e.g. volatility, number of trades, average durations) together with an overall dynamic which has components at different frequencies. The periodic component can be more parsimoniously specified by recurring to some shrinkage estimation as in Brownlees and Gallo (2008). Multivariate extensions are also possible (following Cipollini *et al.* (2008) by retrieving the price-volume dynamics mentioned earlier in order to establish a relationship that can be related to the flow of information at different frequencies, separating it from (possibly common) periodic components.

References

Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *The Journal of Finance*, **51**, 169–204.

Berkowitz, S. A., Logue, D. E., and Noser, E. A. J. (1988). The total cost of transactions on the nyse. *The Journal of Finance*, **43**(1), 97–112.

Białkowski, J., Darolles, S., and Fol, G. L. (2008). Improving vwap strategies: A dynamic volume approach. *Journal of Banking and Finance*, **32**, 1709–1722.

Bollerslev, T. and Ghysels, E. (1996). Periodic autoregressive conditional heteroskedasticity. *Journal of Business and Economic Statistics*, **14**, 139–151.

Brownlees, C. T. and Gallo, G. M. (2006). Financial econometric analysis at ultra–high frequency: Data handling concerns. *Computational Statistics and Data Analysis*, **51**, 2232–2245.

Brownlees, C. T. and Gallo, G. M. (2008). Shrinkage estimation of semiparametric multiplicative error models. Technical report, Dipartimento di Statistica "G.Parenti" – Florence.

Chordia, T., Roll, R., and Subrahmanyam, A. (2008). Why has trading volume increased? Technical report, UCLA.

Cipollini, F., Engle, R. F., and Gallo, G. M. (2008). Semiparametric vector MEM. Working Paper Series 1282155, SSRN eLibrary.

Dufour, A. and Engle, R. F. (2000). Time and the price impact of a trade. *The Journal of Finance*, **55**, 2467–2498.

Engle, R., Ferstenberg, R., and Russell, J. (2006a). Measuring and modeling execution cost and risk. Technical report, Stern - NYU.

Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics*, **17**, 425–446.

Engle, R. F., Sokalska, M. E., and Chanda, A. (2006b). Forecasting intraday volatility in the US Equity Market. multiplicative component garch. Technical report, North American Winter Meetings of the Econometric Society, 2007.

Gouriéroux, C., Jasiak, J., and Fol, G. L. (1999). Intra-day market activity. *Journal of Financial Markets*, **2**, 193–226.

Karpoff, J. M. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis*, **22**, 109–126.

Madhavan, A. (2002). Vwap strategies. *Transaction Performance: The Changing Face of Trading Investment Guides Series*, pages 32–38.

Martens, M., Chang, Y.-C., and Taylor, S. J. (2002). A comparison of seasonal adjustment methods when forecasting intraday volatility. *Journal of Financial Research*, **25**(2), 283–299.

Nelson, D. B. and Cao, C. Q. (1992). Inequality constraints in the univariate garch model. *Journal of Business and Economic Statistics*, **10**, 229–235.

Newey, W. K. and McFadden, D. (1994). *Large sample estimation and hypothesis testing*, chapter 36, pages 2111–2245. Elsevier B.V., doi:10.1016/S1573-4412(05)80005-4.

Wooldridge, J. M. (1994). *Estimation and inference for dependent processes*, chapter 45, pages 2639–2738. Elsevier B.V., doi:10.1016/S1573-4412(05)80014-5.

Ticker	Specification	$\omega^{(\eta)}$	$\alpha^{(\eta)}$	$\gamma^{(\eta)}$	$\beta^{(\eta)}$	$\omega^{(\mu)}$	$\alpha_1^{(\mu)}$	$\alpha_2^{(\mu)}$	$\gamma^{(\mu)}$	$\beta^{(\mu)}$	σ	$pers(\mu)$	LB_1	LB_7	LB_{13}
DIA	base	0.025 4.388	0.320 11.835		0.654 22.287	0.268	0.316 33.73			0.416 23.813	0.674	0.732	0.122	0.000	0.000
	asym	0.023 4.302	0.298 10.814	0.025 2.428	0.666 22.945	0.257	0.275 27.513		0.058 6.651	0.439 25.658	0.673	0.743	0.041	0.000	0.000
	intra2	0.011 2.963	0.180 7.154		0.808 29.663	0.060	0.347 35.181	−0.225 −16.257		0.817 51.039	0.671	0.919	0.159	0.839	0.802
	asym-intra2	0.010 2.957	0.158 6.561	0.026 3.049	0.819 32.175	0.057	0.319 30.864	−0.226 −17.643	0.042 7.966	0.830 60.655	0.668	0.926	0.181	0.802	0.805
QQQQ	base	0.019 3.741	0.363 13.156		0.622 21.716	0.286	0.377 42.346			0.337 21.157	0.483	0.714	0.057	0.000	0.000
	asym	0.019 3.772	0.348 12.327	0.014 1.699	0.63 21.985	0.281	0.350 37.285		0.043 6.416	0.348 22.115	0.481	0.674	0.037	0.000	0.000
	intra2	0.001 0.817	0.089 4.656		0.91 45.851	0.031	0.399 44.393	−0.294 −24.835		0.864 86.833	0.481	0.955	0.463	0.236	0.516
	asym-intra2	0.001 0.812	0.090 4.629	0.006 0.886	0.907 45.243	0.031	0.381 41.268	−0.291 −25.228	0.024 6.962	0.866 91.154	0.479	0.954	0.527	0.239	0.518
SPY	base	0.021 4.368	0.410 13.997		0.569 18.489	0.289	0.360 39.476			0.352 21.203	0.533	0.712	0.032	0.000	0.000
	asym	0.022 4.677	0.384 12.931	0.030 3.158	0.579 18.956	0.281	0.321 33.311		0.055 7.520	0.370 22.641	0.532	0.719	0.013	0.000	0.000
	intra2	0.009 2.718	0.213 7.518		0.778 25.828	0.044	0.384 41.261	−0.273 −21.137		0.845 64.616	0.532	0.936	0.631	0.424	0.516
	asym-intra2	0.008 2.846	0.164 6.251	0.023 2.686	0.816 29.999	0.035	0.359 37.233	−0.279 −23.907	0.032 8.576	0.870 88.53	0.531	0.952	0.705	0.237	0.238

Table 2: Parameter Estimates. Sample period 2002 – 2006 (1248 trading days, 13 daily bins, 16224 observations). t-statistics are reported in parenthesis. LB_l denote p-values of the corresponding Ljung-Box statistics at the l-th lag. $pers(\mu)$ indicates estimated persistence of the dynamic intra-daily component.

	DIA			QQQQ			SPY		
	vol	slice	vwap	vol	slice	vwap	vol	slice	vwap
RM	56.68	3.609	1.373	64.83	2.688	1.796	52.28	2.761	1.129
				Static					
base	53.00***	3.632	1.354**	63.73***	2.672***	1.822	50.45***	2.745***	1.110
asym	53.04***	3.632	1.352**	63.71***	2.672***	1.823	50.45***	2.745***	1.110
intra2	53.00***	3.632	1.335***	63.73***	2.672***	1.803	50.45***	2.745***	1.124
asym-intra2	53.04***	3.632	1.319***	63.71***	2.672***	1.771	50.45***	2.747***	1.120
				Dynamic					
base	49.17***	3.595***	1.240***	58.41***	2.649***	1.753	45.39***	2.703***	1.079**
asym	49.11***	3.594***	1.238***	58.37***	2.649***	1.755	45.31***	2.704***	1.082**
intra2	49.07***	3.568***	1.241***	57.89***	2.617***	1.744	45.16***	2.682***	1.095**
asym-intra2	48.91***	3.566***	1.233***	58.03***	2.613***	1.687*	44.98***	2.679***	1.093**

Table 3: Out–of–Sample Volume, Slicing and VWAP tracking Forecasting Results.