UNIVERSITÀ DEGLI STUDI FIRENZE

DISIA
DIPARTIMENTO DI STATISTICA, INFORMATICA, APPLICAZIONI "GIUSEPPE PARENTI"

Go with the Flow: A GAS model for Predicting Intra-daily Volume Shares

Francesco Calvori, Fabrizio Cipollini,
Giampiero M. Gallo

DISIA WORKING PAPER
2014/01

Go with the Flow: A GAS model for Predicting Intra–daily Volume Shares

Francesco Calvori* Fabrizio Cipollini† Giampiero M. Gallo‡

February 10, 2014

Abstract

The Volume Weighted Average Price (VWAP) mixes volumes and prices at intra-daily intervals and is a benchmark measure frequently used to evaluate a trader's performance. Under suitable assumptions, splitting a daily order according to *ex-ante* volume predictions is a good strategy to replicate the VWAP. To bypass possible problems generated by local trends in volumes, we propose a novel Generalized Autoregressive Score (GAS) model for predicting volume shares (relative to the daily total), inspired by the empirical regularities of the observed series (intra-daily periodicity pattern, residual serial dependence). An application to six NYSE tickers confirms the suitability of the model proposed in capturing the features of intra–daily dynamics of volume shares.

Keywords: High Frequency Financial Data, Prediction, Trading Volumes, Volume Shares, VWAP, GAS.

1 Introduction

Managing asset trading transaction costs represents a crucial concern in implementing investment decisions, especially for institutional investors. Such costs include all expenses related to buying or selling a security and consist of several components with different degrees of transparency and manageability (cf. Kissel and Glantz (2003, ch. 1) for a comprehensive discussion). In wanting to execute an order, there is a trade–off between immediacy (which may move the price, the so–called market impact of a trade) and market risk (waiting increases the uncertainty about future prices): both translate into cost components (Dufour and Engle (2000), Hautsch and Ruihong (2012)). As remarked by Brownlees *et al.* (2011), McCulloch and Kazakov (2007) and Konishi (2002), among others, market impact costs are often mitigated by making reference to Volume Weighted Average Price (VWAP) trading strategies, which benchmark the order execution price against the VWAP calculated over a specific period of time (generally one trading day). In practice, VWAP–based trading strategies spread a certain order over the day (reducing instantaneous liquidity demand) with the aim of achieving an average execution price as close as possible to the VWAP. The theoretical justification for a VWAP benchmark comes from Berkowitz *et al.* (1988), who reckon that an unbiased estimate of the prices involving all sort of traders during the day is a weighted average of transaction prices (Madhavan (2002)).

*DiSIA, Università di Firenze, Italy. e-mail: calvori@disia.unifi.it

†DiSIA, Università di Firenze, Italy. e-mail: cipollini@disia.unifi.it

‡Corresponding author: Dipartimento di Statistica, Informatica, Applicazioni "G. Parenti" (DiSIA), Università di Firenze, Viale G.B. Morgagni, 59 - 50134 Firenze – Italy; e-mail: gallog@disia.unifi.it; ph. +39 055 2751 591. The authors thank the **MIUR PRIN MISURA Project**, 2013-2015, for financial support.

The VWAP for day t is expressed as

$$\mathsf{VWAP}_t := \frac{\sum_{j=1}^{J_t} v_t(j)\, p_t(j)}{\sum_{j=1}^{J_t} v_t(j)} = \frac{\sum_{i=1}^{I} v_{t,i}\overline{p}_{t,i}}{\sum_{i=1}^{I} v_{t,i}} = \sum_{i=1}^{I} w_{t,i}\overline{p}_{t,i}. \tag{1}$$

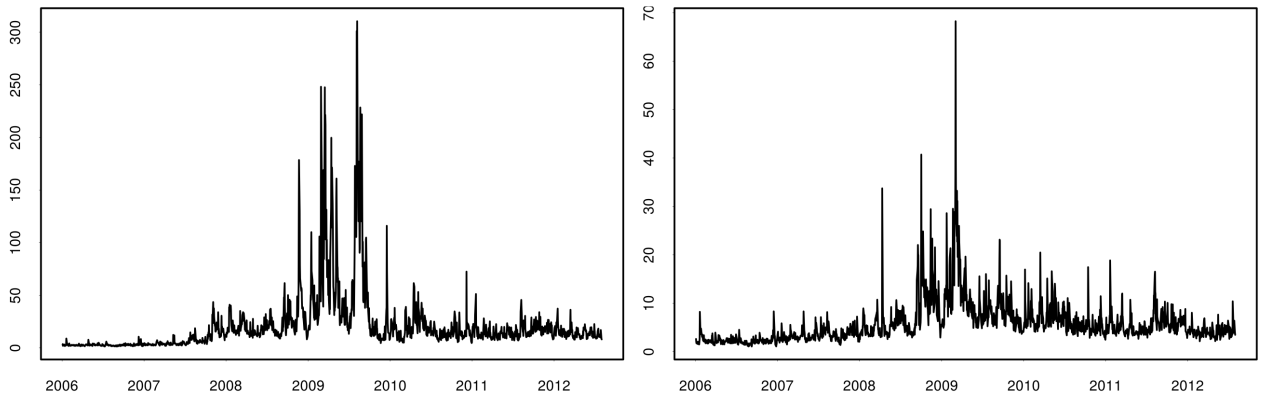
The first term considers all J_t transactions within the t-th day, with $p_t(j)$ and $v_t(j)$ the price and the volume of the j-th transaction of the day, respectively. The further expressions aggregate the same quantity as an average across I equally spaced (in calendar time) bins ($v_{t,i}$, $w_{t,i}$ and $\overline{p}_{t,i}$ are the volume, the volume share and the VWAP, respectively, within the i-th bin).

Considering that prices are substantially unpredictable, one way of interpreting Equation (1) is to concentrate on predicting volumes (at regular bins, and the daily overall value) (cf. Kissel and Glantz (2003, ch. 11) and Białkowski *et al.* (2008)). The latter authors decompose the intra-day turnover (volume relative to outstanding shares, a common practice to stabilize the series, cf. Darolles and Le Fol (2003)) into two parts: one reflects the whole market evolution and aims at capturing the periodic part of the dynamics (represented via a linear latent factor model); the second, stock specific, is derived as a residual and modeled with ARMA or SETAR dynamics. Such an approach is justified with the observation that the periodic dynamics is common across stocks, whereas the remaining non-periodic movements are stock specific. Brownlees *et al.* (2011) model turnover via a Multiplicative Error Model (extending Engle (2002)) with three components (daily, intra–daily periodic and intra–daily non periodic). While their model reproduces the empirical regularities of the observed series, it rests on the adequacy of the chosen dynamics of each component, and may be encumbered by the presence of local trends for the volumes. Such an issue is implicitly recognized by Białkowski *et al.* (2008, Section 3.2), when they choose to analyze their data in 1-month stretches because the factor model proposed is inadequate over longer periods.

Since Equation (1) indicates that volume shares are a crucial ingredient for a VWAP–based trading strategy, in this paper we suggest to model $w_{t,i}$ directly, exploiting several advantages: by construction, they do not have trending patterns, nor do they require a normalization of volumes by the number of outstanding stocks; moreover, they avoid the need to formulate the dynamics of the daily component (cf. the extensive discussion in Section 2). The downside is that $w_{t,i}$ is bounded between zero and one, it has an adding–up constraint and cannot be evaluated before the daily market closure. The novel model we propose belongs to the class of the Generalized Autoregressive Score (GAS) models by Creal *et al.* (2012) (also known as Dynamic Score Models, Harvey (2013)), a fairly general and flexible approach for non-linear, non-Gaussian *observation driven* time series models. In the present context, a GAS approach views the volume shares as linked to the time–varying parameters of a Dirichlet density function. Within this framework we are able to derive several score–based specification tests and to address the issue of how to achieve a parsimonious parameterization. We build the relevant series from tick–by–tick data on six tickers from the NYSE and we model volume–share dynamics achieving a satisfactory in-sample performance and a significant gain when considering out–of–sample forecasts over a more naïve approach.

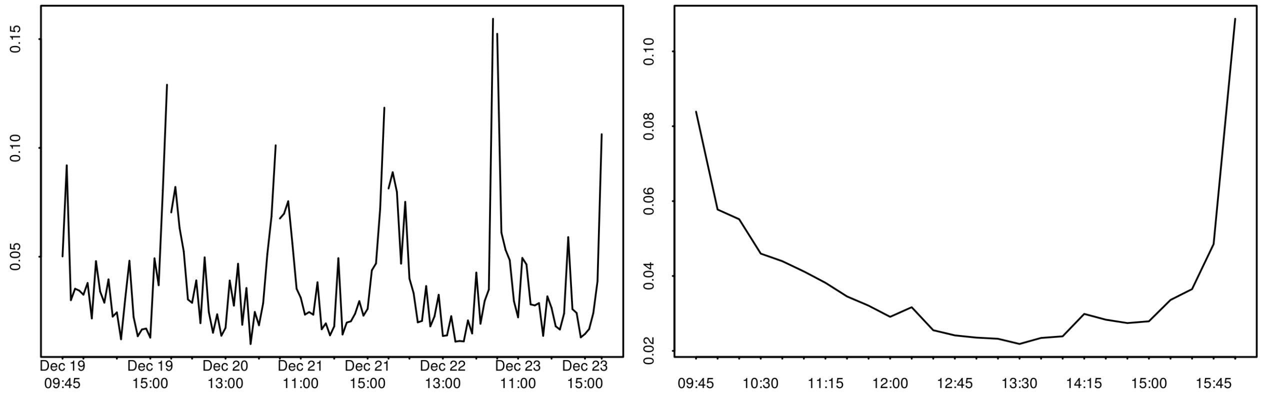
The paper is structured as follows: in Section 2 we discuss the modeling strategy; in Section 3 we present model inference and introduce some model diagnostics tools. In Section 4 we address the issues related to the estimation of several models in the empirical application, the need for a parsimonious parameterization of the model and volume shares forecasting. In Section 5 we notice the difficulty of evaluating share predictions within a VWAP measure since prices are not predictable: standard MSE measures are deceiving in this context because they contain factors which may cancel each other out. As a consequence, we suggest a proper loss function which is illustrated empirically. Section 6 concludes.

Figure 1: Time series of the daily turnover.



(a) Citigroup. (b) General Electric.

Figure 2: Ford ticker: Time series and intra–daily periodicity plots.



(a) Raw series $w_{t,i}$: one week of data. (b) Bin averages (January 2006 – July 2012).

2 Modeling Strategy

The modeling strategy is driven by some empirical stylized facts. The first one is the presence of local trends in volumes, as shown by the two examples in Figure 1. After a substantially stable period, the daily turnover of General Electric (GE) increases sharply from half 2008 until the beginning of 2009; since this moment, it tends to decrease quite regularly but never returning to the pre-2008 levels. The pattern of Citigroup (C) is somewhat similar although markedly more irregular, with frequent and abrupt changes of the mean level; moreover, the average value around which it moves since 2010 is several times higher than the pre-2008 levels. Such considerations indicate that even normalizing by the shares outstanding, working on volumes requires an appropriate formulation of the daily component (around which the intradaily level moves) and that its structure must be enough complex to reproduce such movements. In order to bypass this issue, in this paper we model volume shares directly. The second stylized fact is the existence of a strong intra-daily periodic pattern. Figure 2a, reproducing a typical one week profile of the shares, evidences an apparent periodic component which seems to dominate the dynamic behavior of the series. When such a component is roughly estimated over a larger period (bin averages), it is no surprise that the typical intra–daily U–shaped pattern of market activity emerges (cf. Figure 2b): this feature must be transferred to the model, in terms of some periodic component. A key issue of this paper is whether the data suggest some additional predictable dynamics other than periodic.

Modeling daily bin shares[1] $\boldsymbol{w}_t = (w_{t,1}; \ldots; w_{t,I})$ must take two characteristics into account: they are defined on a simplex (namely $w_{t,i} \geq 0$ and $\mathbb{1}'\boldsymbol{w}_t = 1$); they are unknown until the market closure, since the total daily volume can be computed only after the last transaction of the day. As a consequence, a sensible assumption, maintained throughout the paper, is to take $\boldsymbol{w}_t|\mathcal{F}_{t-1}$ to be Dirichlet distributed, in symbols $\boldsymbol{w}_t|\mathcal{F}_{t-1} \sim \mathrm{Dir}(\boldsymbol{\alpha}_t)$, where $\boldsymbol{\alpha}_t = (\alpha_{t,1}; \ldots; \alpha_{t,I})$ $(\alpha_{t,i} > 0)$. The log probability density function (pdf), its derivative with respect to $\boldsymbol{\alpha}_t$ and the corresponding information matrix, are given, respectively, by

$$l_t = \ln f(\boldsymbol{w}_t|\mathcal{F}_{t-1}) = \ln \Gamma(\alpha_{t,0}) - \sum_{i=1}^{I} \ln \Gamma(\alpha_{t,i}) + \sum_{i=1}^{I} (\alpha_{t,i} - 1) \ln w_{t,i} \tag{2}$$

$$\boldsymbol{\nabla}_t^{(\alpha)} = \frac{\partial \ln f(\boldsymbol{w}_t|\mathcal{F}_{t-1})}{\partial \boldsymbol{\alpha}_t} = \{\psi(\alpha_{t,0}) - \psi(\alpha_{t,i}) + \ln w_{t,i} : i = 1, \ldots, I\} \tag{3}$$

$$\boldsymbol{\mathcal{I}}_t^{(\alpha)} = E\left(\boldsymbol{\nabla}_t^{(\alpha)} \boldsymbol{\nabla}_t^{(\alpha)\prime} | \mathcal{F}_{t-1}\right) = \mathrm{diag}\left\{\dot{\psi}(\alpha_{t,i}) : i = 1, \ldots, I\right\} - \dot{\psi}(\alpha_{t,0}) \mathbb{1}\mathbb{1}' \tag{4}$$

where $\alpha_{t,0} = \mathbb{1}'\boldsymbol{\alpha}_t$, and $\psi(\cdot)$, $\dot{\psi}(\cdot)$ denote, respectively, the digamma and trigamma functions (cf. $\ddot{\psi}(\cdot)$ for the quadrigamma function below).

In order to capture the dynamics of the time series under analysis, $\boldsymbol{\alpha}_t$ is made to depend on $\mathcal{F}_{t-1}$, the information set at the end of day $t-1$, and on a $\boldsymbol{\theta}$, a p vector of parameters, where an appropriate formulation of $\boldsymbol{\alpha}_t$ retrieves some empirical features discussed before. The GAS modeling framework (Creal *et al.* (2012)) has the necessary flexibility to build a model where components with different dynamics may or may not be present. We assume that $\alpha_{t,i} = \exp(\pi_i + \beta_{t,i})$, where π_i is the value of the periodic component (a Fourier sine/cosine function as in Brownlees *et al.* (2011)) and the time-varying parameter $\boldsymbol{\beta}_t = (\beta_{t,1}, \ldots, \beta_{t,I})$ evolves according to[2]

$$\boldsymbol{\beta}_t = \boldsymbol{A}\boldsymbol{s}_{t-1} + \boldsymbol{B}\boldsymbol{\beta}_{t-1}. \tag{5}$$

$\boldsymbol{s}_t$ is the conditional *scaled score* $\boldsymbol{s}_t = \boldsymbol{S}_t \boldsymbol{\nabla}_t^{(\beta)}$ where

$$\boldsymbol{\nabla}_t^{(\beta)} = \frac{\partial \ln f(\boldsymbol{w}_t|\mathcal{F}_{t-1})}{\partial \boldsymbol{\beta}_t} = \mathrm{diag}(\boldsymbol{\alpha}_t) \boldsymbol{\nabla}_t^{(\alpha)}$$

is the conditional *score* and

$$\boldsymbol{S}_t = \boldsymbol{S}(\mathcal{F}_{t-1})$$

is a scaling matrix; $\boldsymbol{A}$ and $\boldsymbol{B}$ are (I, I) matrices depending on some parameters (included into $\boldsymbol{\theta}$). Denoting the information matrix (corresponding to the conditional score defined above) as

$$\boldsymbol{\mathcal{I}}_t = E\left(\boldsymbol{\nabla}_t^{(\beta)} \boldsymbol{\nabla}_t^{(\beta)\prime} | \mathcal{F}_{t-1}\right) = \mathrm{diag}(\boldsymbol{\alpha}_t)\, \boldsymbol{\mathcal{I}}_t^{(\alpha)}\, \mathrm{diag}(\boldsymbol{\alpha}_t),$$

we may compare four different versions of $\boldsymbol{S}_t$:

1. the identity matrix I (no rescaling);
2. $\mathrm{ddiag}(\boldsymbol{\mathcal{I}}_t)^{-1/2}$ (leading to an $\boldsymbol{s}_t$ in which each component of the score is standardized)[3];
3. $\boldsymbol{\mathcal{I}}_t^{-1}$ (similar to the rescaling matrix of the score in the BHHH estimation update);
4. $\boldsymbol{\mathcal{I}}_t^{-1/2}$, the low-triangular matrix in the Cholesky decomposition of $\boldsymbol{\mathcal{I}}_t^{-1}$ (so that $\boldsymbol{s}_t$ is the standardized score).

[1]If $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K$ are matrices with the same number of columns, then $(\boldsymbol{x}_1; \ldots; \boldsymbol{x}_K)$ indicates the matrix obtained stacking the matrices $\boldsymbol{x}_i$'s columnwise.

[2]More lags could be added for a generic GAS(p, q).

[3]$\mathrm{ddiag}(\boldsymbol{M})$ (where $\boldsymbol{M}$ is a square matrix) is the diagonal matrix having $\mathrm{diag}(\boldsymbol{M})$ as main diagonal.

It is interesting to note that the structure of the Dirichlet pdf allows for relatively simple expressions of the scaled score for all the versions above, avoiding an explicit inversion of the non-diagonal scaling matrices $\boldsymbol{\mathcal{I}}_t$ (or, for that matter, of $\boldsymbol{\mathcal{I}}_t^{1/2}$ as well). For example, let us consider $\boldsymbol{s}_t = \boldsymbol{\mathcal{I}}_t^{-1}\,\boldsymbol{\nabla}_t^{(\beta)}$ for the third choice of $\boldsymbol{S}_t$ above; the scaled score is given by

$$\boldsymbol{s}_t = \left\{ \frac{b_t + \nabla_{t,i}^{(\alpha)}}{\alpha_{t,i}\dot{\psi}(\alpha_{t,i})} : i = 1, \ldots, I \right\},$$

where

$$b_t = \frac{\sum_{i=1}^{I} \nabla_{t,i}^{(\alpha)} / \dot{\psi}(\alpha_{t,i})}{1/\dot{\psi}(\alpha_{t,0}) - \sum_{i=1}^{I} 1/\dot{\psi}(\alpha_{t,i})}. \tag{6}$$

Such a choice provides the best results in our applications – see Section 4; similar expressions for $\boldsymbol{s}_t$ are derived in the other cases, but details are available upon request.

3 Model Inference and Diagnostics

Since GAS models are fully parametric, estimation can be performed via Maximum Likelihood (ML). The score function may be derived from the log–likelihood function $l = \sum_{t=1}^{T} l_t$ (cf. (2) above), as

$$\boldsymbol{\nabla}^{(\theta)} = \sum_{t=1}^{T} \boldsymbol{\nabla}_t^{(\theta)} = \sum_{t=1}^{T} \left(\frac{\partial \boldsymbol{\pi}'}{\partial \boldsymbol{\theta}} + \frac{\partial \boldsymbol{\beta}_t'}{\partial \boldsymbol{\theta}} \right) \boldsymbol{\nabla}_t^{(\beta)}. \tag{7}$$

The ensuing information matrix

$$\lim_{T\to\infty} \left[T^{-1} \sum_{t=1}^{T} E\left[\left(\frac{\partial \boldsymbol{\pi}'}{\partial \boldsymbol{\theta}} + \frac{\partial \boldsymbol{\beta}_t'}{\partial \boldsymbol{\theta}} \right) \boldsymbol{\mathcal{I}}_t \left(\frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\theta}'} + \frac{\partial \boldsymbol{\beta}_t}{\partial \boldsymbol{\theta}'} \right) \right] \right]$$

can be estimated by means of the average cross-product of gradients,

$$T^{-1} \sum_{t=1}^{T} \widehat{\boldsymbol{\nabla}}_t^{(\theta)} \widehat{\boldsymbol{\nabla}}_t^{(\theta)\prime}$$

evaluated at the ML estimate $\widehat{\boldsymbol{\theta}}$.

The derivative of $\boldsymbol{s}_t$ is important for the analytical expression of the score function $\boldsymbol{\nabla}^{(\theta)}$. In fact, the intricate point in calculating (7) is the partial derivative

$$\frac{\partial \boldsymbol{\beta}_t'}{\partial \boldsymbol{\theta}} = \frac{\partial \boldsymbol{s}_{t-1}'}{\partial \boldsymbol{\theta}} \boldsymbol{A}' + \frac{\partial \boldsymbol{\beta}_{t-1}'}{\partial \boldsymbol{\theta}} \boldsymbol{B}' + \begin{pmatrix} \boldsymbol{s}_{t-1}' \frac{\partial \boldsymbol{A}'}{\partial \boldsymbol{\theta_1}} \\ \vdots \\ \boldsymbol{s}_{t-1}' \frac{\partial \boldsymbol{A}'}{\partial \boldsymbol{\theta_p}} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\beta}_{t-1}' \frac{\partial \boldsymbol{B}'}{\partial \boldsymbol{\theta_1}} \\ \vdots \\ \boldsymbol{\beta}_{t-1}' \frac{\partial \boldsymbol{B}'}{\partial \boldsymbol{\theta_p}} \end{pmatrix},$$

because of its dependence on the derivative (lagged),

$$\frac{\partial \boldsymbol{s}_t'}{\partial \boldsymbol{\theta}} = \left(\frac{\partial \boldsymbol{\pi}'}{\partial \boldsymbol{\theta}} + \frac{\partial \boldsymbol{\beta}_t'}{\partial \boldsymbol{\theta}} \right) \operatorname{diag}(\boldsymbol{\alpha}_t) \frac{\partial \boldsymbol{s}_t'}{\partial \boldsymbol{\alpha}_t}.$$

Suppressing the index t for notational simplicity, we can show that the derivative $\partial \boldsymbol{s}'/\partial \boldsymbol{\alpha}$ has elements

$$\frac{\partial s_i}{\partial \alpha_j} = \frac{b\ddot{\psi}(\alpha_0)/\dot{\psi}^2(\alpha_0) - c_j \alpha_j}{b_D \alpha_i \dot{\psi}(\alpha_i)} - \delta_{ij} \left(c_i + \frac{s_i + 1}{\alpha_i} \right)$$

where $c_i = s_i \ddot{\psi}(\alpha_i)/\dot{\psi}(\alpha_i)$ and δ_{ij} is 1 if $i = j$ and 0 otherwise.

Estimation can be complemented with adequate diagnostic tools in order: a) to ascertain (*ex-ante*) the possible presence of some predictable dynamics ($\boldsymbol{\beta}_t \neq \mathbf{0}$) in addition to the periodic component; and b) to check (*ex-post*) if the estimated model adequately captures all dynamic features of the data (mirroring residual diagnostics analysis). However, in either case, a traditional approach would revolve around the concept that estimation *residuals* should not contain any useful information for prediction. In the present context, the definition of a residual is not trivial, given that the model assumptions do not isolate an explicit error term entering either in an additive or multiplicative fashion. According to the GAS modeling philosophy, the attention is shifted to the behavior of the score, which provides the base of different tests.

The GAS Lagrange Multiplier is a test (Calvori *et al.* (2013)) where, under the null, no dynamics other than periodic is present. More formally, the null hypothesis states that $\boldsymbol{\beta}_t = \mathbf{0}$ against a $\boldsymbol{\beta}_t$ driven only by the lagged scores; specifically, in order to avoid the well–known identification issue (cf. Section 2.1.1 in Bollerslev *et al.* (1994)), under the alternative we assume only GAS$(p, 0)$ effects. To fix ideas, consider Equation (5) constraining $\boldsymbol{B} = \mathbf{0}$, with $\boldsymbol{A} = \mathbf{0}$ under the null hypothesis and $\boldsymbol{A} \neq \mathbf{0}$ under the alternative. The test statistic is given by

$$LM_{GAS} = T\widehat{\overline{\boldsymbol{\nabla}}}^{(\theta_0)\prime} \widehat{\boldsymbol{D}}^{-1} \widehat{\overline{\boldsymbol{\nabla}}}^{(\theta_0)} \xrightarrow{d} \chi^2(df), \tag{8}$$

where

$$\widehat{\boldsymbol{D}} = T^{-1} \sum_{t=1}^{T} \widehat{\boldsymbol{\nabla}}_t^{(\theta_0)} \widehat{\boldsymbol{\nabla}}_t^{(\theta_0)\prime};$$

$\widehat{\overline{\boldsymbol{\nabla}}}^{(\theta_0)}$ denotes the average score under the alternative evaluated at the maximum likelihood estimate under the null; df is the number of parameters set to zero under the null, but not under the alternative. Such a test provides evidence as of whether the vector $\boldsymbol{\alpha}_t$ would not depend on t and each of its components be a deterministic function of the time of day (i-th bin). Rejecting the null points to the presence of additional latent dynamics.

The issue of adequately specifying the model dynamics entails the lack of serial correlation in some objects related to estimation, to be suitably defined; following the same spirit, therefore, one can think of score-based specification tests of autocorrelation adapted to the present context for both *ex-ante* and *ex-post* diagnostics.

A score based test (Newey (1985), White (1987)) aims at testing if some moment conditions on $\boldsymbol{\nabla}_t^{(\theta)}$, implied by the model under correct specification, are satisfied by the data. In principle, capturing the dynamics implies that the components of the score are uncorrelated with one another; therefore, among possible choices, it is reasonable to consider the two moment conditions

$$E\left(\boldsymbol{\nabla}_t^{(\theta)\prime} \boldsymbol{\nabla}_{t-1}^{(\theta)} | \mathcal{F}_{t-1}\right) = 0 \tag{9}$$

$$E\left(\boldsymbol{\nabla}_t^{(\theta)} \odot \boldsymbol{\nabla}_{t-1}^{(\theta)} | \mathcal{F}_{t-1}\right) = \mathbf{0} \tag{10}$$

where $\odot$ indicates the element-by-element product. Denoting the argument of the expectations (9) and (10) with a common symbol $\boldsymbol{m}_t$ (thus a scalar or a p vector, respectively), following White (1987), the corresponding tests for model misspecification can be based on the chi-square statistic

$$ST(df) = T\widehat{\overline{\boldsymbol{m}}}' \widehat{\boldsymbol{V}}^{-1} \widehat{\overline{\boldsymbol{m}}} \xrightarrow{d} \chi^2(df) \tag{11}$$

where:

$$\widehat{\overline{\boldsymbol{m}}} = T^{-1} \sum_{t=1}^{T} \widehat{\boldsymbol{m}}_t$$

is an estimate of the corresponding expectation;

$$\widehat{\boldsymbol{V}} = T^{-1}\left(\widehat{\boldsymbol{M}}'\widehat{\boldsymbol{M}} - \widehat{\boldsymbol{M}}'\widehat{\boldsymbol{\nabla}}(\widehat{\boldsymbol{\nabla}}'\widehat{\boldsymbol{\nabla}})^{-1}\widehat{\boldsymbol{\nabla}}'\widehat{\boldsymbol{M}}\right)$$

(with $\widehat{\boldsymbol{M}} = (\widehat{\boldsymbol{m}}_1, \ldots, \widehat{\boldsymbol{m}}_T)'$ and $\widehat{\boldsymbol{\nabla}} = (\widehat{\boldsymbol{\nabla}}_1^{(\theta)}, \ldots, \widehat{\boldsymbol{\nabla}}_T^{(\theta)})'$) is a consistent estimator of the asymptotic variance matrix of $\sqrt{T}\widehat{\overline{\boldsymbol{m}}}$; df is the size of $\boldsymbol{m}_t$ (1 for (9) and p for (10)). Under general conditions (see White (1987), Davidson and MacKinnon (1990)), a regression based test statistic, asymptotically equivalent to $ST(df)$, can be computed as T times the R^2 of the regression of the unit constant on the vector $(\widehat{\boldsymbol{\nabla}}_t^{(\theta)}; \widehat{\boldsymbol{m}}_t)$.

4 Empirical Application

The model defined in Section 2 is applied to the volume shares, computed every 15 minutes (26 bins per day) between January 3, 2006 to July 31, 2012, on the following tickers: ANF (Abercrombie & Fitch), BAC (Bank of America), C (Citigroup), F (Ford Motor), GE (General Electric), JNJ (Johnson & Johnson). The tick–by–tick data used for computing the shares come from the TAQ database on trades and are cleaned according to Brownlees and Gallo (2006); days lacking one or more bins (because of trading halts or anticipated closures) are removed from the analysis.

4.1 Specification, Estimation, and Diagnostics

We consider the model defined in Section 2 with the following alternative parameter specifications for $\boldsymbol{\beta}_t = \boldsymbol{A}\boldsymbol{s}_{t-1} + \boldsymbol{B}\boldsymbol{\beta}_{t-1}$ (Equation (5)):

$$\mathcal{M}_0 : \boldsymbol{\beta}_t = \boldsymbol{0} \tag{12}$$

$$\mathcal{M}_1 : \boldsymbol{\beta}_t = a\boldsymbol{s}_{t-1} + b\boldsymbol{\beta}_{t-1} \tag{13}$$

$$\mathcal{M}_2 : \boldsymbol{\beta}_t = \text{diag}(a; \ldots; a; a_{I-1}; a_I)\boldsymbol{s}_{t-1} + b\boldsymbol{\beta}_{t-1} \tag{14}$$

$$\mathcal{M}_3 : \boldsymbol{\beta}_t = \text{diag}(\boldsymbol{a})\boldsymbol{s}_{t-1} + b\boldsymbol{\beta}_{t-1}. \tag{15}$$

$\mathcal{M}_0$ corresponds to a model including only the periodic component (i.e. no GAS effects). The other specifications involve GAS dynamics with two polar cases, $\mathcal{M}_1$ and $\mathcal{M}_3$, and an intermediate model $\mathcal{M}_2$: $\mathcal{M}_1$ has the same parameters across bins; $\mathcal{M}_3$ allows for parameters all different across bins for the score component; $\mathcal{M}_2$ builds on the considerations that the dynamics toward the end of the day may be more adequately captured by allowing the score parameters of the last two bins to be different from the remaining ones. The suitability of either model must be empirically evaluated by means of hypothesis testing and diagnostics, together with an evaluation of the forecasting performance.

In addition to the periodic component, a highly significant non-periodic dynamics is present in the data, as shown by the results of the LM_{GAS}, $ST(1)$ and $ST(p)$ specification tests introduced in Section 3 (Table 1), fully justifying the introduction of the time-varying parameter $\boldsymbol{\beta}_t$.

Table 2 reports the GAS coefficients in Equation (5) for the $\mathcal{M}_1$ parameterization (Equation (13)). The results are remarkably similar across tickers: b, which can be interpreted as a *persistence* parameter, is above 0.99 in all cases (but smaller than 1, ensuring stationarity); a, the parameter of the lagged score, ranges from 0.02 to 0.03. In all cases such parameters are largely significant, as a consequence of small standard errors (the total number of observations is fairly high). The $\mathcal{M}_3$ parameterization of Equation (15) (Table 4), indicates that the a_i parameters tend to be relatively stable across bins around 0.02, increasing to about 0.04 for the last bin; the b estimates are again about 0.99. Such a pattern is confirmed by the estimation of the more parsimonious model $\mathcal{M}_2$ of Equation (14), where just the last two bins are allowed to have their own

Table 1: *Ex-ante* detection of non periodic dynamics. P-values of specification tests on 15 minute data. LM_{GAS} denotes the Lagrange Multiplier test for no GAS effects (Equation (8)); $ST(1)$ and $ST(p)$ indicate the score based specification tests (Equation (11)) derived from the moment Equations (9) and (10), respectively. The reference model includes the periodic component only ($\mathcal{M}_0$).

Ticker	LM_{GAS}	$ST(1)$	$ST(p=26)$
ANF	0.0000	0.0000	0.0000
BAC	0.0000	0.0000	0.0000
C	0.0000	0.0000	0.0000
F	0.0000	0.0000	0.0000
GE	0.0000	0.0000	0.0000
JNJ	0.0000	0.0000	0.0000

Table 2: Estimates of the GAS parameters a and b in Model $\mathcal{M}_1$: $\boldsymbol{\beta}_t = a\boldsymbol{s}_{t-1} + b\boldsymbol{\beta}_{t-1}$.

Coeff.	ANF	BAC	C	F	GE	JNJ
a	0.0230	0.0247	0.0314	0.0253	0.0300	0.0215
	(0.0009)	(0.0008)	(0.0008)	(0.0008)	(0.0009)	(0.0006)
b	0.9926	0.9945	0.9945	0.9959	0.9901	0.9954
	(0.0007)	(0.0005)	(0.0004)	(0.0003)	(0.0006)	(0.0004)

Table 3: Estimates of the GAS parameters a, a_{I-1}, a_I and b in Model $\mathcal{M}_2$: $\boldsymbol{\beta}_t = diag(a;\ldots;a;a_{I-1};a_I)\boldsymbol{s}_{t-1} + b\boldsymbol{\beta}_{t-1}$.

Coeff.	ANF	BAC	C	F	GE	JNJ
a_1	0.0194	0.0216	0.0301	0.0240	0.0242	0.0147
	(0.0010)	(0.0007)	(0.0007)	(0.0008)	(0.0008)	(0.0005)
a_{I-1}	0.0295	0.0345	0.0438	0.0343	0.0445	0.0241
	(0.0041)	(0.0032)	(0.0039)	(0.0039)	(0.0041)	(0.0026)
a_I	0.0408	0.0405	0.0420	0.0371	0.0491	0.0334
	(0.0029)	(0.0018)	(0.0019)	(0.0024)	(0.0018)	(0.0012)
b	0.9889	0.9959	0.9948	0.9958	0.9906	0.9946
	(0.0010)	(0.0004)	(0.0004)	(0.0003)	(0.0006)	(0.0004)

Table 4: Estimates of the GAS parameters $a_i, i = 1, \ldots, 26$ and b in Model $\mathcal{M}_3$: $\boldsymbol{\beta}_t = diag(\boldsymbol{a})\boldsymbol{s}_{t-1} + b\boldsymbol{\beta}_{t-1}$.

Coeff.	ANF	BAC	C	F	GE	JNJ
a_1	0.0189	0.0172	0.0229	0.0258	0.0204	0.0189
	(0.0022)	(0.0016)	(0.0017)	(0.0019)	(0.0018)	(0.0019)
a_2	0.0184	0.0174	0.0268	0.0227	0.0235	0.0107
	(0.0031)	(0.0025)	(0.0028)	(0.0026)	(0.0032)	(0.0021)
a_3	0.0211	0.0165	0.0271	0.0237	0.0233	0.0123
	(0.0034)	(0.0024)	(0.0025)	(0.0028)	(0.0027)	(0.0023)
a_4	0.0155	0.0192	0.0293	0.0209	0.0190	0.0114
	(0.0038)	(0.0033)	(0.0033)	(0.0031)	(0.0034)	(0.0028)
a_5	0.0181	0.0149	0.0299	0.0224	0.0182	0.0107
	(0.0041)	(0.0030)	(0.0033)	(0.0034)	(0.0034)	(0.0027)
a_6	0.0196	0.0172	0.0285	0.0220	0.0206	0.0115
	(0.0047)	(0.0032)	(0.0036)	(0.0034)	(0.0042)	(0.0028)
a_7	0.0212	0.0164	0.0248	0.0220	0.0181	0.0102
	(0.0052)	(0.0030)	(0.0031)	(0.0035)	(0.0042)	(0.0028)
a_8	0.0245	0.0214	0.0287	0.0204	0.0199	0.0070
	(0.0058)	(0.0037)	(0.0036)	(0.0035)	(0.0041)	(0.0027)
a_9	0.0257	0.0172	0.0288	0.0201	0.0164	0.0063
	(0.0056)	(0.0037)	(0.0039)	(0.0037)	(0.0041)	(0.0029)
a_{10}	0.0155	0.0166	0.0290	0.0200	0.0196	0.0093
	(0.0053)	(0.0037)	(0.0041)	(0.0037)	(0.0047)	(0.0032)
a_{11}	0.0144	0.0170	0.0285	0.0204	0.0251	0.0098
	(0.0051)	(0.0036)	(0.0041)	(0.0034)	(0.0054)	(0.0033)
a_{12}	0.0173	0.0185	0.0300	0.0216	0.0206	0.0110
	(0.0059)	(0.0039)	(0.0045)	(0.0044)	(0.0054)	(0.0034)
a_{13}	0.0161	0.0225	0.0318	0.0238	0.0251	0.0102
	(0.0057)	(0.0047)	(0.0048)	(0.0050)	(0.0059)	(0.0035)
a_{14}	0.0191	0.0207	0.0295	0.0211	0.0249	0.0124
	(0.0059)	(0.0043)	(0.0047)	(0.0046)	(0.0058)	(0.0041)
a_{15}	0.0164	0.0205	0.0319	0.0228	0.0246	0.0121
	(0.0059)	(0.0042)	(0.0050)	(0.0042)	(0.0052)	(0.0036)
a_{16}	0.0186	0.0235	0.0338	0.0213	0.0211	0.0136
	(0.0061)	(0.0048)	(0.0055)	(0.0046)	(0.0054)	(0.0038)
a_{17}	0.0179	0.0240	0.0343	0.0231	0.0221	0.0135
	(0.0054)	(0.0045)	(0.0051)	(0.0043)	(0.0049)	(0.0041)
a_{18}	0.0134	0.0256	0.0345	0.0245	0.0264	0.0145
	(0.0052)	(0.0045)	(0.0049)	(0.0046)	(0.0048)	(0.0037)
a_{19}	0.0134	0.0246	0.0340	0.0235	0.0262	0.0186
	(0.0048)	(0.0039)	(0.0043)	(0.0040)	(0.0047)	(0.0037)
a_{20}	0.0190	0.0229	0.0357	0.0255	0.0271	0.0187
	(0.0059)	(0.0037)	(0.0047)	(0.0040)	(0.0046)	(0.0036)
a_{21}	0.0194	0.0254	0.0331	0.0239	0.0285	0.0191
	(0.0054)	(0.0042)	(0.0044)	(0.0044)	(0.0051)	(0.0037)
a_{22}	0.0212	0.0281	0.0381	0.0280	0.0304	0.0199
	(0.0059)	(0.0046)	(0.0053)	(0.0050)	(0.0054)	(0.0040)
a_{23}	0.0237	0.0264	0.0416	0.0293	0.0323	0.0214
	(0.0054)	(0.0039)	(0.0050)	(0.0047)	(0.0050)	(0.0035)
a_{24}	0.0235	0.0319	0.0385	0.0319	0.0320	0.0226
	(0.0051)	(0.0042)	(0.0046)	(0.0045)	(0.0047)	(0.0035)
a_{25}	0.0295	0.0335	0.0442	0.0339	0.0438	0.0238
	(0.0045)	(0.0035)	(0.0043)	(0.0041)	(0.0045)	(0.0029)
a_{26}	0.0410	0.0393	0.0422	0.0362	0.0483	0.0331
	(0.0030)	(0.0017)	(0.0020)	(0.0024)	(0.0018)	(0.0012)
b	0.9888	0.9962	0.9942	0.9965	0.9911	0.9942
	(0.0010)	(0.0004)	(0.0004)	(0.0003)	(0.0006)	(0.0004)

Table 5: P-values of likelihood ratio tests on 15 minute data. Specifications $\mathcal{M}_j$, $j = 0, \ldots, 3$ of the $\boldsymbol{\beta}_t$'s are defined in (12)-(15). LR indicates the likelihood ratio test between the models as detailed in the top row.

Ticker	$\mathcal{M}_1$ vs. $\mathcal{M}_0$	$\mathcal{M}_2$ vs. $\mathcal{M}_1$	$\mathcal{M}_3$ vs. $\mathcal{M}_2$
ANF	0.0000	0.0000	0.9166
BAC	0.0000	0.0000	0.0000
C	0.0000	0.0000	0.0005
F	0.0000	0.0000	0.6511
GE	0.0000	0.0000	0.0213
JNJ	0.0000	0.0000	0.0000

Table 6: Average Information Criteria $IC = (2l - penalty)/T$, where l is the maximum log-likelihood, T is the number of days in the data, $penalty$ is $2p$ for AIC and $\ln(T)p$ for BIC, on 15 minute data. Higher values are preferred. Specifications $\mathcal{M}_j$, $j = 0, \ldots, 3$ of the $\boldsymbol{\beta}_t$'s are defined in (12)-(15).

	Average AIC				Average BIC			
Ticker	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$
ANF	144.2968	145.1582	145.1891	145.1698	144.2112	145.0660	145.0903	144.9953
BAC	150.8371	152.1229	152.1800	152.1899	150.7514	152.0306	152.0812	152.0153
C	145.1327	147.6534	147.6789	147.6826	145.0471	147.5612	147.5801	147.5081
F	139.3947	141.0750	141.0946	141.0786	139.3091	140.9828	140.9958	140.9041
GE	152.0116	153.1825	153.2710	153.2666	151.9260	153.0903	153.1722	153.0921
JNJ	154.0000	155.2014	155.3017	155.3278	153.9144	155.1093	155.2030	155.1534

coefficients in the $\boldsymbol{A}$ matrix (Table 3). The values of the first coefficient are about 0.02 while the second to the last are higher and the last are generally the highest around 0.04.

As far as the different specifications are concerned, the likelihood ratio tests reported in Table 5 indicate that the simplest GAS formulation, $\mathcal{M}_1$, provides a better fit than the no GAS effect specification $\mathcal{M}_0$, and, more in general, that larger models perform better than their simpler counterparts (except, in some cases, for $\mathcal{M}_3$ versus $\mathcal{M}_2$).

Somewhat similar remarks can be derived from Table 6 where we report some information criteria (higher values are preferred): AIC favors the richer parameterization $\mathcal{M}_3$ and the intermediate $\mathcal{M}_2$ in an equal number of cases, while BIC supports $\mathcal{M}_2$; at any rate, both criteria show that the GAS based formulations ($\mathcal{M}_1$ to $\mathcal{M}_3$) are clearly superior to $\mathcal{M}_0$.

Although the previous statistics support the GAS based formulations and, among them, the most complex specifications $\mathcal{M}_2$ and $\mathcal{M}_3$, we still need to check if they are able to capture the dynamics of the data adequately. Table 7 shows the p-values of the specification diagnostics discussed in Section 3. The score based tests indicate that the models proposed capture well the predictable dynamics shown by the data in addition to the periodic component, with an overwhelming increase of the p-values from the previous Table 1 (which refers to $\mathcal{M}_0$). With the exception of Citigroup, the $ST(p)$ test statistics point to a reduction in the p-values leading in most cases to a rejection of the null. This is maybe due to the separate consideration of the moment conditions, rather than their sum (which may imply some algebraic compensation). Differently from the previous diagnostics, the score based specification tests seem to support the simplest GAS formulation $\mathcal{M}_1$.

A few comments are also necessary about the empirical performance of the various scaling matrices of the score introduced in Section 3. Following Creal *et al.* (2012) the issue of scaling is crucial for having stable numerical properties of the estimation algorithm.

Table 7: *Ex-post* detection of residual dynamics. P-values of the score based specification tests on 15 minute data. $ST(1)$ and $ST(p)$ indicate the score based specification tests (Equation (11)) derived from the moment Equations (9) and (10), respectively. Specifications $\mathcal{M}_j$, $j = 1, \ldots, 3$ of the $\boldsymbol{\beta}_t$'s are defined in (13)-(15).

	$\mathcal{M}_1$		$\mathcal{M}_2$		$\mathcal{M}_3$	
Ticker	$ST(1)$	$ST(p=28)$	$ST(1)$	$ST(p=30)$	$ST(1)$	$ST(p=53)$
ANF	0.7033	0.3555	0.8831	0.4565	0.5369	0.0558
BAC	0.0413	0.0044	0.0225	0.0204	0.0654	0.0040
C	0.0018	0.0705	0.0006	0.0742	0.0064	0.1994
F	0.1606	0.1333	0.1721	0.0616	0.2528	0.0291
GE	0.0607	0.0158	0.1875	0.1786	0.4871	0.3627
JNJ	0.0391	0.0002	0.1202	0.0068	0.3266	0.0044

In our case, the identity matrix and $\text{ddiag}(\boldsymbol{\mathcal{I}}_t)^{-1/2}$ turn out to be a bad choice, since they produce no convergence for more than half of the estimated models across tickers. Between the other two, $\boldsymbol{\mathcal{I}}_t^{-1}$ leads to better results than $\text{ddiag}(\boldsymbol{\mathcal{I}}_t)^{-1/2}$ in terms of computational burden, value of the log-likelihood function and model diagnostics. All reported results are obtained with this scaling matrix.

4.2 Forecasting

We compare the out–of–sample prediction ability of the different model formulations in a forecasting exercise (August, 1 to May, 31, 2013 – 206 days) performing Diebold and Mariano (1995) tests on 1-day ahead forecasts. We use three different loss functions:

- minus the Dirichlet log-likelihood,

$$L_t^{\mathsf{LL}} = -\left[\ln\Gamma(\widehat{\alpha}_{t,0}) - \sum_{i=1}^{I}\ln\Gamma(\widehat{\alpha}_{t,i}) + \sum_{i=1}^{I}(\widehat{\alpha}_{t,i}-1)\ln w_{t,i}\right]; \tag{16}$$

- the *Slicing* loss function proposed by Brownlees *et al.* (2011),

$$L_t^{\mathsf{SL}} = \sum_{i=1}^{I} w_{t,i}\left(\ln w_{t,i} - \ln\widehat{w}_{t,i}\right); \tag{17}$$

- the Squared Error loss function,

$$L_t^{\mathsf{SE}} = \sum_{i=1}^{I}\left(w_{t,i} - \widehat{w}_{t,i}\right)^2; \tag{18}$$

where $\widehat{w}_{t,i} = \widehat{\alpha}_{t,i}/\widehat{\alpha}_{t,0}$.

Table 8 shows the p-values to test whether the loss of the smaller model is greater than the one of the larger model. The results, similar across loss functions, show that the intermediate model $\mathcal{M}_1$ improves significantly over $\mathcal{M}_0$. The larger $\mathcal{M}_2$ model shows some significant improvement only for two tickers (GE and JNJ). Additionally, allowing for an even richer specification with model $\mathcal{M}_3$ improves only marginally relative to $\mathcal{M}_2$ (GE across loss functions and JNJ just for one). We can thus suggest that the substantial increase in the number of parameters does not always pay off in terms of out–of–sample prediction accuracy: the issue is all the more important in the perspective of analyzing data at smaller intervals (i.e. we may not need the number of parameters to increase 'proportionally' to the number of bins).

Table 8: P-values of the Diebold-Mariano test for 1-day ahead predictions (Aug. 1, 2012 – May 31, 2013) on 15 minute data, considering alternative hypotheses $H_1 : L(\mathcal{M}_{j-1}) > L(\mathcal{M}_j)$, where $L(\cdot)$ denotes the loss computed on the model within parentheses (header). The parameter specifications $\mathcal{M}_j$, $j = 0, \ldots, 3$ of the $\boldsymbol{\beta}_t$ Equation (5) are defined in (12)-(15). The loss functions are defined in (16)-(18).

	$H_1 : L(\mathcal{M}_0) > L(\mathcal{M}_1)$			$H_1 : L(\mathcal{M}_1) > L(\mathcal{M}_2)$			$H_1 : L(\mathcal{M}_2) > L(\mathcal{M}_3)$		
Ticker	L^{LL}	L^{SL}	L^{SE}	L^{LL}	L^{SL}	L^{SE}	L^{LL}	L^{SL}	L^{SE}
ANF	0.0000	0.0000	0.0000	0.9510	0.9400	0.9652	0.9753	0.9402	0.9874
BAC	0.0000	0.0000	0.0000	0.3976	0.2698	0.4660	0.4259	0.4290	0.2280
C	0.0000	0.0000	0.0000	0.1744	0.1970	0.2379	0.3125	0.2958	0.2328
F	0.0000	0.0000	0.0000	0.3519	0.0685	0.1223	0.7853	0.5997	0.3960
GE	0.0000	0.0000	0.0000	0.0021	0.0021	0.0061	0.0247	0.0273	0.0004
JNJ	0.0000	0.0000	0.0000	0.0395	0.0130	0.0094	0.7182	0.1590	0.0065

5 Is VWAP a Suitable Benchmark for Share Models?

A more performance–oriented evaluation of the share forecasts would include them in *predictions* of the VWAP to be compared against the corresponding historical values. Of course one should consider that future prices by bin are unpredictable, to the point that if one were to follow a *genuine* one-day ahead approach, using $p^{(L)}_{t-1,I}$ (the last recorded price of day $t-1$) as forecast of all prices throughout the t-th day, the resulting 1–day VWAP forecasts would be the previous day's closing price irrespective of share forecasts.

Instead of predictions, therefore, we must resort to *virtual* VWAP values, computed by replacing each price forecast with a corresponding reference price *observed ex-post*, such as, for example, the VWAP of the previous bin (cf. Białkowski *et al.* (2008)) or of the current one; a price randomly sampled within the bin; the last price of the previous or of the same bin (cf. Brownlees *et al.* (2011)). Considering the last option, the virtual VWAP is given by

$$\widetilde{\mathsf{VWAP}}_t = \sum_{i=1}^{I} \widehat{w}_{t,i} p^{(L)}_{t,i}, \tag{19}$$

where $\widehat{w}_{t,i}$ is the one-day ahead forecast of the corresponding volume share and $p^{(L)}_{t,i}$ is the last recorded price of the bin. The latter provides a *neutral* reference price which is simple to get and is a sensible choice from an operational point of view, because it does not involve the knowledge of the actual shares within the bin as bin VWAPs would require. Notice, though, that an actual strategy with the allocation of certain shares to certain bins is not guaranteed to deliver the prices observed ex-post, since the allocation of a substantial share in a specific bin could have a strong market impact and then change subsequent prices and even the end–of–day result. This fact has some relevance, as discussed below.

Let us analyze the relevance of the share forecasts when evaluating measures of discrepancy between the VWAP and the virtual VWAP (Equation (19)), by considering the out–of–sample MSE,

$$MSE = \tau^{-1} \sum_{t=T+1}^{T+\tau} \left(\mathsf{VWAP}_t - \widetilde{\mathsf{VWAP}}_t \right)^2 = \tau^{-1} \sum_{t=T+1}^{T+\tau} \left[\sum_{i=1}^{I} \left(w_{t,i} \overline{p}_{t,i} - \widehat{w}_{t,i} p^{(L)}_{t,i} \right) \right]^2 . \tag{20}$$

Notice that its value depends on four factors:

1. The differences between $\widehat{w}_{t,i}$ and $w_{t,i}$ (share forecasting errors) already discussed;
2. The differences between $p^{(L)}_{t,i}$ and $\overline{p}_{t,i}$ (related to the intra-bin volatility);

3. The variability of the $\overline{p}_{t,i}$'s during the day (intra-day volatility);
4. Intra-day algebraic compensations.

Such factors cannot be easily disentangled because they interact in a non-additive way, but some ideas about their relevance can be given.

Points 1. and 2. above are interrelated: let us consider the following decomposition of the MSE,

$$MSE = MSE_1 + MSE_2 + DP \tag{21}$$

where

$$MSE_1 = \tau^{-1} \sum_{t=T+1}^{T+\tau} \left[\sum_{i=1}^{I} (w_{t,i} - \widehat{w}_{t,i})\, \overline{p}_{t,i} \right]^2 \tag{22}$$

$$MSE_2 = \tau^{-1} \sum_{t=T+1}^{T+\tau} \left[\sum_{i=1}^{I} \widehat{w}_{t,i} \left(\overline{p}_{t,i} - p_{t,i}^{(L)} \right) \right]^2 \tag{23}$$

$$DP = 2\tau^{-1} \sum_{t=T+1}^{T+\tau} \left[\sum_{i=1}^{I} (w_{t,i} - \widehat{w}_{t,i})\, \overline{p}_{t,i} \sum_{i=1}^{I} \widehat{w}_{t,i} \left(\overline{p}_{t,i} - p_{t,i}^{(L)} \right) \right]. \tag{24}$$

The MSE is here expressed as the sum of three addends: MSE_1, equal to MSE but with the bin VWAP prices $\overline{p}_{t,i}$ in place of the closing bin prices $p_{t,i}^{(L)}$, summarizes the effect of the forecasting errors; MSE_2, equal to MSE but with the shares predictions $\widehat{w}_{t,i}$ in place of the corresponding $w_{t,i}$ values, captures the effect of the price differences; to establish equality with the MSE (20), we need to consider a double product term, DP.[4] The first four panels of Table 9 report the values of the MSE and of its components for the tickers analyzed in the paper. In general, the contribution of the share's forecasting errors (MSE_1) dominates the other two components; moreover, the effect of the price differences (MSE_2) and of the double product term (DP) tend to cancel each other out. Only JNJ tells a partially different story, in the sense that the price differences have a non-negligible impact and that their contribution does not tend to cancel out the double product component. We can then conclude that the share's forecasting errors play a dominant role, in comparison with the price differences, in explaining the MSE values; noting, in addition, that exact share forecasts give a zero value of MSE_1, the interest in producing good volume share predictions seems fully justified.

Moving to the third factor in the list, the relevance of the daily volatility is easily explained observing that when $\overline{p}_{t,i}$ is a constant, say $\overline{p}_{t,i} = \overline{p}_t$, then

$$MSE_1 = \tau^{-1} \sum_{t=T+1}^{T+\tau} \overline{p}_t^2 \left[\sum_{i=1}^{I} (w_{t,i} - \widehat{w}_{t,i}) \right]^2 = 0$$

irrespective of the quality of the share forecasts while, other factors being equal, it tends to increase with the variability of the $\overline{p}_{t,i}$'s. This implies that, in the case of low volatility, accurate share forecasting may be irrelevant while, when volatility is high, it may gain importance.

Table 10 reports the MSE and the MSE_1 values computed on the top 20% highest and the bottom 20% lowest plain vanilla 15 minute realized volatility days. As expected, when volatility is high, the error measures are considerably larger than in days with low volatility; likewise, the differences in the MSE's between models tend to increase when the volatility goes up. This effect is acknowledged by practitioners (cf. Hobson (2006))

[4]The identity is easily proved by adding and subtracting $\widehat{w}_{t,i}\overline{p}_{t,i}$ inside the parentheses of (20). An alternative decomposition can be obtained adding and subtracting $w_{t,i}p_{t,i}^{(L)}$.

Table 9: Values (×100) of different versions of out–of–sample MSE's (see Equations (20), (22), (23), (24), (26) and (27) for their definition) on 15 minutes data. $\mathcal{U}$ indicates an uniform allocation of the stocks across bins; $\mathcal{M}_j$, $j = 0, \ldots, 3$ denote allocation strategies computed according to the model described in Section 2 and whose $\boldsymbol{\beta}_t$ equations are defined in (12)-(15).

Error Measure	Strategy	ANF	BAC	C	F	GE	JNJ
	$\mathcal{U}$	1.1919	0.0258	0.2685	0.0290	0.0497	0.2199
	$\mathcal{M}_0$	0.8468	0.0130	0.1464	0.0173	0.0226	0.1013
MSE	$\mathcal{M}_1$	0.8009	0.0150	0.1617	0.0167	0.0215	0.0966
	$\mathcal{M}_2$	0.8195	0.0152	0.1618	0.0165	0.0211	0.0949
	$\mathcal{M}_3$	0.8213	0.0154	0.1621	0.0165	0.0213	0.0938
	$\mathcal{U}$	1.2459	0.0235	0.2585	0.0267	0.0517	0.2126
	$\mathcal{M}_0$	0.8654	0.0111	0.1402	0.0159	0.0216	0.0784
MSE_1	$\mathcal{M}_1$	0.8013	0.0146	0.1580	0.0147	0.0190	0.0662
	$\mathcal{M}_2$	0.8205	0.0148	0.1582	0.0144	0.0185	0.0637
	$\mathcal{M}_3$	0.8231	0.0150	0.1583	0.0144	0.0186	0.0631
	$\mathcal{U}$	0.0276	0.0011	0.0099	0.0012	0.0015	0.0095
	$\mathcal{M}_0$	0.0646	0.0024	0.0264	0.0023	0.0043	0.0295
MSE_2	$\mathcal{M}_1$	0.0715	0.0029	0.0299	0.0027	0.0055	0.0381
	$\mathcal{M}_2$	0.0703	0.0029	0.0301	0.0027	0.0056	0.0390
	$\mathcal{M}_3$	0.0703	0.0029	0.0299	0.0027	0.0056	0.0394
	$\mathcal{U}$	-0.0815	0.0012	0.0001	0.0011	-0.0035	-0.0022
	$\mathcal{M}_0$	-0.0833	-0.0006	-0.0203	-0.0009	-0.0033	-0.0066
Dp	$\mathcal{M}_1$	-0.0719	-0.0025	-0.0262	-0.0007	-0.0030	-0.0077
	$\mathcal{M}_2$	-0.0714	-0.0025	-0.0265	-0.0007	-0.0029	-0.0079
	$\mathcal{M}_3$	-0.0720	-0.0024	-0.0261	-0.0007	-0.0029	-0.0087
	$\mathcal{U}$	209.5426	10.3935	148.5851	12.6187	64.9818	877.4284
	$\mathcal{M}_0$	95.6124	3.5566	48.8251	5.5259	20.1175	227.3815
MSE^*	$\mathcal{M}_1$	85.4159	3.1968	45.5394	4.8106	16.0862	173.5318
	$\mathcal{M}_2$	85.9518	3.1945	45.4130	4.7976	15.8019	170.6608
	$\mathcal{M}_3$	86.0525	3.1931	45.4219	4.7979	15.7616	170.0171
	$\mathcal{U}$	209.5237	10.3946	148.5802	12.6204	64.9745	877.4463
	$\mathcal{M}_0$	95.6001	3.5569	48.8141	5.5272	20.1082	227.3591
MSE_1^*	$\mathcal{M}_1$	85.4297	3.1966	45.5325	4.8117	16.0800	173.5057
	$\mathcal{M}_2$	85.9587	3.1942	45.4068	4.7987	15.7963	170.6330
	$\mathcal{M}_3$	86.0591	3.1929	45.4155	4.7991	15.7561	169.9893

Table 10: Values (×100) of two versions of the out–of–sample MSE (see Equations (20) and (22) for their definition) computed on 15 minutes data, considering separately the 20% of days with the highest/smallest (High/Low) realized volatility. $\mathcal{U}$ indicates an uniform allocation of the stocks across bins; $\mathcal{M}_j$, $j = 0, \ldots, 3$ denote allocation strategies computed according to the model described in Section 2 and whose $\boldsymbol{\beta}_t$ equations are defined in (12)-(15).

Volatility	Error Measure	Strategy	ANF	BAC	C	F	GE	JNJ
High	MSE	$\mathcal{U}$	4.0303	0.0668	0.6221	0.0849	0.1477	0.5706
		$\mathcal{M}_0$	2.8127	0.0385	0.4014	0.0453	0.0724	0.2335
		$\mathcal{M}_1$	2.6809	0.0459	0.4524	0.0435	0.0681	0.2076
		$\mathcal{M}_2$	2.7262	0.0470	0.4571	0.0428	0.0665	0.2012
		$\mathcal{M}_3$	2.7293	0.0483	0.4630	0.0427	0.0670	0.1996
	MSE_1	$\mathcal{U}$	4.2816	0.0626	0.6414	0.0797	0.1586	0.5933
		$\mathcal{M}_0$	2.8896	0.0372	0.4098	0.0403	0.0791	0.2279
		$\mathcal{M}_1$	2.7277	0.0480	0.4547	0.0379	0.0718	0.1790
		$\mathcal{M}_2$	2.7650	0.0491	0.4598	0.0372	0.0699	0.1711
		$\mathcal{M}_3$	2.7702	0.0503	0.4665	0.0371	0.0704	0.1706
Low	MSE	$\mathcal{U}$	0.2144	0.0071	0.1117	0.0076	0.0094	0.1108
		$\mathcal{M}_0$	0.1097	0.0021	0.0349	0.0044	0.0044	0.0743
		$\mathcal{M}_1$	0.0844	0.0017	0.0323	0.0043	0.0047	0.0734
		$\mathcal{M}_2$	0.0871	0.0018	0.0325	0.0042	0.0047	0.0738
		$\mathcal{M}_3$	0.0879	0.0018	0.0329	0.0042	0.0047	0.0734
	MSE_1	$\mathcal{U}$	0.2363	0.0067	0.0925	0.0068	0.0079	0.1009
		$\mathcal{M}_0$	0.1118	0.0012	0.0128	0.0041	0.0014	0.0536
		$\mathcal{M}_1$	0.0834	0.0011	0.0116	0.0037	0.0012	0.0512
		$\mathcal{M}_2$	0.0865	0.0011	0.0116	0.0036	0.0011	0.0521
		$\mathcal{M}_3$	0.0873	0.0011	0.0117	0.0036	0.0011	0.0513

who frequently scale the daily VWAP error by the daily price range. In our framework, the MSE can be turned into its volatility-adjusted counterpart as

$$MSE^{(s)} = \tau^{-1} \sum_{t=T+1}^{T+\tau} \left[\sum_{i=1}^{I} \left(\frac{w_{t,i}\overline{p}_{t,i} - \widehat{w}_{t,i} p_{t,i}^{(L)}}{p_t^{\max} - p_t^{\min}} \right) \right]^2 , \qquad (25)$$

where $p_t^{\max}$ ($p_t^{\min}$) is the maximum (minimum) price in the t-th day. Table 11 reports values for the out–of–sample volatility–adjusted MSE, evidencing nothing substantially different from Table 9, except that the model ranking seems now to favor $\mathcal{M}_0$ against $\mathcal{M}_1$ (better in 4 out of 6 cases), reversing in this sense the evidence of the latter table. We note incidentally that, in comparison with the original measures, scaling the daily VWAP differences by the corresponding price ranges tends to reduce differences in the MSE values among tickers.

Considering finally the last factor in the list, what we named intra-daily compensations is a consequence of the fact that 'the square of a sum' is not 'the sum of squares'. In fact, the differences appearing into MSE and MSE_1 are computed on daily aggregated quantities, so that possibly high bin differences (in absolute value) but with a different sign, may cancel each other out, once they are aggregated daily. Such an effect on shares is somewhat to be expected: for example, considering MSE_1 for simplicity, it is likely that overestimating the true volume share in, say, the first bin will tend to produce underestimation in some of the remaining bins of the day, because of the zero lower bound and the adding up constraint.[5] In order to appreciate the numerical relevance of

[5] In general, any combination of weights $\widehat{\boldsymbol{w}}_t$ that, at the same time lies in the I-dimensional simplex and

Table 11: Values (×100) of the out–of–sample volatility–adjusted MSE's (Equations (25 and (28)) on 15 minutes data. $\mathcal{U}$ indicates an uniform allocation of the stocks across bins; $\mathcal{M}_j$, $j = 0, \ldots, 3$ denote allocation strategies computed according to the model described in Section 2 and whose $\boldsymbol{\beta}_t$ equations are defined in (12)-(15).

Error Measure	Strategy	ANF	BAC	C	F	GE	JNJ
$MSE^{(s)}$	$\mathcal{U}$	0.3242	0.2658	0.2742	0.2669	0.3226	0.3628
	$\mathcal{M}_0$	0.2136	0.1208	0.1341	0.1604	0.1337	0.1706
	$\mathcal{M}_1$	0.1994	0.1266	0.1405	0.1614	0.1358	0.1695
	$\mathcal{M}_2$	0.2033	0.1271	0.1399	0.1602	0.1350	0.1692
	$\mathcal{M}_3$	0.2040	0.1277	0.1392	0.1602	0.1360	0.1674
$MSE^{(s)*}$	$\mathcal{U}$	5514.8	6755.4	7450.9	8302.3	28456.8	70966.2
	$\mathcal{M}_0$	2491.3	2152.8	2380.1	3477.4	8387.8	17730.4
	$\mathcal{M}_1$	2180.3	1872.9	2147.5	2891.3	6383.5	13366.7
	$\mathcal{M}_2$	2201.9	1864.4	2143.1	2869.5	6205.9	12936.7
	$\mathcal{M}_3$	2206.2	1861.1	2146.0	2865.2	6179.7	12882.4

the compensations, we compare MSE, MSE_1 and $MSE^{(v)}$ with the corresponding 'sum of squares' versions[6]

$$MSE^* = \tau^{-1} \sum_{t=T+1}^{T+\tau} \sum_{i=1}^{I} \left(w_{t,i}\overline{p}_{t,i} - \widehat{w}_{t,i} p_{t,i}^{(L)} \right)^2 \tag{26}$$

$$MSE_1^* = \tau^{-1} \sum_{t=T+1}^{T+\tau} \sum_{i=1}^{I} \left[\left(w_{t,i} - \widehat{w}_{t,i} \right) \overline{p}_{t,i} \right]^2 \tag{27}$$

$$MSE^{(v)*} = \tau^{-1} \sum_{t=T+1}^{T+\tau} \sum_{i=1}^{I} \left(\frac{w_{t,i}\overline{p}_{t,i} - \widehat{w}_{t,i} p_{t,i}^{(L)}}{p_t^{\max} - p_t^{\min}} \right)^2 . \tag{28}$$

Such a comparison (Table 9) shows that the impact of the intra-daily compensations is so overwhelming as to obscure the remaining factors. For example, the MSE^* and MSE_1^* measures reveal that the $\mathcal{M}_1$-based strategy gives a substantial reduction in the error, when compared to $\mathcal{M}_0$; on the contrary, MSE and MSE_1 tend to hide such a result, sometimes altering the model ranking provided by the starred counterparts. Similar considerations come from Table 11 regarding the volatility–adjusted MSE's.

Summarizing, the analysis of the VWAP MSE reveals that its very structure tends to obscure (and even to cancel out) the gains provided by better share predictions: depending on *unpredictable* combined features of the share and price dynamics, models with a worse performance in terms of share predictions may produce virtual VWAP values indistinguishable from better performing share models. More importantly, poor share forecasts (corresponding to a high MSE^*) cannot really be associated with historically observed bin prices, given that getting the 'wrong' timing of order execution is likely to have a relevant market impact. In other words, using historically observed prices with the 'wrong'

satisfies $\sum_{i=1}^{I} \widehat{w}_{t,i}\overline{p}_{t,i} = \sum_{i=1}^{I} w_{t,i}\overline{p}_{t,i}$ guarantees that the contribution of the t-th day to MSE_1 is zero.

[6]Since, in general,

$$\left(\sum_i a_i \right)^2 = \sum_i a_i^2 + \sum_i \sum_{j \neq i} a_i a_j,$$

compensations in the LHS between positive and negative differences (represented by the a_i's terms in above formula) transform themselves into negative $a_i a_j$ terms in the RHS because of the different signs. As a consequence, the difference between the 'square of sum' and the 'sum of squares' versions of the MSE measures the incidence of the compensations.

Table 12: P-values of the Diebold-Mariano test for 1-day ahead 15 minute virtual VWAP (Aug. 1, 2012 – May 31, 2013), considering alternative hypotheses $H_1 : L(\mathcal{M}_{j-1}) > L(\mathcal{M}_j)$, where $L(\cdot)$ denotes the loss computed on the model within parentheses (header of the table). The parameter specifications $\mathcal{M}_j$, $j = 0, \ldots, 3$ of the $\boldsymbol{\beta}_t$ Equation (5) are defined in (12)-(15). The loss functions L^{SE}, L^{SE^*} are defined in (29)-(30).

	$H_1 : L(\mathcal{U}) > L(\mathcal{M}_0)$		$H_1 : L(\mathcal{M}_0) > L(\mathcal{M}_1)$		$H_1 : L(\mathcal{M}_1) > L(\mathcal{M}_2)$		$H_1 : L(\mathcal{M}_2) > L(\mathcal{M}_3)$	
Ticker	L^{SE}	L^{SE^*}	L^{SE}	L^{SE^*}	L^{SE}	L^{SE^*}	L^{SE}	L^{SE^*}
ANF	0.0288	0.0000	0.0803	0.0000	0.9763	0.9256	0.9806	0.9780
BAC	0.0000	0.0000	0.8972	0.0000	0.8046	0.4098	0.7716	0.3634
C	0.0000	0.0000	0.9765	0.0000	0.5158	0.1474	0.5760	0.5530
F	0.0002	0.0000	0.0785	0.0000	0.0021	0.1976	0.4829	0.5546
GE	0.0000	0.0000	0.1149	0.0000	0.0645	0.0061	0.8588	0.0006
JNJ	0.0000	0.0000	0.2258	0.0000	0.0990	0.0107	0.0001	0.0050

volume shares is a counterfactual which would never be of practical importance.[7] In this respect, therefore, we advocate starred error measures as being closer in spirit to the need to reduce the market impact of trades, given that a small 'sum of squares' implies a low 'square of sum' while giving, at the same time, positive value to reduced share forecasting errors.

The comparison between the two sets of measures is performed according to the Diebold-Mariano test applied to two loss functions

$$L_t^{SE} = \left[\sum_{i=1}^{I} \left(w_{t,i}\overline{p}_{t,i} - \widehat{w}_{t,i}p_{t,i}^{(L)}\right)\right]^2 \quad (29)$$

$$L_t^{SE^*} = \sum_{i=1}^{I} \left(w_{t,i}\overline{p}_{t,i} - \widehat{w}_{t,i}p_{t,i}^{(L)}\right)^2. \quad (30)$$

related to MSE and MSE^*, respectively, as in Table 12. The results show a clear dominance of models over the uniform share strategy for both measures, while a discrepancy is observed for the relevance of the GAS effects. In fact, the second set of results (inclusion of GAS effects) shows a better performance of $\mathcal{M}_1$ according to the starred criterion, while such a superiority would not be detected should one use L^{SE}. As it happened with share forecasting diagnostics (cf. Table 8), the evidence about which parameterization performs better is ticker–dependent, showing a preference for richer models for GE and JNJ (both $\mathcal{M}_2$ over $\mathcal{M}_1$, and $\mathcal{M}_3$ over $\mathcal{M}_2$) when adopting the starred criterion.

6 Conclusions

In this paper we outline a modeling strategy for intra–daily volume shares expressed as a proportion of the total daily volume. The issue is relevant for a trader who wishes to avoid the price impact of a trade as much as possible by dispersing an overall order during the day "going with the flow", i.e. in smaller chunks of different sizes, placing higher portion of his orders where the activity is higher.

The data can be constructed dividing the trading day in bins as in, say, Brownlees *et al.* (2011) and aggregating volume data. The raw data shares present a strong intra–daily periodic pattern but also some significant additional predictable dynamics, worth of an appropriate modeling treatment. In so doing, we assume that the volume shares of

[7] From a practitioner's point of view, who actually implements a VWAP–based trading strategy, it is actually possible to verify at the end of the day the sequence of volumes and prices paid (hence a realized VWAP) as the result of her strategy (including slippage effects – difference between expected and actual prices) and the official VWAP.

the different bins within the same day follow a Dirichlet distribution whose parameters are made time-varying taking advantage of the flexibility of a novel class of models called Generalized Autoregressive Score (Creal *et al.* (2012)). We also propose some score–based specification tests in order to check the model adequacy in reproducing the pattern of the data. Model diagnostics show that the proposed model performs very well in capturing the dynamics of the data. The Diebold-Mariano out–of–sample tests indicate that the model including GAS effects (namely with a non–periodic time–varying dynamics of the Dirichlet parameters outperforms significantly the model with no GAS effects. Some parameter restrictions are possible and accommodate a seemingly different dynamics around the end of the trading day without a substantial loss in the forecasting performance. This specification parsimony looks promising since it allows for a reduction in the time interval (for example from 15 to 5 minutes) without resorting to a cumbersome increase in the number of parameters.

We have discussed extensively how traditional measures of loss functions for forecasting performance are not appropriate when VWAP predictions are to be evaluated: MSE is made of several components linked to different factors which may variously combine with one another, leading to the wrong impression that a more sophisticated share prediction may not deliver a better performance in terms of VWAP values. The issue is solved by showing that the compensating elements may cloud the picture, essentially because in considering actual prices, one is not capable of measuring the counterfactual market impact of trades, i.e. deriving what prices would be observed as the result of a 'wrong' allocation of an order. We suggest therefore an alternative measure which avoids compensations and we show that the GAS model successfully captures relevant dynamics in intra-daily volume shares even when prices are considered.

References

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**, 97–112.

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

Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994). *ARCH Models, Handbook of Econometrics*, volume IV, chapter 49, pages 2959–3038. Elsevier Science B.V.

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., Cipollini, F., and Gallo, G. M. (2011). Intra-daily volume modeling and prediction for algorithmic trading. *Journal of Financial Econometrics*, **9**, 489–518.

Calvori, F., Creal, D., Koopman, S. J., and André, L. (2013). Testing for parameter instability in competing modeling frameworks. Technical report, Timbergen Institute Discussion Paper, Roetersstraat 31, 1018 WB Amsterdam, The Nederlands.

Creal, D., Koopman, S. J., and Lucas, A. (2012). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, (available online).

Darolles, S. and Le Fol, G. (2003). Trading volume and arbitrage. Technical report, CREST.

Davidson, R. and MacKinnon, J. G. (1990). Specification tests based on artificial regressions. *Journal of the American Statistical Association*, **85**, 220–227.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, **13**, 253–263.

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

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

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Society Monographs. Cambridge University Press, Cambridge.

Hautsch, N. and Ruihong, H. (2012). The market impact of a limit order. *Journal of Economic Dynamics and Control*, **36**, 501–522.

Hobson, D. D. (2006). VWAP and volume profiles. *Journal of Trading*, **1**, 38–42.

Kissel, R. and Glantz, M. (2003). *Optimal Trading Strategies*. AMACOM, New York (NY).

Konishi, H. (2002). Optimal slice of a VWAP trade. *Journal of Financial Markets*, **5**, 197–221.

Madhavan, A. (2002). *VWAP Strategies*, pages 32–38. Investment Guides Series. Institutional Investor Inc.

McCulloch, J. and Kazakov, V. (2007). Optimal VWAP trading strategy and relative volume. Technical report, University of Technology, Sydney. Research Paper 201.

Newey, W. K. (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica*, **53**, 1047–1070.

White, H. (1987). Specification testing in dynamic models. In T. F. Bewley, editor, *Advances in Econometrics, Fifth World Congress*, volume 1, chapter 9, pages 1–58. Cambridge University Press.