High-Frequency Volatility Modelling: A Markov-Switching Autoregressive Conditional Intensity Model

Yifan Li* Ingmar Nolte† Sandra Nolte‡

University of Manchester Lancaster University Lancaster University

Abstract

We develop a Markov-Switching Autoregressive Conditional Intensity (MS-ACI) model with time-varying transitional probability, and show that it can be reliably estimated via the Stochastic Approximation Expectation-Maximization algorithm. Applying our model to high-frequency transaction data, we detect two distinct regimes in the intraday volatility process: a dominant volatility regime that is observable throughout the trading day representing the risk-transferring trading activity of investors, and a minor volatility regime that concentrates around market liquidity shocks which mainly capture impacts of firm-specific news arrivals. We propose a novel daily volatility decomposition based on the two detected volatility regimes.

Keywords: Regime Switch, Intensity Modelling, Invariance, Stock Return Volatility.

*Corresponding author: Alliance Manchester Business School, Booth Street West, Manchester, M15 6PB, UK. e-mail: yifan.li@manchester.ac.uk.

†Lancaster University Management School, Bailrigg, Lancaster, LA1 4YX, UK. Phone +44 152 459 2644, e-mail: I.Nolte@lancaster.ac.uk

‡Lancaster University Management School, Bailrigg, Lancaster, LA1 4YX, UK. Phone +44 152 459 3634, e-mail: S.Nolte@lancaster.ac.uk.

1 Introduction

Since the seminal work of Engle and Russell (1998), a growing amount of literature has emerged on parametric modelling of intraday financial data. An important strand of this literature concentrates on the parametric modelling of intraday price volatility using point processes, including Gerhard and Hautsch (2002), Tse and Yang (2012), Li et al. (2019), Hong et al. (2020), etc. These papers show that point process based volatility estimators can provide valid intraday volatility estimates that are comparable to the popular Realized Variance (RV) estimates of Andersen et al. (2001).

An interesting feature of the point process-based approach is that its parametric structure provides a useful framework to examine intraday volatility interaction with market microstructure covariates without compromising the quality of volatility estimates. Existing frameworks are either incapable of incorporating other variables due to their non-parametric design (e.g. RV framework), or are considered inappropriate for intraday volatility estimation (e.g. intraday GARCH framework).

Under the point process-based framework, we propose a Markov-switching extension to the Autoregressive Conditional Intensity (ACI) model of Russell (1999) for the modelling of intraday volatility. To the best of our knowledge, we are among the first to develop such an extension to a conditional intensity model. In fact, only few studies consider Markov-switching extensions to autoregressive models for point processes, e.g. Hujer et al. (2002); De Luca and Zuccolotto (2006); Gallo and Otranto (2012), which are all developed from the Autoregressive Conditional Duration (ACD) model of Engle and Russell (1998). We therefore fill this gap by providing an intensity-based autoregressive model with a Markov-switching feature for the modelling of point processes.

The lack of Markov-switching autoregressive models for point processes is possibly due to the fact that a non-linear autoregressive structure is required to ensure the positivity of the durations and the conditional intensity. This introduces a 'path dependency problem' in the construction of the likelihood function[1], which greatly complicates the estimation of such model. The most widely applied strategy to circumvent the path dependency problem is to approximate the observed likelihood function by a feasible version, for example Gray (1996); Kim (1994); Dueker (1997); Klaassen (2002); Haas et al. (2004) in a MS-GARCH framework, and Hujer et al. (2002); De Luca and Zuccolotto (2006); Gallo and Otranto (2012) for the MS-ACD model. However, these approaches do not solve the path dependency problem directly, and the quality of these approximations is difficult to verify empirically (Billio et al., 2014).

Distinct from the aforementioned studies, a direct solution to the path dependency problem typically relies on simulation and data augmentation techniques. For example, Bauwens et al. (2010, 2014);

[1]Intuitively, as the Markov state variables are unobserved, the observed likelihood is computed by integrating out the full path of the latent states. Since possible realizations of the Markov chain grows exponentially w.r.t. the number of observations, direct computation of the observed likelihood quickly become intractable as the sample size expands.

Billio et al. (2014) develop Bayesian estimation techniques for the MS-GARCH model. Augustyniak (2014) introduces a Monte Carlo Expectation-Maximization (MCEM) algorithm (Wei and Tanner, 1990) for maximum likelihood estimation of the MS-GARCH model. Inspired by these approaches, we develop a maximum likelihood estimator of the Markov-switching ACI (MS-ACI) model based on the Stochastic Approximation Expectation-Maximization (SAEM) algorithm of Celeux and Diebolt (1992); Delyon et al. (1999), which overcomes the path dependency problem and provides estimates of the variance-covariance matrix of the estimated parameter and the most probable state vector. Our approach can be computationally more efficient than the method of Augustyniak (2014) as the SAEM algorithm utilizes the simulated data more efficiently than the MCEM algorithm (Delyon et al., 1999). Via simulation, we show that our SAEM algorithm provides empirically feasible and reliable parameter estimates of the MS-ACI model.

We apply the MS-ACI model on the high-frequency Trade and Quote (TAQ) data of ten frequently traded securities (including a market index ETF, SPY) for the year 2016, timestamped at milliseconds. We model the dynamics of price durations, that is, the amount of time for the log-price to change by a given amount, and examine the contemporaneous relationship between price durations and the cumulative trading volume within each price duration. Our MS-ACI model detects two distinct regimes in the intraday volume-duration relationship for individual stocks: a dominant regime that is observable throughout the trading day in which the duration and volume exhibits a strong power law relationship, and a minor regime that concentrates around large bid-ask spread events with a much weaker connection between price duration and volume. However, the latter regime cannot be observed for SPY, the stock index ETF. These findings are further corroborated by empirical results based on an extended sample that covers all Dow Jones constituents in 2016, which we present in the supplementary material of the paper.

Our empirical findings provide new insights into the dynamics of intraday volatility processes. Firstly, we show that intraday volatility processes for individual stocks exhibit regime-switching behaviour on an intraday level, which is likely to be caused by unpredicted shocks to the liquidity state of the market as a result of firm-specific information arrivals. Secondly, we conjecture that the dominant regime discussed above is likely to summarize risk transfers between market participants, and the power law relationship between volume and price duration may hold across assets and time as a result of the Market Microstructure Invariance hypothesis of Kyle and Obizhaeva (2016). Finally, we propose a novel decomposition of daily volatility into two components based on the two detected regimes, which allows us to disentangle different driving forces of volatility on a daily basis.

The rest of this paper is structured as follows: Section 2 introduces some basic point process theory and the original ACI model. In Section 3, the specification and estimation technique of the MS-ACI model are discussed. Simulation evidence is provided in Section 4, with the empirical application following in Section 5. Section 6 concludes.

2 Conditional Intensity Modelling

2.1 Basic Point Process Theory

This section briefly summarizes fundamental point process theory used in this paper following Hautsch (2012). For an in-depth textbook treatment we refer to Karr (1991) and Daley and Vere-Jones (2003) among others.

Definition 1. *On a filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$*, let* t *denote the physical time. A (simple) point process is defined as the sequence of random event arrival times* $\{t_i\}_{i=0,1,2,\ldots}$*, subject to* $t_0 = 0$ *and* $t_i < t_{i+1}, \forall i$*, almost surely. Each* t_i *represents the arrival time of the* i*-th event. For a sample size* T*, the complete observed sequence of the point process can be denoted as* $\{t_i\}_{i=1:T}$*.*

A point process can be uniquely characterized by three processes. The first one is the counting process, denoted as $N(t) := \sum_{i\geq 1} \mathbb{1}_{\{t_i \leq t\}}$ for the right-continuous version and $\breve{N}(t) := \sum_{i\geq 1} \mathbb{1}_{\{t_i < t\}}$ for the left-continuous one. The second process is the duration process defined as $x_i = t_i - t_{i-1}$ for $i > 1$ and $x_1 = t_1$. A related definition is the backward recurrence time, defined as $x(t) = t - t_{\breve{N}(t)}$. Let $\mathcal{F}_t$ denote the information set available till time t to which the point process is adapted, the third process is the $\mathcal{F}_t$-conditional intensity process, defined formally as:

Definition 2. *(Hautsch, 2012, p. 71) Let* $N(t)$ *be a simple point process on* $[0, \infty)$ *that is adapted to some history* $\mathcal{F}_t$ *and assume that* $\lambda(t|\mathcal{F}_t)$ *is a positive-valued process with sample paths that are left-continuous and have right-hand limits. Then the process*

$$\lambda(t|\mathcal{F}_t) \approx \lambda(t + |\mathcal{F}_t) = \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \mathrm{E}[N(t+\Delta) - N(t)|\mathcal{F}_t], \quad \lambda(t+) > 0, \forall t, \tag{2.1}$$

with $\lambda(t + |\mathcal{F}_t) := \lim_{\Delta \downarrow 0} \lambda(t + \Delta|\mathcal{F}_t)$*, is called the* $(\mathcal{F}_t)$*-conditional intensity process of the counting process* $N(t)$*.*

The following crucial property of the conditional intensity process is exploited in constructing models for conditional intensity. Let us denote the integrated intensity between two events as:

$$\Lambda_i = \int_{t_{i-1}}^{t_i} \lambda(s|\mathcal{F}_s) ds.$$

According to the Random Time Change theorem (RTCT hereafter, see e.g. Bowsher (2007)), the process $\{\Lambda_i\}_{i=1,2,\ldots}$ is the duration process of a unit rate Poisson process with:

$$\Lambda_i \sim i.i.d. exp(1). \tag{2.2}$$

This property serves as a useful tool in constructing intensity-based models, and is used in constructing residuals and diagnostic tests for point processes.

The log-likelihood function of a point process can be constructed solely based on the conditional intensity process (Karr, 1991):

$$\ln\ \mathcal{L}(\theta;\mathbb{Y}) = \sum_{i=1}^{T}\big[-\Lambda_i + \ln\ \lambda(t_i|\mathcal{F}_{t_i})\big]. \tag{2.3}$$

Here $\mathbb{Y} = \{t_i\}_{i=1:T}$ and θ is the unique parameter vector for some parametrized conditional intensity process.

2.2 The ACI Model

The seminal paper of Russell (1999) proposes the bivariate ACI model with applications to high-frequency financial data. In this section, we describe the main ingredients of the original ACI model in a univariate framework as in Hautsch (2012).

The ACI model is a fully parametric model, which specifies the conditional intensity in a multiplicative form as follows:

$$\lambda(t|\mathcal{F}_t) = \Phi(t)\lambda_0(t). \tag{2.4}$$

The $\Phi(t)$ is an autoregressive component that could include time-varying covariates, and $\lambda_0(t)$ is the baseline intensity function. Additional components can be included multiplicatively to capture other effects of interest on the conditional intensity (e.g. a seasonality component), which we ignore in this paper for simplicity. To ensure the non-negativity of the conditional intensity, $\Phi(t)$ is usually parametrized as the exponential form of an ARMA-type structure, as an example:

$$\Phi(t) = e^{\tilde{\Phi}_{\breve{N}(t)+1}+\boldsymbol{\eta}'\boldsymbol{Z}(t)}, \tag{2.5}$$

$$\tilde{\Phi}_i = \sum_{j=1}^{q}\alpha_j\tilde{\varepsilon}_{i-j} + \sum_{k=1}^{p}\beta_k\tilde{\Phi}_{i-k}, \tag{2.6}$$

in which $\breve{N}(t)+1 = i$ for $t_{i-1} < t \leq t_i$, $\boldsymbol{Z}(t)$ is a matrix of covariates (can include both time-varying and time-invariant covariates) and $\boldsymbol{\eta}$ is the corresponding parameter vector. $\tilde{\Phi}_i$ is a zero mean ARMA-type process, and the weak stationarity condition for $\tilde{\Phi}_i$ is that all the roots of the polynomial $\beta(z) = 1 - \sum_{k=1}^{q}\beta_k z^k$ lie outside the unit circle. The innovation terms, $\tilde{\varepsilon}_i$, can be defined as:

$$\tilde{\varepsilon}_i = -\gamma - \ln\Lambda_i. \tag{2.7}$$

in which γ is the Euler-Mascheroni constant. According to Eq. (2.2), since Λ_i is i.i.d. unit exponential if the ACI model is correctly specified, $-\ln\Lambda_i$ follows a standard type-I Gumbel distribution with mean γ hence $\tilde{\varepsilon}_i$ is a zero mean martingale.

The baseline intensity component $\lambda_0(t)$ can be specified in various ways. Because a closed form solution of the integrated conditional intensity and thus the error term is more computationally convenient, the following specifications are popular: (1) the exponential baseline $\lambda_0(t) = e^w$, (2) the

Weibull baseline $\lambda_0(t) = ae^{wa}x(t)^{a-1}$ with $a > 0$ (3) the Burr-type baseline $\lambda_0(t) = e^{\kappa}\frac{ax(t)^{a-1}}{e^{-wa}+x(t)^a}$ with $a > 0$. The model can be estimated by standard maximum likelihood approach with the log-likelihood function given in Eq. (2.3).

2.3 Stationarity of the ACI Model

The stationarity concept discussed in this paper refers to weak (covariance) stationarity unless stated otherwise. Previous studies focus on discussing the stationarity of the $\tilde{\Phi}_i$ component (see Russell (1999), Hautsch (2012) for instance), and the stationarity condition of $\tilde{\Phi}_i$ is considered as a sufficient stationarity condition for the ACI model. We show that for a plain ACI model with Weibull baseline and the error term specified as in Eq. (2.7), the stationarity of $\tilde{\Phi}_i$ is insufficient for the conditional intensity process $\lambda(t_i|\mathcal{F}_{t_i})$ or the duration process x_i to be stationary. Our findings are summarized in the theorem below.

Theorem 2.1. *The sufficient conditions for the stationarity of the conditional intensity and duration processes generated by an ACI(p,q) model defined as in Eqs. (2.5) to (2.7) and $\lambda_0(t) = ae^{wa}x(t)^{a-1}$ with $\boldsymbol{Z}(t) = 0$ are:*

1. *All roots of the polynomial $\beta(z)$ lie outside the unit circle, where $\beta(z)$ is the polynomial in the lag operator form of $\tilde{\Phi}_i$: $\beta(L)\tilde{\Phi}_i = \alpha(L)\tilde{\varepsilon}_{i-1}$, in which $\beta(L) = 1-\sum_{i=1}^{p}\beta_i L^i$ and $\alpha(L) = \sum_{j=1}^{q}\alpha_j L^j$.*
2. *Let $\psi(L) = \frac{\alpha(L)}{\beta(L)} = \sum_{i=1}^{\infty}\psi_j L^j$. The following conditions are required:*

$$a > \frac{2}{3}, \quad \sup|\psi_i| < \frac{a}{2}. \tag{2.8}$$

Intuitively, condition 1 can be viewed as the stationarity condition for the ARMA component and condition 2 is the moment condition since power transformed unit exponential variables do not necessary have finite second moments. Note that the stationarity of durations does not imply the stationarity of the conditional intensity, and vice versa. We thus augment the results in Russell (1999) and Hautsch (2012) by showing that extra parameter constraints are required for the ACI model to be stationary.

3 Markov Switching ACI Model

In this section we propose the Markov Switching ACI (MS-ACI) model by augmenting the original ACI model with a Markov switching structure.

Let $\mathbb{S} = \{s_i\}_{i=1:T}, s_i \in \mathcal{M} = \{1, \ldots, M\}$ denote a M-state first order Markov chain, understood as marks attached to each point arrivals $\{t_i\}_{i=1:T}$. The Markov chain is assumed to be ergodic with the transition probability $\mathbf{P}(s_i = m|s_{i-1} = l) = \pi_{lm}$ for $l, m \in \mathcal{M}$ and an invariant probability measure

π_i. The MS(M)-ACI(p, q) model is specified as:

$$\lambda(t; \mathcal{F}_t) = \Phi(t)\lambda_0(t), \tag{3.1}$$

$$\Phi(t) = e^{\tilde{\Phi}_{\breve{N}(t)+1}(s_{\breve{N}(t)+1}) + \boldsymbol{\eta}(s_{\breve{N}(t)+1})\boldsymbol{Z}(t)}, \tag{3.2}$$

$$\tilde{\Phi}_i(s_i) = \sum_{j=1}^{q} \alpha_j(s_i)\tilde{\varepsilon}_{i-j}(s_{i-j}) + \sum_{k=1}^{p} \beta_k(s_i)\tilde{\Phi}_{i-k}(s_{i-k}), \tag{3.3}$$

$$\tilde{\varepsilon}_i(s_i) = -\gamma - \ln \int_{t_{i-1}}^{t_i} \lambda(u; \mathcal{F}_u)du, \tag{3.4}$$

$$\mathbf{P}(s_i = m | s_{i-1} = l) = \pi_{lm}, \quad l, m \in \mathcal{M}, \tag{3.5}$$

in which γ is the Euler-Mascheroni constant, $\boldsymbol{Z}(t)$ is a matrix of some possible covariates, and $\boldsymbol{\eta}(s_{\breve{N}(t)+1})$ is the corresponding regime-specific coefficient vector. The component $\boldsymbol{\eta}(s_{\breve{N}(t)+1})\boldsymbol{Z}(t)$ enables state-specific relationships between the covariates and the intensity process.

In this paper we consider a Weibull baseline function for the MS-ACI model:

$$\lambda_0(t) = a(s_{\breve{N}(t)+1})e^{w(s_{\breve{N}(t)+1})a(s_{\breve{N}(t)+1})}x(t)^{a(s_{\breve{N}(t)+1})-1}. \tag{3.6}$$

We restrict ourselves to the Weibull baseline which nests the exponential baseline, because it is in the exponential family which allows for the convergence of the Stochastic Approximation EM algorithm according to Delyon et al. (1999) and Allassonnière et al. (2010). We will denote all the dynamic and baseline parameters with the parameter vector θ and the transition parameters of the Markov chain with the parameter vector Π.

In Eq. (3.5), we assume that the transition parameters are constant over time. This can be relaxed by assuming that the transition parameters depend on a set of covariates associated with each event arrival $\{\boldsymbol{Q}_{lm,i}\}_{i=1:T,\, l,m\in\mathcal{M}}$. Following the approach in Filardo (1994), we specify the structure of $\pi_{lm,i}$ through a logistic link function:

$$\pi_{lm,i} = \frac{e^{c_{lm} + \boldsymbol{\gamma}'_{lm}\boldsymbol{Q}_{lm,i-1}}}{1 + e^{c_{lm} + \boldsymbol{\gamma}'_{lm}\boldsymbol{Q}_{lm,i-1}}}, \tag{3.7}$$

where c_{lm} controls for the baseline transition probability from state l to state m, and $\boldsymbol{\gamma}'_{lm}$ is a vector of coefficients that captures the impact of $\boldsymbol{Q}_{lm,i-1}$ on the transition probability from state l to state m at the i-th observation.

Similar to a plain ACI model, the stationarity of a MS-ACI model requires both the stationarity of $\tilde{\Phi}_i$ and a moment condition for the conditional intensity and duration. Francq and Zakoïan (2001) and Stelzer (2009) provide the strict and weak stationarity conditions for $\tilde{\Phi}_i$. For conciseness we do not present this condition, and refer the reader to Theorem 2.1 in Stelzer (2009) as a reference. For the MS-ACI model defined by Eqs. (3.1) to (3.6) and assume the weak stationarity of $\tilde{\Phi}_i$, a sufficient condition that ensures the existence of the second moment of the MS-ACI model is given by:

$$\sup_{m\in\mathcal{M}}\{a(m)\} > \frac{2}{3}, \quad \sup|\psi_i(\mathbb{S})| < \frac{1}{2}\sup_{m\in\mathcal{M}}\{a(m)\}, \tag{3.8}$$

where $\psi_i(\mathbb{S})$ are the MA(∞) coefficients of $\tilde{\Phi}_i$ given some state vector $\mathbb{S}$, and the supremum of $|\psi_i(\mathbb{S})|$ is taken over all i and all possible choices of $\mathbb{S}$. The proof of the above condition is analogous to that of Theorem 2.1 by conditioning on $\mathbb{S}$, and is thus omitted. However, the requirement on $\sup|\psi_i(\mathbb{S})|$ is difficult to verify in practice for a general MS-ACI model due to the dependence on $\mathbb{S}$. A feasible criteria can be checked based on the following relation:

$$\left.\begin{array}{l} \sup_{m\in\mathcal{M},k\in\{1,\dots,p\}}|\beta_k(m)|<1, \\ \sup_{m\in\mathcal{M},k\in\{1,\dots,q\}}|\alpha_k(m)|<\frac{1}{2}\sup_{m\in\mathcal{M}}\{a(m)\} \end{array}\right\} \Rightarrow \sup|\psi_i(\mathbb{S})|<\frac{1}{2}\sup_{m\in\mathcal{M}}\{a(m)\}, \tag{3.9}$$

which holds true due to the MS-ARMA structure of $\tilde{\Phi}_i$.

3.1 Model Estimation

We rely on the maximum likelihood estimation (MLE) method to estimate the parameter vector θ and the state probability parameters Π. A standard implementation of MLE maximizes the observed log-likelihood of the data, which is the marginal log-likelihood of the observed data $\mathbb{Y}$ given θ:

$$\ln\mathcal{L}(\theta;\mathbb{Y})=\ln\sum_{\mathbb{S}}\mathcal{L}(\theta,\Pi;\mathbb{Y},\mathbb{S}). \tag{3.10}$$

Note that all covariates in the above likelihood function are considered as conditionally exogenous and omitted for conciseness. Eq. (3.10) is empirically very difficult to maximize for two reasons: (1) the functional form of the likelihood function is difficult to maximize by standard gradient/score methods; (2) the dimensionality of $\mathbb{S}$ growth exponentially and summing over the entire space of $\mathbb{S}$ becomes infeasible even in a relatively small sample. We hence apply the Stochastic Approximation Expectation-Maximization (SAEM) algorithm developed by Celeux and Diebolt (1992) and further analysed by Kuhn and Lavielle (2004) to overcome the two difficulties in maximizing Eq. (3.10).

We firstly explain three relevant likelihood functions in the estimation process. The conditional log-likelihood of $\mathbb{Y}$ given the state vector $\mathbb{S}$:

$$\ln\mathcal{L}(\theta;\mathbb{Y}|\mathbb{S})=\sum_{i=1}^{T}\Big[-\Lambda_i+\ln\lambda(t_i|\mathcal{F}_{t_i})\Big]. \tag{3.11}$$

This log-likelihood can be easily maximized since it is in a log-linear form. The complete data log-likelihood for the joint density of $\{\mathbb{Y},\mathbb{S}\}$ can be decomposed as:

$$\ln\mathcal{L}(\theta,\Pi;\mathbb{Y},\mathbb{S})=\ln\mathcal{L}(\theta;\mathbb{Y}|\mathbb{S})+\ln\mathcal{L}(\Pi;\mathbb{S}), \tag{3.12}$$

in which $\ln\mathcal{L}(\Pi;\mathbb{S})$ is the marginal log-likelihood for the Markov chain, given by:

$$\ln\mathcal{L}(\Pi;\mathbb{S})=\sum_{i=2}^{T}\ln\pi_{s_{i-1}s_i,i}. \tag{3.13}$$

Note that since Π is independent of $\mathcal{L}(\theta; \mathbb{Y}|\mathbb{S})$, the complete likelihood can be maximized by separately maximising $\mathcal{L}(\theta; \mathbb{Y}|\mathbb{S})$ and $\mathcal{L}(\Pi; \mathbb{S})$. Both log-likelihood functions can be maximized by gradient-based algorithms such as the Newton-Raphson method. We provide the analytical gradient of $\ln \mathcal{L}(\theta; \mathbb{Y}|\mathbb{S})$ in Appendix B, which allows for a faster and more efficient optimization compared to pure numerical algorithms. The marginal likelihood of $\mathbb{Y}$, which cannot be directly maximized as discussed in the previous section, can be expressed as:

$$\ln \mathcal{L}(\theta; \mathbb{Y}) = \ln \sum_{\mathbb{S}} \mathcal{L}(\theta; \mathbb{Y}|\mathbb{S}) \mathcal{L}(\Pi; \mathbb{S}). \tag{3.14}$$

To apply the SAEM algorithm, we draw a random sample from the conditional density of the state given the current parameter estimate $\theta^{(n)}$, the data $\mathbb{Y}$ and the current state vector $\mathbb{S}^{(n)}$ using the single move Gibbs sampler developed by Bauwens et al. (2010). The single move sampler exploits the following conditional density of the i-th state:

$$p(s_i | s_{1:i-1}^{(n+1)}, s_{i+1:T}^{(n)}, \theta^{(n)}, \mathbb{Y}) \propto p(s_i | s_{i-1}^{(n+1)}, s_{i+1}^{(n)}, \Pi^{(n)}) f(y_{i:T} | s_i, s_{1:i-1}^{(n+1)}, s_{i+1:T}^{(n)}, \theta^{(n)}). \tag{3.15}$$

At every n, by iterating i from 2 to T, we obtain a random draw of the vector $\mathbb{S}^{(n+1)}$ conditioning on the previous state vector. Based on the Gibbs sampler we can implement the SAEM algorithm to estimate the MS-ACI model, which we summarize as follows.

Simulation Step (S-step): At the n-th iteration, given the current parameter estimate $\theta^{(n)}$, $\Pi^{(n)}$ and the current draws of state $\mathbb{S}^{(n,k)}$, draw $\mathbb{S}^{(n+1,k)}$ from the following density for $i = 1 : T$ and $k = 1 : K_n$:

$$p(s_i = l | s_{-i}^{(n,k)}, \theta^{(n)}, \mathbb{Y}) = \frac{p(s_i = l | s_{i-1}^{(n+1,k)}, s_{i+1}^{(n,k)}, \Pi^{(n)}) f(y_{i:T} | s_i = l, s_{-i}^{(n,k)}, \theta^{(i)})}{\sum_m^{\mathcal{M}} p(s_i = m | s_{i-1}^{(n+1,k)}, s_{i+1}^{(n,k)}, \Pi^{(n)}) f(y_{i:T} | s_i = m, s_{-i}^{(n,k)}, \theta^{(i)})}, \tag{3.16}$$

in which $s_{-i}^{(n,k)} = \{s_{1:i-1}^{(n+1,k)}\} \cup \{s_{i+1:T}^{(n,k)}\}$.

Stochastic Approximation: Let $\vartheta^{(n)}$ denote the vector that combines $\theta^{(n)}$ and all free probability parameters in $\Pi^{(n)}$, update the quantity:

$$Q_n(\vartheta | \vartheta^{(n)}) = (1 - \gamma_n) Q_{n-1}(\vartheta | \vartheta^{(n-1)}) + \frac{\gamma_n}{K_n} \sum_{k=1}^{K_n} \ln \mathcal{L}(\vartheta; \mathbb{Y}, \mathbb{S}^{(n+1,k)}), \tag{3.17}$$

in which γ_n is a positive step size that gradually decreases to zero as $n \to \infty$.

Maximization Step (M-step): Maximize $Q_n(\vartheta | \vartheta^{(n)})$ w.r.t. ϑ to obtain $\vartheta^{(n+1)}$. Repeat until a termination criterion is reached.

Note that maximizing $Q_n(\vartheta | \vartheta^{(n)})$ w.r.t. ϑ is equivalent to maximizing a weighted sum of complete log-likelihoods. Therefore it can be decomposed into a weighted sum of conditional log-likelihoods of the data given the state vector and a weighted sum of log-likelihoods of simulated Markov chains, both of which can be maximized using gradient-based algorithms.

We now discuss the choices of tuning parameters. For the choice of the step size γ_n, the general guideline is that, as explained by Jank (2006), small (large) step sizes reduce (inflate) the Monte Carlo error and yield slower (faster) convergence. Our choice step size γ_n is of the following form:

$$\gamma_n = \begin{cases} 1, & n \leq n_0, \\ \frac{1}{(n-n_0)^{0.75}}, & n > n_0. \end{cases} \tag{3.18}$$

The initial n_0 steps are understood as a burn-in period which will produce a sequence of parameter estimates $\{\vartheta^{(i)}\}_{i=1:n_0}$ that will converge fast but have large Monte Carlo error. We then use a local average of parameters from the burn-in period as the starting values of the SAEM iterations, and 0.75 is a moderate convergence speed that balances the Monte Carlo error and the convergence speed of the algorithm.

As to K_n, it controls for the Monte Carlo size drawn at each iteration. Intuitively, the term $\frac{1}{K_n}\sum_{k=1}^{K_n} \ln \mathcal{L}(\vartheta; \mathbb{Y}, \mathbb{S}^{(n+1,k)})$ is a Monte Carlo approximation of $\ln \mathcal{L}(\theta; \mathbb{Y})$. Larger K_n accelerates the speed of convergence but also increases the computational burden. Unlike the Monte Carlo EM of Wei and Tanner (1990) which requires $K_n \to \infty$ for the algorithm to converge, in the SAEM algorithm only a fixed K_n is needed. In our implementation we choose $K_n = 1$ in the burn-in period and $K_n = 20$ in SAEM iterations, which are chosen to fit our computational power and time constraints.

To ensure that the algorithm does not terminate prematurely, we set the termination criterion as follows:

$$\Big|\Big|\frac{1}{H}\sum_{h=1}^{H}(\vartheta^{(n-h+1)} - \vartheta^{(n-H-h+1)})\Big|\Big|_\infty \leq \epsilon, \tag{3.19}$$

where $||\cdot||_\infty$ denotes the supremum norm. This criterion compares the average parameter estimates calculated from the most recent H iterations to those using H estimates prior to the most recent H iterations. If the algorithm converges, we would expect that the criteria is close to zero. In our analysis we choose $H = 5$ and $\epsilon = 5e^{-3}$.

The initial values of the burn-in period, namely $\vartheta^{(0)}$ and $\mathbb{S}^{(0)}$, are important factors in the estimation procedure. We find that the estimation scheme is robust to choices of $\mathbb{S}^{(0)}$, $\Pi^{(0)}$ and all ARMA parameters, but depends crucially on the baseline parameters. This is due to the fact that the regime identification is to a large extent determined by the baseline parameters. Guidances on choices of the initial values are provided in Section 4.

The SAEM algorithm only provides a point estimate for the parameter vector. As suggested by Delyon et al. (1999) and Kuhn and Lavielle (2004), we can also obtain variance-covariance matrix estimates for the parameter estimates. From the posterior probability of the states conditioning on the parameter estimates, we are able to provide an estimate of the most probable state vector. The detailed estimation procedures for the variance-covariance matrix and the most probable state vector are presented in Appendix C.

3.2 Post-Estimation Diagnostics

According to the RTCT, the transformed residual process $\{\hat{\Lambda}_i\}_{i=1:T}$ follows an i.i.d. unit exponential process if the model specification is correct. As a result, diagnostic tests of the original ACI model usually involve testing the unit exponentiality and the presence of autocorrelation in $\{\hat{\Lambda}_i\}_{i=1:T}$. These tests, however, are not directly applicable to the MS-ACI model with only parameter estimates $\hat{\vartheta}$ from the SAEM algorithm. This is because a state vector $\mathbb{S}$ is required for the residual series to be constructed. Thanks to the estimation of the most probable state sequence, we are able to obtain a conditional residual series $\{\hat{\Lambda}_i|\hat{\mathbb{S}}\}_{i=1:T}$ by plugging in the estimated (most probable) state sequence $\hat{\mathbb{S}}$, which allows us to use standard diagnostic tests such as Ljung-Box tests or empirical density function tests to evaluate the goodness-of-fit of our model. However, distribution of these tests statistics may become non-standard due to the conditioning and thus test results only provide indicative rather than conclusive evidence.

To provide a descriptive statistic that reflects the strength of the regime identification, we focus on the $T \times M$ posterior probability matrix $\mathcal{P}$ conditioning on the estimated parameters $\hat{\vartheta}$ and the estimated state vector $\hat{\mathbb{S}}$. The element at the intersection of row i and column m in $\mathcal{P}$, denoted by $\mathcal{P}_{i,m} = p(s_i = m|\hat{s}_{1:i-1}, \hat{s}_{i+1:T}, \mathbb{Y}, \hat{\vartheta})$, is the posterior probability of the i-th state being classified as state m conditioning on $\hat{s}_{1:i-1}$, $\hat{s}_{i+1:T}$, $\mathbb{Y}$ and $\hat{\vartheta}$ calculated similarly as in Eq. (3.15). Based on this matrix, we construct a statistic named the 'Significance of Regimes' (SoR hereafter) which serves as an indicator of the overall significance of the regime-switching structure. It is calculated as follows:

$$SoR = T^{-1} \sum_{i=1}^{T} \max_{m \in \mathcal{M}} \mathcal{P}_{i,m}. \tag{3.20}$$

Intuitively, SoR is the average of the largest probability in every row of $\mathcal{P}$. It measures the average (conditional) probability of each state being classified into the most probable states. The rationale behind this statistic is that, assuming the DGP consists of M distinct regimes with densities far apart from each other, the probability of any observation being classified into its corresponding true state, hence the SoR, will be close to one. On the contrary, when all M densities are identical, all the elements in the matrix $\mathcal{P}$, and therefore the SoR, reduce to M^{-1}. Therefore, conditioning on $\hat{\mathbb{S}}$ and $\hat{\vartheta}$, a SoR close to 1 indicates that every observation is being assigned to a regime with a probability close to 1, hence the regime identification is strong. Nevertheless, when SoR is close to its lower bound M^{-1}, the model is indifferent about assigning each observation into any of the M regimes (so the probability of being assigned to any regime is M^{-1}), which implies a very weak regime identification. The SoR allows for easy comparisons across models with different number of regimes and baseline specifications. Moreover, we can calculate SoR for each regime to compare their relative significance. The SoR for the l-th regime is defined as:

$$SoR(l) = \Big(\sum_{i=1}^{T} \mathbb{1}_{\{\arg\max_{m \in \mathcal{M}}(\mathcal{P}_{i,m}) = l\}} \Big)^{-1} \sum_{i=1}^{T} \max_{m \in \mathcal{M}} \mathcal{P}_{i,m} \mathbb{1}_{\{\arg\max_{m \in \mathcal{M}}(\mathcal{P}_{i,m}) = l\}}, \tag{3.21}$$

which basically calculates the average of the largest element in every row of $\mathcal{P}$ if the largest element belongs to the l-th state. We provide some guidance on how to interpret the SoR in the following section.

4 Monte Carlo Simulation Study

In this section, we conduct a Monte Carlo simulation study to highlight the estimation performance of our SAEM algorithm under various parameter specifications. We mainly focus on the impact of different baseline parameters and transition parameters based on MS(2)-ACI(1,1) models without covariates.

The selection of parameter constellations is designed to examine the aforementioned relationship between specifications and estimation quality. The size of the Monte Carlo study for each specification is 1000. For each specification, we choose $T = 1000$ and $n_0 = 10$. We initialize $\alpha_1^{(0)}(1) = \alpha_1^{(0)}(2) = 0$, $\beta_1^{(0)}(1) = \beta_1^{(0)}(2) = \pi_{11}^{(0)} = \pi_{22}^{(0)} = 0.5$ and $\mathbb{S}^{(0)}$ to be a random draw of T independent fair coin tosses. The initial values of the scale parameter $w^{(0)}(l)$ are carefully set to mitigate the label-switching problem. Specifically, in the two-regime case, assume the true scale parameters $w(1) > w(2)$, for $\hat{w}(1)$ to converge towards $w(1)$ (instead of $w(2)$ or being eliminated by the algorithm), we recommend setting $w^{(0)}(1) > w(1) > w(2) > w^{(0)}(2)$ or $w(1) > w^{(0)}(1) \gg w^{(0)}(2) > w(2)$. On a Windows machine with a 2.3GHz CPU, it takes about 1-3 minutes to obtain one set of parameter estimates based on our SAEM algorithm implemented in MATLAB R2020a[2].

The quality of the parameter estimates is assessed through the bias and the root mean squared error (RMSE) of the Monte Carlo parameter estimates. In detail, for each parameter specification, we benchmark the bias and the RMSE of the Monte Carlo parameter estimates against the bias and the RMSE from a Monte Carlo simulation of the same model given the latent state vector (referred to as the complete model hereafter, and the incomplete model refers to the model with a latent state vector). As parameter estimates from the complete model are statistically optimal, we are able to depict the loss of accuracy associated with both parameter specifications and the observability of the Markov chain by comparing the quality of parameter estimates of the incomplete model to its corresponding complete counterpart.

The main results are presented in Table 1. We base our simulation on six different specifications, with spec. 1-5 using exponential baselines and spec. 6 using a Weibull baseline. The first part of the table shows the bias and RMSE of the Monte Carlo parameter estimates of the incomplete model, with the bias and RMSE of the complete counterparts included in the second part of the table. The

[2]Note that the time consumption is depends on the specification of the model. In detail, models with a larger SoR converge faster as there are less randomness in the parameter estimates. Also, the amount of computation grows quadratically with the number of observations due to the single move Gibbs sampler which sweeps through the entire sample. For a sample size of 3000, the required time to obtain 1 set of parameter estimates is about 10-15 minutes.

Table 1: Monte Carlo simulation results of parameter estimates of MS(2)-ACI(1,1) models and the corresponding complete models based on 1000 random draws

Spec.	$\tilde{\alpha}_1(1)$ Bias RMSE	$\tilde{\alpha}_1(2)$ Bias RMSE	$\tilde{\beta}_1(1)$ Bias RMSE	$\tilde{\beta}_1(2)$ Bias RMSE	$\tilde{w}(1)$ Bias RMSE	$\tilde{w}(2)$ Bias RMSE	$\tilde{a}(1)$ Bias RMSE	$\tilde{a}(2)$ Bias RMSE	$\tilde{\pi}_{11}$ Bias RMSE	$\tilde{\pi}_{22}$ Bias RMSE	$\overline{SoR}$ $\sigma(SoR)$
						Results for Incomplete Models					
1	0.05	0.1	0.8	0.95	0	-5	1	1	0.5	0.5	
	-0.0019	0.0028	-0.0092	-0.0618	0.0020	0.0039	.	.	-0.0010	-0.0011	0.9805
	0.0328	0.0309	0.1668	0.1604	0.0654	0.0621	.	.	0.0248	0.0247	0.0019
2	0.05	0.1	0.8	0.95	0	-3	1	1	0.5	0.5	
	0.0030	0.0030	-0.0371	-0.0878	0.0051	0.0034	.	.	-0.0008	-0.0012	0.9084
	0.0479	0.0422	0.2241	0.2150	0.0802	0.0640	.	.	0.0319	0.0309	0.0059
3	0.05	0.1	0.8	0.95	0	-3	1	1	0.9	0.9	
	-0.0009	0.0066	-0.1032	-0.0305	0.0088	0.0195	.	.	-0.0049	-0.0043	0.9759
	0.0366	0.0343	0.2637	0.0910	0.0650	0.0847	.	.	0.0169	0.0167	0.0041
4	0.05	0.1	0.8	0.95	0	-1	1	1	0.5	0.5	
	0.0041	0.0028	0.0493	-0.1126	-0.1899	-0.1520	.	.	0.1832	-0.1318	0.7549
	0.0434	0.0494	0.1539	0.2194	0.2158	0.1778	.	.	0.1998	0.1608	0.0352
5	0.05	0.1	0.8	0.95	0	0	1	1	0.9	0.9	
	0.0267	-0.0217	0.0960	-0.0955	0.1301	-0.1952	.	.	-0.2018	-0.5489	0.6986
	0.0410	0.0530	0.1284	0.1737	0.1515	0.2154	.	.	0.2075	0.5543	0.0385
6	0.05	0.1	0.8	0.95	0	-1	1	2	0.5	0.5	
	-0.0014	0.0062	-0.0205	-0.0680	0.0353	-0.0123	0.0290	0.0702	0.0008	0.0036	0.7827
	0.0508	0.0456	0.1997	0.1753	0.1032	0.0463	0.0611	0.1503	0.0546	0.0561	0.0234
						Results for Complete Models					
1	0.05	0.1	0.8	0.95	0	-5	1	1	0.5	0.5	
	-0.0018	0.0014	0.0042	-0.0365	0.0068	0.0063	.	.	-0.0013	-0.0011	
	0.0256	0.0262	0.1141	0.1045	0.0592	0.0624	.	.	0.0221	0.0221	
2	0.05	0.1	0.8	0.95	0	-3	1	1	0.5	0.5	
	-0.0004	0.0021	-0.0005	-0.0428	0.0054	0.0072	.	.	-0.0006	-0.0005	
	0.0256	0.0256	0.1148	0.1189	0.0569	0.0619	.	.	0.0224	0.0228	
3	0.05	0.1	0.8	0.95	0	-3	1	1	0.9	0.9	
	-0.0001	0.0007	-0.0350	-0.0142	0.0033	0.0030	.	.	-0.0010	-0.0005	
	0.0270	0.0238	0.1563	0.0558	0.0579	0.0765	.	.	0.0136	0.0138	
4	0.05	0.1	0.8	0.95	0	-1	1	1	0.5	0.5	
	-0.0009	0.0021	-0.0031	-0.0369	0.0044	0.0032	.	.	0.0000	-0.0006	
	0.0253	0.0265	0.1193	0.1044	0.0607	0.0631	.	.	0.0213	0.0220	
5	0.05	0.1	0.8	0.95	0	0	1	1	0.9	0.9	
	-0.0010	0.0006	-0.0303	-0.0148	0.0061	0.0047	.	.	-0.0001	-0.0011	
	0.0269	0.0235	0.1387	0.0593	0.0562	0.0763	.	.	0.0137	0.0138	
6	0.05	0.1	0.8	0.95	0	-1	1	2	0.5	0.5	
	-0.0004	0.0025	-0.0049	-0.0365	-0.0007	0.0017	0.0071	0.0097	-0.0013	-0.0008	
	0.0246	0.0258	0.1184	0.1126	0.0636	0.0336	0.0366	0.0727	0.0224	0.0218	
						RMSE Ratios					
1	1.2785	1.1790	1.4619	1.5354	1.1060	0.9939	.	.	1.1252	1.1167	
2	1.8702	1.6479	1.9516	1.8081	1.4099	1.0354	.	.	1.4235	1.3592	
3	1.3576	1.4404	1.6870	1.6327	1.1224	1.1068	.	.	1.2438	1.2064	
4	1.7152	1.8597	1.2906	2.1021	3.5552	2.8180	.	.	9.3974	7.3268	
5	1.5217	2.2556	0.9261	2.9292	2.6942	2.8216	.	.	15.1995	40.2603	
6	2.0667	1.7680	1.6866	1.5572	1.6217	1.3775	1.6694	2.0692	2.4374	2.5743	

Note: $\tilde{\theta}$ denotes the DGP value of θ. $\overline{SoR}$ and $\sigma(SoR)$ are the mean and standard deviation of the significance of regime defined in equation Eq. (3.20). RMSE is the root mean square error of the Monte Carlo parameter estimates. The RMSE ratio section presents the RMSE of the corresponding parameter of the incomplete model divided by the RMSE of its complete counterpart.

RMSE ratio calculated by dividing the RMSE of each parameter estimate of the incomplete model by that of the complete model allows for an easy comparison. Generally, the bias and RMSE of the incomplete model tend to fluctuate across specifications, while those for the complete model are much more consistent.

Results for the incomplete model shows that, by comparing spec. 1, 2 and 4, the closer the gap between the two regime-specific densities implied by the scale parameters $\tilde{w}(1)$ and $\tilde{w}(2)$, the smaller the mean SoR. According to the RMSE ratios, it is evident that the quality of the parameter estimates deteriorates with the mean SoR implied by the increasingly narrowing gap between the DGP scale parameters.

A more persistent latent Markov chain improves the mean SoR and the estimation performance, as can be seen from the RMSE ratios of spec. 2 and 3. However, a persistent latent Markov chain alone cannot guarantee reliable parameter estimates, as is shown in spec. 5. It is worth mentioning that in spec. 5, the regime-specific scale parameters are identical, and the only information on the latent state vector implied by the observable data is the differences in the ARMA parameters, which do not generate enough discrepancy in the density for the latent Markov chain to be well identified. This is explained by the large downward biases in the transitional parameters in spec. 5. Moreover, the mean parameter estimates for the ARMA parameters of the incomplete model are very close to each other, suggesting a label-switching problem caused by a common baseline.

Finally, comparing spec. 4 and 6, we observe that a more complex baseline function can improve the mean SoR by generating more flexible discrepancies between the regime-specific densities. In spec. 6, by changing the Weibull parameter, the Markov chain parameters are estimated with much better accuracy compared to spec. 4 (smaller bias and RMSE ratio).

An important message from Table 1 is that, since the latent Markov chain will inevitably result in a loss of information, the quality of the parameter estimates depends crucially on how much we can learn about the latent states from the observed data. Complexity of baselines and the persistence of the latent Markov chain largely influence the amount of information available in the observed data, and will have a significant impact on the quality of the parameter estimates. However, the predominant factors are the DGP scale parameters of the baseline functions, which generally control the location of the distribution and the ability of our sampler to correctly classify the states. This rationalizes the use of the SoR as an indicator of the quality of the parameter estimates, as it is a measure of confidence of the Gibb's sampler in classifying the observations. For reliable parameter estimates of the MS-ACI model, we recommend a SoR of at least 90% so that the hidden Markov chain is on average properly identified (as can be observed from the RMSE ratios for the last three specifications).

5 Application to High-Frequency Stock Prices

We apply the MS-ACI model to high-frequency stock prices to analyse regime shifts in the intraday trading activities. Our sample comprises of 9 highly liquid stocks and a stock index ETF traded on the U.S. stock market, namely AIG, CVX, GM, INTC, JPM, PFE, T, VZ, WMT and SPY, whose detailed descriptions are shown in Table 2. The raw data is obtained from the Trade and Quote[3] (TAQ) dataset, and is cleaned according to Holden and Jacobsen (2014) and Barndorff-Nielsen et al. (2009). The sampling period ranges from 1 Jan 2016 to 31 Dec 2016. The trade dataset consists of prices and trading volumes timestamped at milliseconds, and the trades are matched with the quotes using the Lee and Ready's (1991) algorithm to determine trade directions and prevailing bid-ask spread at each trade. To conserve space, we mainly focus on the empirical results for AIG, VZ and SPY, which are representative for all assets considered in our sample. Detailed empirical results for all 10 stocks are relegated to the supplementary material of the paper, which also includes empirical analyses based on an extended sample of 24 individual stocks that cover all Dow Jones constituents in 2016.

Table 2: Description of the sampled stocks/ETFs and their ticker symbols

Ticker	Corporate/ETF name	DJ 2016
AIG	American International Group, Inc.	N
CVX	Chevron Corporation	Y
GM	General Motors Corporation	N
INTC	Intel Corporation	Y
JPM	JPMorgan Chase & Co.	Y
PFE	Pfizer Inc.	Y
T	AT&T Inc.	N
VZ	Verison Communications Inc.	Y
WMT	Walmart Inc.	Y
SPY	SPDR S&P500 ETF Trust	N

Note: the DJ 2016 column indicates whether the stock belongs to the Dow Jones 30 index constituents in 2016.

We apply our MS-ACI model to the point process of absolute price change events constructed as follows. For a sequence of trade observations with its associated log-price $\{t_j, P(t_j)\}_{j=1:J}$, we construct a sequence of 'price events' based on a price change threshold δ using the following algorithm:

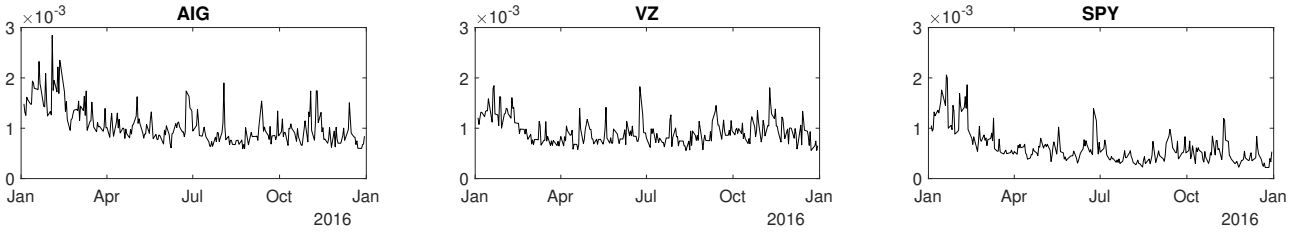
1. From $j = 1$, set $t_0 = t_1$. Set the value of δ.
2. Let $t_i^{(\delta)} = \inf_{t_j > t_{i-1}^{(\delta)}} \{|P(t_j) - P(t_{i-1}^{(\delta)})| \geq \delta\}$.
3. Iterate until the end of the sample.

[3] https://www.nyse.com/market-data/historical/daily-taq

We use mid-quotes in our empirical analysis to alleviate the problem of bid-ask bounce effects. The process $\{t_i^{(\delta)}\}_{i=1:T}$ records the arrival times of each price event, and is known as the δ-related absolute price change process. We will also denote $x_i^{(\delta)} \equiv t_i^{(\delta)} - t_{i-1}^{(\delta)}$ as the price duration between two price events. Since we only sample the price process whenever the absolute price change is equal or larger than δ, intuitively the accumulated price volatility is approximately δ^2 for each price duration. In fact, under a continuous martingale framework, it is proved by Li et al. (2019) proved that the integrated variance between consecutive price events are i.i.d. with mean δ^2 and variance $2\delta^4/3$. Therefore, we can interpret price durations as an inverse measure of average spot volatility within each interval.

For our analysis, we choose a daily δ which produces a mean daily price duration that is closest to a five-minute interval. Empirical results in Tse and Yang (2012) show that volatility estimates constructed based on this sampling frequency perform reasonably well against noise-robust measures of RV estimates. Moreover, this choice of δ forces the daily number of price events to be identical up to a random discretization perturbation, which removes daily dynamics from the price durations and thus greatly simplifies our model specification. In fact, the daily δ^2 is an estimate of the average 5-minute integrated variance under the continuous martingale price assumption. Therefore, each price duration can be interpreted as the time elapse for the volatility to accumulate by 1/78 of the day's total volatility. An example of daily choices of δ is presented in Figure 1.

Figure 1: Daily choices of δ for AIG, VZ and SPY



Note: The figure plots daily price change threshold δ for AIG, CVX and SPY. For each asset, δ is calculated as the maximum threshold that produces a mean price duration which is closest to a 5-minute interval.

In this paper, we demonstrate that there exists an intraday regime-switching behaviour between price durations ($x_i^{(\delta)}$) and the cumulative trading volume within each price duration (Vol_i). Due to the well-documented diurnal pattern in the intraday trading process, we firstly extract the diurnal component of $x_i^{(\delta)}$ and Vol_i using a flexible Fourier regression estimated on a monthly basis following the approach in Andersen and Bollerslev (1997) and Engle and Russell (1998). Taking $x_i^{(\delta)}$ as an example:

$$\frac{x_i^{(\delta)}}{\bar{x}_i^{(\delta)}} = c_0 + c_1 \bar{t}_{i-1} + \sum_{n=1}^{3} c_{s,n} \sin(2n\pi \bar{t}_{i-1}) + c_{c,n} \cos(2n\pi \bar{t}_{i-1}) + \epsilon_i, \tag{5.1}$$

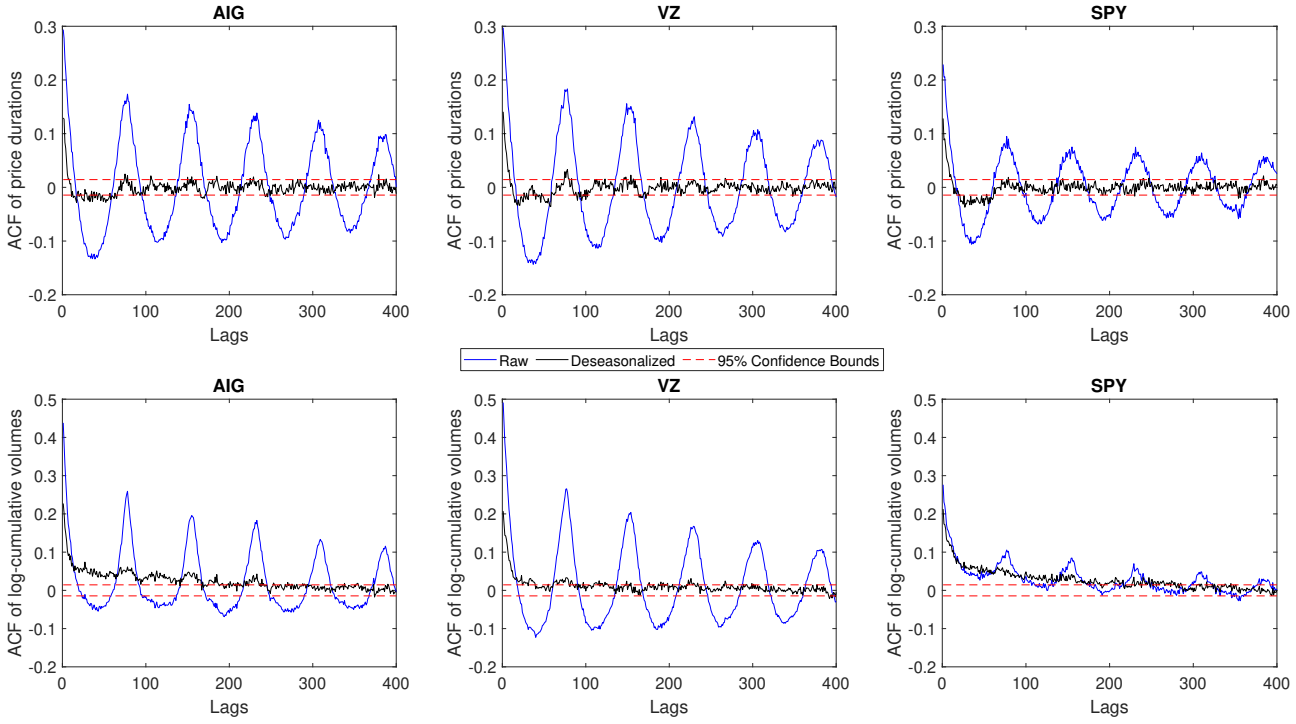
where $\bar{t}_i$ is the fraction of trading hours normalized to be between [0,1], and $\bar{x}_i^{(\delta)}$ is the sample mean of the price durations in the regression above. Let us denote by $\hat{s}_i$ the fitted values from Eq. (5.1), the

deseasonalized price duration is then computed as:

$$\tilde{x}_i^{(\delta)} = x_i^{(\delta)} / \hat{s}_i.$$

For the cumulative volume, we firstly compute $\ln Vol_i$ and deseasonalize it using Eq. (5.1) to obtain $\ln \tilde{Vol}_i$. We plot the average monthly autocorrelations for raw and deseasonalized price durations and log-volumes in Figure 2. The figure shows clearly that both raw duration and volume have a sinusoid-shaped autocorrelation function which peaks at multiples of roughly 78, the average number of price durations per day. This is completely removed by our deseasonalization procedure. The deseasonalized price durations do not appear to have long memory, which indicates that the long-range dependence of volatility is subsumed into our daily δ sequences.

Figure 2: Correlograms of the raw and deseasonalized price duration and log-cumulative volume



Note: The figure plots monthly averaged autocorrelations of raw and deseasonalized price durations and log-cumulative volumes for AIG, VZ and SPY.

We fit a MS(2)-ACI(1,1) model to the deseasonalized price durations series on a monthly basis to capture the regime-switching behaviour of price durations upon information arrivals. The model is summarized by Eqs. (3.1) to (3.4), (3.6) and (3.7) with $p = q = 1$. To capture a regime-switching contemporaneous duration-volume relationship, we specify $\boldsymbol{Z}(t) = \ln \widetilde{Vol}_{\breve{N}(t)+1}$. We set the transition probability covariates $\boldsymbol{Q}_{11,i} = \boldsymbol{Q}_{22,i} = \ln BAS_i$, where BAS_i is the prevailing bid-ask spread at time $t_i^{(\delta)}$. The parameters being estimated are: $\vartheta = \{\alpha_1(l), \beta_1(l), a(l), w(l), \eta(l), c_{ll}, \gamma_{ll}\}'_{l\in\{1,2\}}$, where $\eta(l)$ captures the impact of contemporaneous log-volume on the conditional intensity in state l and γ_{ll} captures the impact of the bid-ask spread on the transitional probability from state l to l.

To reveal some basic features of the data used in the model estimation, we present descriptive statistics of the deseasonalized durations and volumes along with the bid-ask spread for the 10 sampled assets in Table 3. The table shows that the mean duration is indeed close to 5 minute as expected. Interestingly, the moments and quantiles of $\tilde{x}_i^{(\delta)}$ are very similar across all assets, which is a consequence of the daily δ sequences that homogenize the price durations across stock-days. SPY's mean $\ln \widetilde{Vol}_i$ is considerably higher than the individual stocks as it is much more heavily traded in the market. Finally, BAS_i are highly skewed to the right with a large kurtosis, which is due to the fact that the prevailing bid-ask spread for most of the price durations (at least 75% for most of the sampled stocks) are \$0.01, which is the smallest possible bid-ask spread.

For illustrative purposes, we present our estimation outputs for three representative stock-months in Table 4. In the table, we can see that for AIG and VZ, our model detects two distinct regimes with high $\widehat{SoR}$ (overall $> 98\%$). Regime 1 only accounts for a small proportion (less than 10%) of the data, with a less persistent intensity dynamics ($\hat{\beta}_1(1) < \hat{\beta}_1(2)$) and a much smaller interaction with trading volume ($|\hat{\eta}(1)| < |\hat{\eta}(2)|$). The results on $\hat{c}_{11}$ and $\hat{\gamma}_{11}$ is somewhat mixed, but $\hat{\gamma}_{22}$ are negative and significant for both stocks, which suggest that durations with larger prevailing bid-ask spreads are less likely to be assigned into regime 2.

As to the estimation output for SPY, our findings are drastically different compared to the individual stocks. Firstly, the $\widehat{SoR}$ is much smaller relative to that of AIG and VZ, suggesting that the model is less confident about the assignment of regimes. The estimated $\hat{\eta}(l)$ are similar for both regimes. The bid-ask spread does not seem to play any role in the classification of regimes, resulting in insignificant $\hat{\gamma}_{ll}$ estimates. In fact, since $\hat{c}_{11} \approx \hat{c}_{22}$, the model suggests a fair coin toss for the regime classification. Summarizing from the diagnostics statistics of the three estimation outputs, the MS-ACI model performs reasonably well in capturing the dynamics of price durations.

To demonstrate that the regimes we detect are consistent in the sampling period considered, we present the evolution of monthly parameter estimates of $\hat{\eta}(l)$, $\hat{\gamma}_{ll}$ and $\widehat{SoR}(l)$ in Figure 3 for AIG, VZ and SPY. From the figure, we can clearly see that both regimes for AIG and VZ behave consistently throughout the sampling period, with parameter estimates of $\hat{\eta}(l)$ and $\hat{\gamma}_{ll}$ following the pattern in Table 1. The estimated $\widehat{SoR}(l)$ are all above 90% and in fact very close to 1, indicating a very strong regime identification of the MS-ACI model. For SPY, the discrepancy between of $\hat{\eta}(1)$ and $\hat{\eta}(2)$ is much weaker, and the estimates of $\hat{\gamma}_{ll}$ appear noisier comparing to AIG and VZ. Importantly, $\widehat{SoR}(l)$ for SPY is significantly smaller than that of AIG and VZ except for the result in 2016-11, consistent with our findings in Table 1. More comprehensive analyses in the supplementary material of the paper show that this finding is robust to the choice of δ, and in general hold true for all individual stocks considered in our sample and the extended sample.

The most interesting finding from Table 4 and Figure 3 is the difference between the regimes we

Table 3: Descriptive statistics of $\tilde{x}_i^{(\delta)}$, $\ln \widetilde{Vol}_i$ and BAS_i for the sampled assets

Ticker	Variable	#Obs.	Mean	Std. Dev.	Skew.	Kurt.	Min	Q(25%)	Median	Q(75%)	Max
	$\tilde{x}_i^{(\delta)}$	19379	297.498	315.326	2.931	18.258	0.000	95.296	201.589	389.780	4687.967
AIG	$\ln \widetilde{Vol}_i$	19379	10.365	1.181	-0.382	5.009	0.798	9.637	10.396	11.122	15.990
	BAS_i	19379	0.017	0.031	8.240	86.933	0.010	0.010	0.010	0.010	0.490
	$\tilde{x}_i^{(\delta)}$	19595	298.593	306.085	2.619	14.403	0.000	98.037	203.700	393.132	3960.754
CVX	$\ln \widetilde{Vol}_i$	19595	10.662	1.102	-0.415	4.657	3.545	9.989	10.693	11.382	16.400
	BAS_i	19595	0.027	0.042	5.678	43.757	0.010	0.010	0.020	0.030	0.570
	$\tilde{x}_i^{(\delta)}$	19078	298.537	338.925	3.386	25.589	0.001	89.036	193.135	384.597	6247.964
GM	$\ln \widetilde{Vol}_i$	19078	11.026	1.165	-0.351	4.346	2.301	10.312	11.056	11.780	16.104
	BAS_i	19078	0.013	0.014	11.146	169.889	0.010	0.010	0.010	0.010	0.360
	$\tilde{x}_i^{(\delta)}$	18677	301.785	353.832	3.314	22.753	0.000	83.880	191.422	391.219	6134.462
INTC	$\ln \widetilde{Vol}_i$	18677	11.586	1.208	-0.305	5.477	1.871	10.870	11.613	12.323	18.512
	BAS_i	18677	0.012	0.014	9.377	115.052	0.010	0.010	0.010	0.010	0.310
	$\tilde{x}_i^{(\delta)}$	19497	297.057	323.500	3.428	27.295	0.000	95.010	198.490	384.909	6127.337
JPM	$\ln \widetilde{Vol}_i$	19497	11.370	1.089	-0.390	5.080	0.000	10.691	11.394	12.080	16.758
	BAS_i	19497	0.014	0.019	10.259	154.256	0.010	0.010	0.010	0.010	0.500
	$\tilde{x}_i^{(\delta)}$	18790	304.966	348.679	3.463	26.603	0.000	89.944	196.880	394.796	6036.711
PFE	$\ln \widetilde{Vol}_i$	18790	11.874	1.255	-0.489	5.539	1.248	11.111	11.902	12.666	17.676
	BAS_i	18790	0.012	0.011	13.388	285.449	0.010	0.010	0.010	0.010	0.470
	$\tilde{x}_i^{(\delta)}$	18765	305.012	360.357	4.347	56.314	0.000	87.733	193.495	390.823	10548.260
T	$\ln \widetilde{Vol}_i$	18765	11.636	1.174	-0.430	5.483	0.000	10.930	11.666	12.379	17.648
	BAS_i	18765	0.012	0.009	10.060	142.817	0.010	0.010	0.010	0.010	0.240
	$\tilde{x}_i^{(\delta)}$	19101	297.310	341.632	3.480	27.493	0.001	85.895	190.414	383.705	6748.658
VZ	$\ln \widetilde{Vol}_i$	19101	11.071	1.192	-0.317	5.122	1.282	10.356	11.093	11.820	17.254
	BAS_i	19101	0.014	0.018	10.593	152.611	0.010	0.010	0.010	0.010	0.460
	$\tilde{x}_i^{(\delta)}$	19399	295.006	351.566	7.275	229.075	0.000	85.828	188.136	382.683	16181.508
WMT	$\ln \widetilde{Vol}_i$	19399	10.656	1.187	-0.409	5.576	0.000	9.933	10.683	11.400	17.006
	BAS_i	19399	0.019	0.032	7.667	78.031	0.010	0.010	0.010	0.020	0.490
	$\tilde{x}_i^{(\delta)}$	19443	301.465	306.120	3.303	30.796	0.161	99.394	210.126	404.082	6133.803
SPY	$\ln \widetilde{Vol}_i$	19443	13.272	0.937	-0.220	3.104	5.738	12.645	13.316	13.935	16.355
	BAS_i	19443	0.011	0.004	32.729	2108.344	0.010	0.010	0.010	0.010	0.320

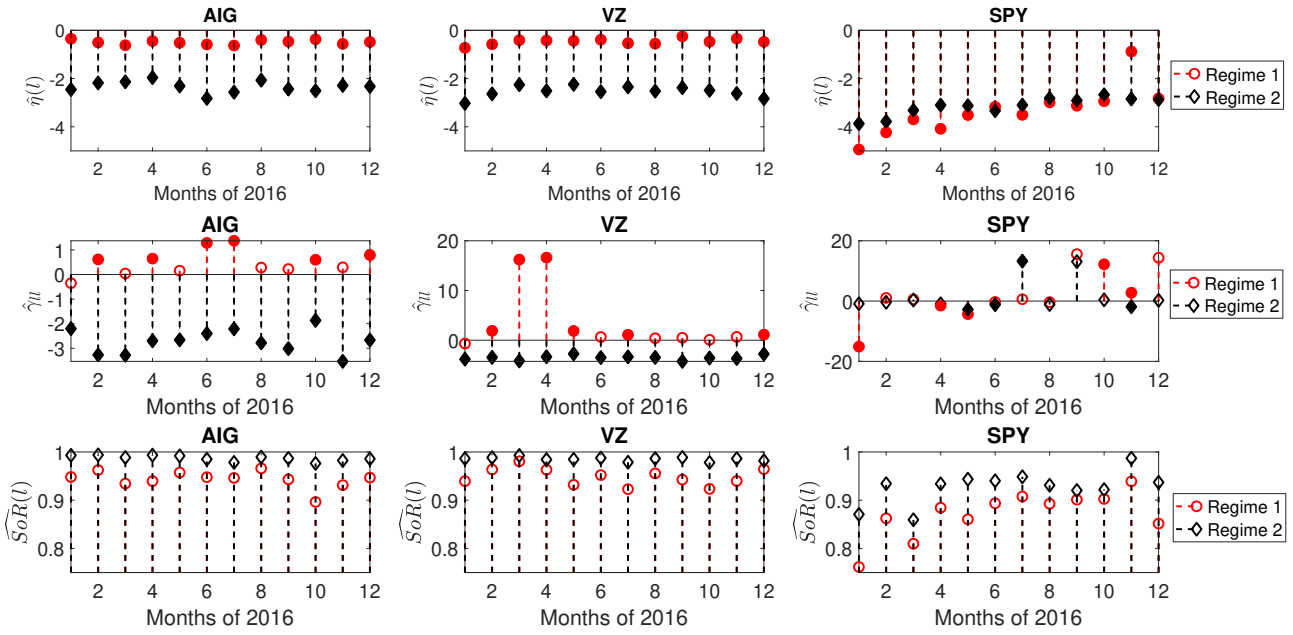
Note: The statistics in the table are computed from $\tilde{x}_i^{(\delta)}$, $\ln \widetilde{Vol}_i$ and BAS_i for the 10 sampled assets constructed from all trading days in 2016. #Obs. is the number of observations in the sample. Std. Dev., Skew. and Kurt. represent sample standard deviation, sample skewness and sample kurtosis, respectively.

Table 4: Estimation outputs for AIG 2016-01, VZ 2016-05 and SPY 2016-03

	AIG 2016-01		VZ 2016-05		SPY 2016-03	
$l =$	1	2	1	2	1	2
$\hat{\alpha}_1(l)$	0.662^{***}	0.231^{***}	0.421^{***}	0.242^{***}	0.204^{***}	0.456^{***}
	(0.077)	(0.018)	(0.057)	(0.019)	(0.049)	(0.017)
$\hat{\beta}_1(l)$	0.548^{***}	0.904^{***}	0.234^{**}	0.939^{***}	0.976^{***}	0.865^{***}
	(0.109)	(0.014)	(0.117)	(0.011)	(0.022)	(0.027)
$\hat{w}(l)$	-0.263	3.650^{***}	2.020^{***}	4.109^{***}	11.343^{***}	7.046^{***}
	(0.804)	(0.117)	(0.600)	(0.136)	(0.444)	(0.306)
$\hat{a}(l)$	0.907^{***}	2.956^{***}	0.738^{***}	2.541^{***}	2.993^{***}	3.523^{***}
	(0.075)	(0.070)	(0.049)	(0.063)	(0.131)	(0.165)
$\hat{\eta}(l)$	-0.350^{***}	-2.454^{***}	-0.435^{***}	-2.240^{***}	-3.693^{***}	-3.319^{***}
	(0.079)	(0.064)	(0.047)	(0.065)	(0.195)	(0.119)
$\hat{c}_{ll}$	-0.463	-5.010^{***}	8.570^{***}	-8.685^{***}	3.064	2.932
	(0.798)	(0.809)	(1.595)	(1.534)	(3.435)	(3.260)
$\hat{\gamma}_{ll}$	-0.353	-2.204^{***}	1.865^{***}	-2.746^{***}	0.575	0.341
	(0.215)	(0.217)	(0.358)	(0.363)	(0.729)	(0.696)
Regime Statistics						
T	1477		1594		1708	
$N(s_i = l)$	90	1387	154	1440	563	1145
$\widehat{SoR}$	0.990		0.980		0.845	
$\widehat{SoR}(l)$	0.948	0.993	0.932	0.985	0.810	0.859
Diagnostic Statistics						
$\mathcal{L}(\hat{\theta}; \mathbb{Y} \vert \hat{\mathbb{S}})$	-8491.725		-9176.952		-9575.877	
$\mathcal{L}(\hat{\vartheta}; \mathbb{Y}, \hat{\mathbb{S}})$	-8613.934		-9371.204		-10470.729	
$\mathrm{E}[\hat{\Lambda}_i \vert \hat{\mathbb{S}}]$	1.012		1.012		0.987	
$\mathrm{Var}[\hat{\Lambda}_i \vert \hat{\mathbb{S}}]$	1.263		1.058		0.954	
AD-stat	3.235^{**}		0.496		1.746	
KS-stat	0.032^{*}		0.020		0.026	
LB(20)	18.026		20.207		29.606^{*}	

Note: Standard errors are in parentheses. ***, ** and * represent significance at 1%, 5% and 10% respectively. $N(s_i = l)$ counts the number of durations in state l. AD-stat and KS-stat are Andersen-Darling and Kolmogorov-Smirnov test statistics for unit exponential distribution constructed on $\hat{\Lambda}_i$. LB(20) is the Ljung-Box test statistics at lag 20. Definitions of $\widehat{SoR}$ and $\widehat{SoR}(l)$ can be found in Eq. (3.20) and Eq. (3.21).

Figure 3: Monthly estimates of $\hat{\eta}(l)$, $\hat{\gamma}_{ll}$ and $\widehat{SoR}(l)$ for AIG, VZ and SPY



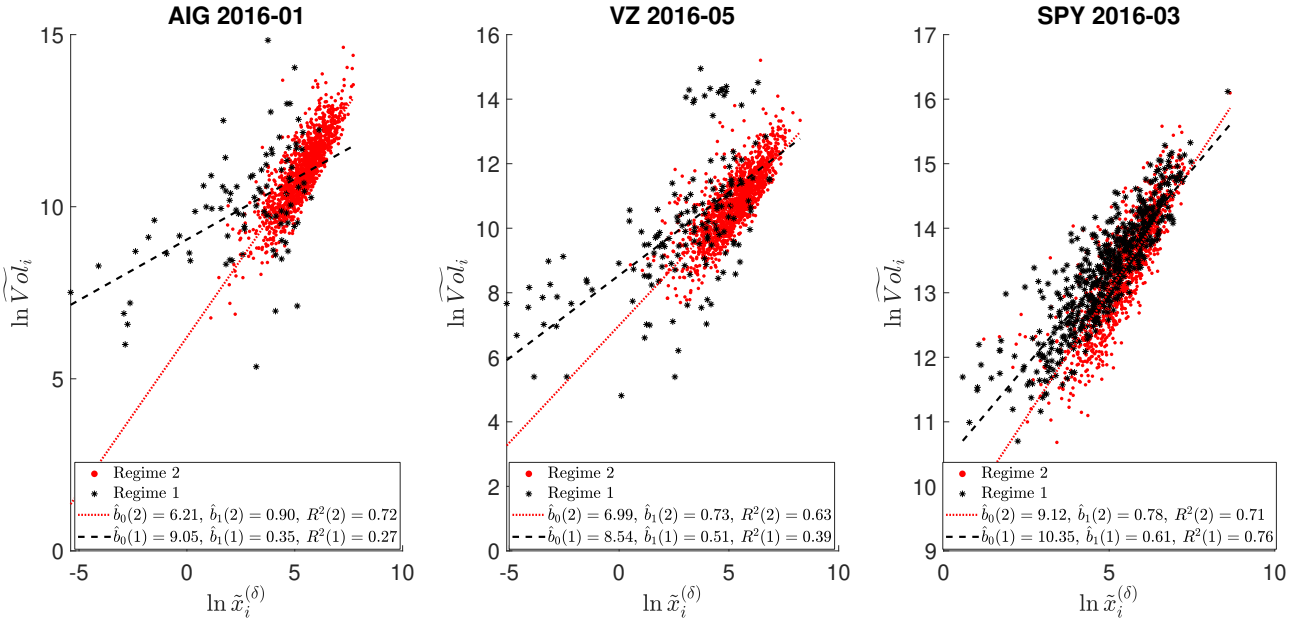
Note: We plot monthly estimated $\hat{\eta}(l)$, $\hat{\gamma}_{ll}$ and $\widehat{SoR}(l)$ for AIG, VZ and SPY in the sampling period. Parameter estimates for regime 1 ($l = 1$) and regime 2 ($l = 2$) are represented by red circles and black diamonds, respectively. For $\hat{\eta}(l)$ and $\hat{\gamma}_{ll}$, filled markers represent parameter estimates that are significant at the 5% level.

capture for the individual stock and SPY. To explain this difference, we plot $\ln \tilde{x}_i^{(\delta)}$ against $\ln \widetilde{Vol}_i$ for two estimated regimes in Figure 4. In this figure, we can clearly see that for AIG and VZ, $\ln \tilde{x}_i^{(\delta)}$ and $\ln \widetilde{Vol}_i$ are highly co-linear with an $R^2(2)$ of $> 60\%$ in regime 2, while this linear relationship is much less obvious in regime 1. However, we can observe similar linear relationships for both regimes 1 and 2 in SPY. This is clear evidence supporting a regime-switching volume-duration relationship in AIG and VZ, but not SPY, which partly explains the lower $\widehat{SoR}$ of SPY.

Our findings in Figure 4 suggest a clear power law relation between $\tilde{x}_i$ and $\widetilde{Vol}_i$ for regime 2. To demonstrate this, we run the regime-specific OLS regression $\ln \widetilde{Vol}_i = b_0(l) + b_1(l) \ln \tilde{x}_i^{(\delta)} + u_i$ for $l = 1, 2$ and provide descriptive statistics of the estimated $\hat{b}_1(l)$ and $R^2(l)$ in Table 5. From the table we can clearly observe that there is a big discrepancy between $R^2(1)$ and $R^2(2)$ for individual stocks in the sense that $\ln \widetilde{Vol}_i$ and $\ln \tilde{x}_i^{(\delta)}$ are highly co-linear in regime 2 with an average $R^2(2)$ of 67%, while $R^2(1)$ is only 31% for regime 1. This result cannot be observed from SPY with similar averages of $R^2(1)$ and $R^2(2)$. The findings in Table 5 is robust to different choices of δ and holds true when we include stocks in the extended sample, which is demonstrated in the supplementary material.

More importantly, the power law relationship between $\tilde{x}_i^{(\delta)}$ and $\widetilde{Vol}_i$ as described in Table 5 resembles the market microstructure invariance hypothesis of Kyle and Obizhaeva (2016), which also suggests a power law between the number of *bets* placed by the investors and trading activity for a given time interval that holds for any asset across time. Following Kyle and Obizhaeva (2016), we propose the following hypothesis:

Figure 4: Scatter plots of $\ln \tilde{x}_i^{(\delta)}$ against $\ln \widetilde{Vol}_i$ for regimes 1 and 2



Note: We plot $\ln \tilde{x}_i^{(\delta)}$ against $\ln \widetilde{Vol}_i$ for regimes 1 and 2, where the regime classification is estimated monthly by the MS(2)-ACI(1,1) model. The dotted and dashed lines are fitted plots for the regression $\ln \widetilde{Vol}_i = b_0(l) + b_1(l) \ln \tilde{x}_i^{(\delta)} + u_i$ for $l = 1, 2$, with estimated parameters and R^2 given in the legend.

Table 5: Descriptive statistics of $\hat{b}_1(l)$ and $R^2(l)$ for regimes 1 and 2

	Mean	Std. Dev.	Q(25%)	Median	Q(75%)
All Individual Stocks					
$\hat{b}_1(1)$	0.4680	0.1185	0.3908	0.4592	0.5411
$\hat{b}_1(2)$	0.7999	0.0476	0.7732	0.8040	0.8311
$R^2(1)$	0.3080	0.1164	0.2325	0.3020	0.3997
$R^2(2)$	0.6706	0.0660	0.6375	0.6788	0.7151
SPY Only					
$\hat{b}_1(1)$	0.5727	0.0779	0.5513	0.5732	0.6111
$\hat{b}_1(2)$	0.7807	0.0318	0.7504	0.7849	0.7983
$R^2(1)$	0.7126	0.1810	0.7308	0.7538	0.8013
$R^2(2)$	0.6813	0.0566	0.6240	0.6932	0.7403

Note: the table presents descriptive statistics of estimated $\hat{b}_1(l)$ and $R^2(l)$ from $\ln \widetilde{Vol}_i = b_0(l) + b_1(l) \ln \tilde{x}_i^{(\delta)} + u_i$ with $l = 1, 2$ for all 120 stock-months in our sample. The regime classification is estimated monthly by the MS(2)-ACI(1,1) model. Results for SPY are excluded in the 'All Individual Stocks' panel and are presented separately in the 'SPY Only' panel. $Q(x\%)$ is the $x\%$ quantile.

Hypothesis 1 (Intraday Volume-Duration Invariance)**.** For a sufficiently large level of price change δ, the (seasonality adjusted) price duration $\tilde{x}_i^{(\delta)}$ and cumulative trading volume generated by bets

satisfies the following relationship for all stocks across time:

$$\widetilde{Vol}_i \propto (\tilde{x}_i^{(\delta)})^{\frac{4}{5}}. \tag{5.2}$$

We conjecture that regime 2 captures price movements triggered by the *bets* placed by the investors, which represents the sharing of risk between different types of market participants that are typically not associated with news arrivals. We note that the results in Kyle and Obizhaeva (2016) do not directly imply Eq. (5.2) without any assumptions on the arrival rate of bets, and the exponent 4/5 is also likely to be dependent on the average sampling interval which is determined by δ. Further research is required to examine the economic connection between Eq. (5.2) and the invariance relationship documented in Kyle and Obizhaeva (2016), which is beyond the scope of this paper.

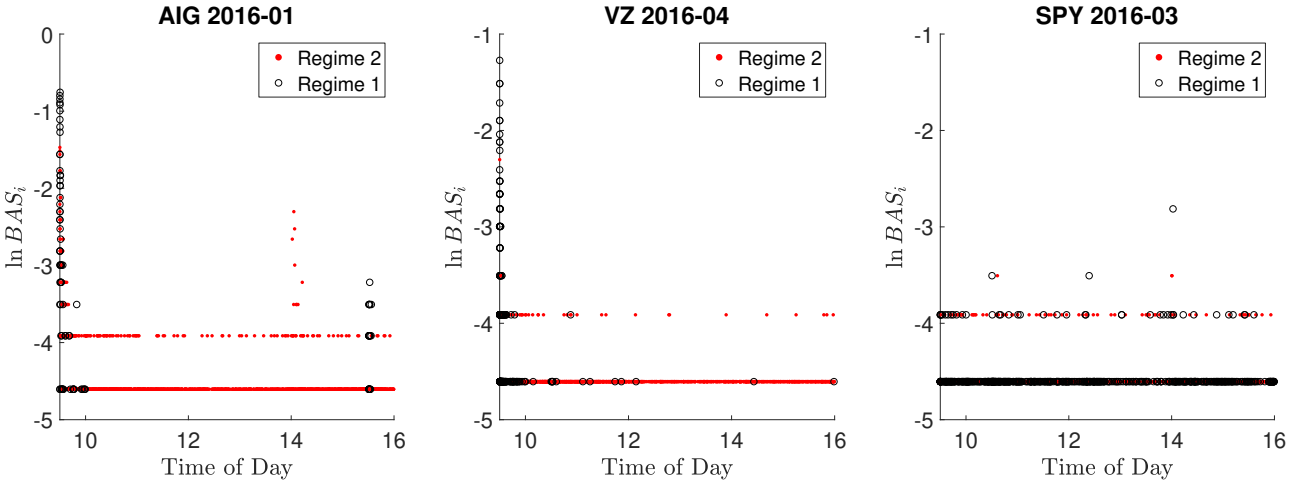
To provide further insight into the regimes we detected, we plot the log bid-ask spread against time of day[4] for the estimated regimes in Figure 5. The figure shows that bid-ask spread to a good extent explains the classification of regime 1 for AIG and VZ. Observations in regime 1 are mostly concentrated at the start of the trading day when the bid-ask spread widens, while observations in regime 2 spread across the whole trading day with a smaller bid-ask spread. As the bid-ask spread of SPY is more evenly distributed throughout the trading day, it appears that both regimes in SPY behave similarly to regime 2 of the individual stocks. The small bid-ask spread of regime 2 observations is also consistent with our previous bet-based argument, as investors planning to submit large bets into the market typically slice their bets into sequences of small orders with the aim to reduce market impact of the bets. Therefore, it is likely for these investors to only submit their orders when the market is sufficiently liquid with a small bid-ask spread.

The cluster of regime 1 observations at the beginning of a trading day can be explained by overnight firm-specific information accumulation. Information-based market microstructure theory (e.g. Glosten and Milgrom (1985), Easley et al. (1996)) predicts that the market makers widen the bid-ask spread in the presence of private information to mitigate their potential loss trading with informed traders. Therefore, trading volume from the informed traders adjusts the price level rapidly when the market opens, which lowers the liquidity of the market, increases the bid-ask spread and causes the volume-duration relationship to deviate from the intraday volume-duration invariance. This also explain why we cannot observe the same effect from SPY, as firm-specific news overnight is unlikely to have an impact on the stock index ETF.

To support our news-based explanation of regime 1, we plot the duration adjusted volume $\widetilde{Vol}_i/(\tilde{x}_i^{(\delta)})^{\frac{4}{5}}$ alongside with the prevailing bid-ask spread against time of day for AIG on Jan 22 and 27, 2016 in Figure 6. On both days there are very few regime 1 observations at the beginning of the trading day. On 22-Jan 2016, we see a cluster of regime 1 observations at around 15:30 in the

[4]Note that the x-axis covers the regular trading session of a typical U.S. security exchange on a trading day, which is from 9:30 to 16:00 Eastern Standard Time.

Figure 5: Scatter plots of BAS_i against time of day for regimes 1 and 2



Note: We plot $\ln BAS_i$ against $t_i^{(\delta)}$ for regimes 1 and 2, where the regime classification is estimated monthly by the MS(2)-ACI(1,1) model. Note that the regime of $\ln BAS_i$ is determined by s_{i+1}.

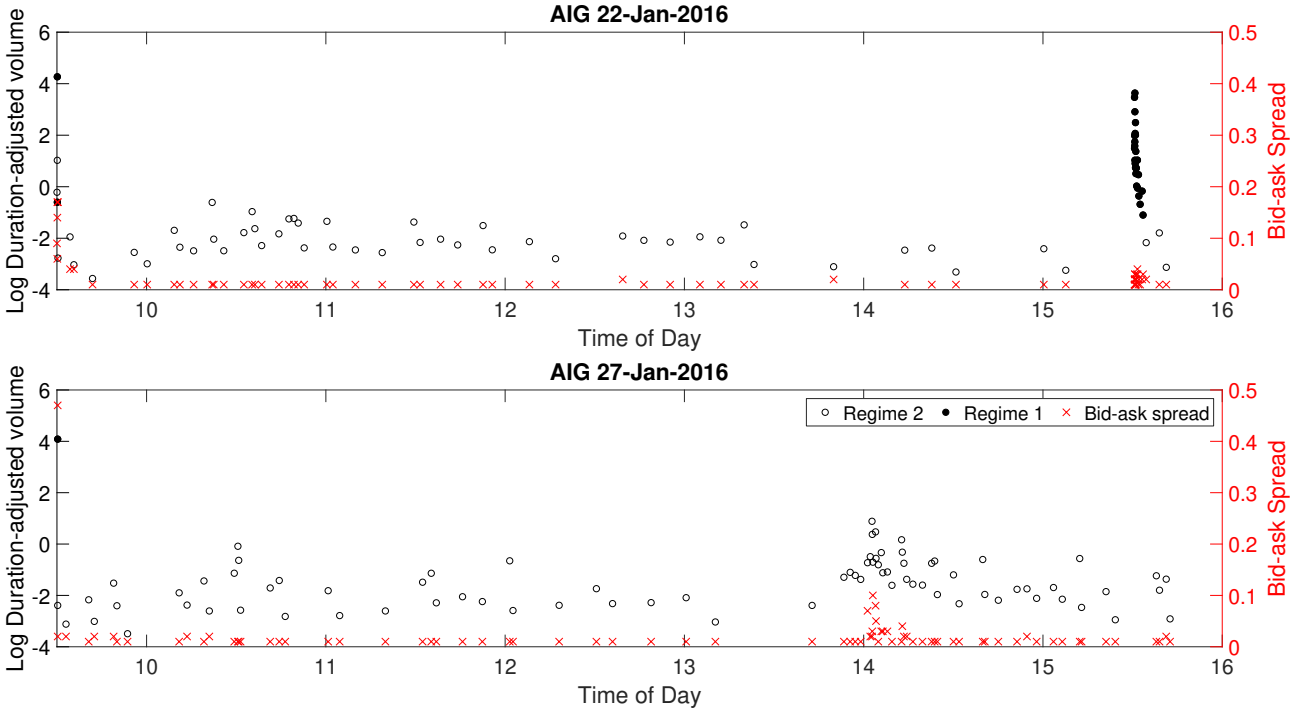
afternoon, accompanied by an increase in the bid-ask spread. This is likely to be triggered by AIG's announcement of a spin-off of mortgage unit. We track this event using the RavenPack[5] dataset which collects all publicly available firm-specific news. The first entry of this event is recorded at 15:32:47 on Jan 22, 2016 with RavenPack unique story ID "CB6E9267993A172986E45E07EBFDC723". This event triggers fast price adjustments with large duration-adjusted volume, which is identified by our model as regime 1.

By contrast, the lower panel of Figure 6 also presents a cluster of widened bid-ask spread right after 14:00, and the duration-adjusted volume is also slightly elevated. This is possibly caused by the Federal Open Market Committee (FOMC) meetings scheduled at 14:00 on 27-Jan 2016. As this market-wide event does not have enough firm-specific information content about AIG, our model does not detect a regime shift in the duration-adjusted volume. Therefore, observations right after 14:00 are not classified into regime 1, despite a significantly widened bid-ask spread. This is direct evidence supporting our argument that regime 1 is mainly capturing market impact of firm-specific news arrivals.

The behaviour of regime 1 observations has a close connection to the findings in Jiang et al. (2010) and Christensen et al. (2014). In these papers, they describe the phenomenon that a burst of volatility in the market is often accompanied by a shock to liquidity, which is often identified as jumps with less frequent sampling (e.g. once every 5 minutes). Plotting the price path and bid-ask spread of AIG around 15:30 on 22-Jan-2016 in Figure 7, we see that a larger cluster of price events is observed after 15:30 with a significantly widened bid-ask spread, which is triggered by a significant increase in price volatility. Importantly, the cluster of price events is not a result of a single jump which only triggers

[5] https://www.ravenpack.com/

Figure 6: Example of regime classification for AIG on 22 and 27, Jan 2016



Note: The duration-adjusted volume is defined as $\widetilde{Vol}_i/(\tilde{x}_i^{(\delta)})^{\frac{4}{5}}$. The the regime classification is estimated monthly by the MS(2)-ACI(1,1) model. For each duration-adjusted volume plotted on the right y-axis, we plot the corresponding prevailing bid-ask spread BAS_{i-1} on the left axis.

one price event. Therefore, our findings can provide a method to detect bursts of volatility proposed by Christensen et al. (2014) in an ultra high-frequency context.

As an application to the two regimes detected by our MS-ACI model, we propose a news-based decomposition of daily volatility estimates for individual stocks. Based on Tse and Yang (2012) and Li et al. (2019), daily volatility estimates can be directly derived from $\hat{\Lambda}_i$. Let $\{\hat{\Lambda}_i\}_{i=1:T_d}$ denote the fitted integrated intensity over the i-th price duration from the MS-ACI model for day d, the Integrated Conditional Variance (ICV) estimate of day d is constructed as:

$$ICV_d = \delta_d^2 \sum_{i=1}^{T_d} F^{-1}(1 - e^{-\hat{\Lambda}_i}), \tag{5.3}$$

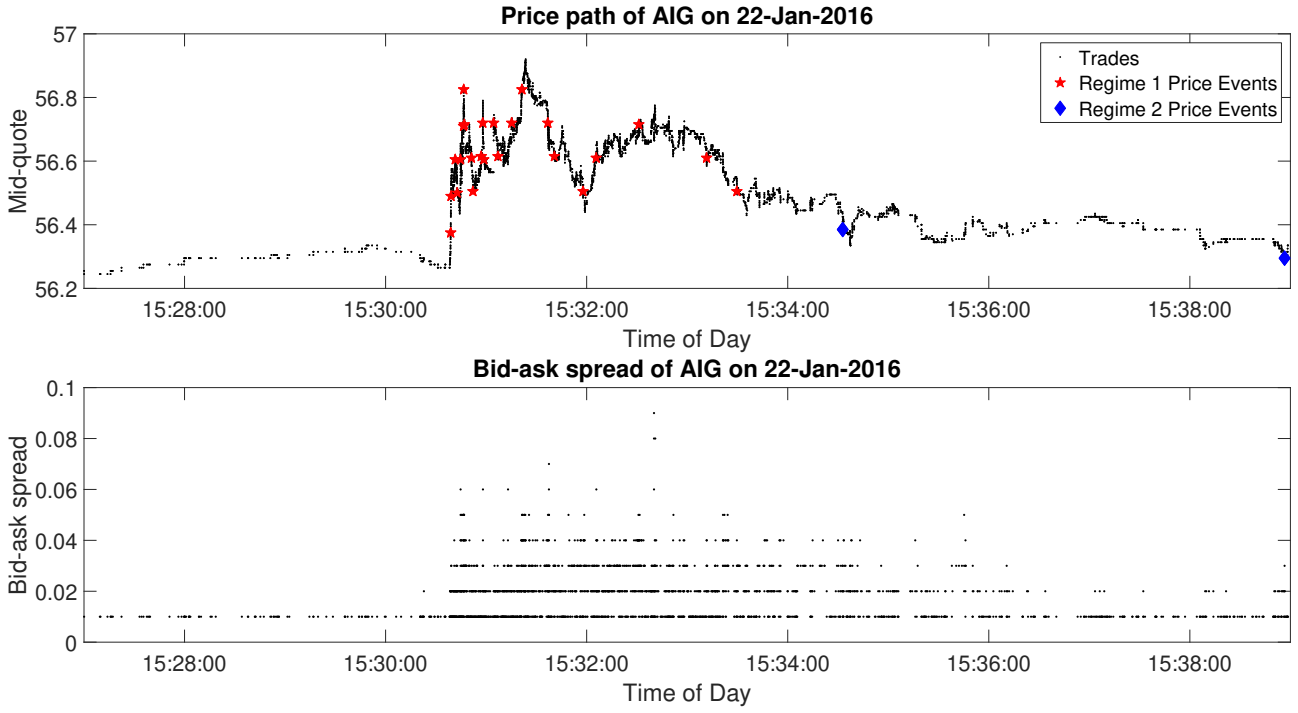
where δ_d is the daily choice of price threshold and $F^{-1}(x)$ is the inverse function of $F(x) = 2 - 2\sum_{k=-\infty}^{\infty} \text{erf}\left(\frac{1+4k}{\sqrt{2x}}\right)$, which is the cumulative density function of the first exit time of a standard Brownian motion to exit the $[-1, 1]$ bound. As each duration is classified into one regime based on the estimated states $\hat{s}_i$, we propose the following decomposition of total daily ICV:

$$ICV_d = ICV_d(1) + ICV_d(2), \tag{5.4}$$

where $ICV_d(l) = \delta_d^2 \sum_{i=1}^{T_d} F^{-1}(1 - e^{-\hat{\Lambda}_i}) \mathbb{1}_{\{\hat{s}_i = l\}}$ is the volatility contribution from regime l in day d.

We plot the ICV decomposition for AIG and VZ in Figure 8. The figure shows that, for most of the trading days, the total ICV is dominated by $ICV_d(2)$ that is not related to news arrivals. However,

Figure 7: AIG price and bid-ask spread plot on 22-Jan-2016



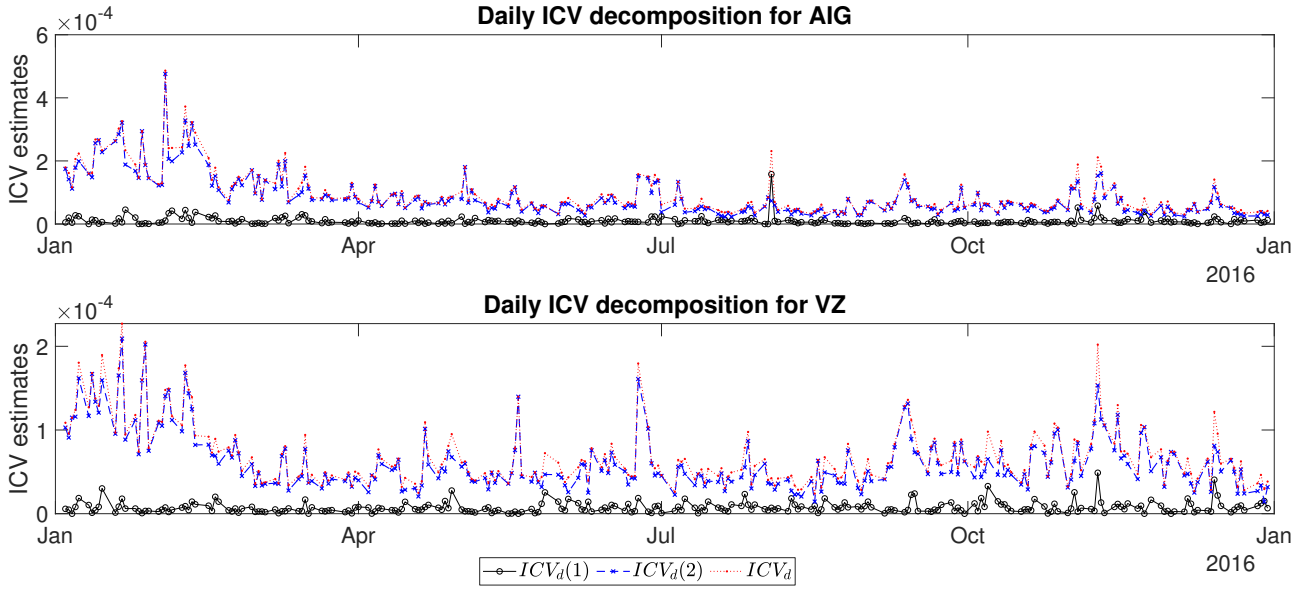
Note: each black dot is associated with a trade plotted in chronological order. The corresponding mid-quote and bid-ask spread at every trade is plotted in the upper and lower panel, respectively. The price events are indicated by red pentagrams and purple diamonds, with a corresponding $\delta = 0.00178$.

in some cases $ICV_d(1)$ can be even higher than $ICV_d(2)$, which can be an indicator of important firm-specific events occurring either during the overnight period or within the trading hours that act as shocks to the liquidity of the market.

The component $ICV_d(1)$ has drastically different statistical properties in comparison to $ICV_d(2)$. In Table 6 we present some descriptive statistics for the ICV decomposition for all 9 individual stocks in our sample. The table shows that, firstly, $ICV_d(1)$ is on average only a small (about 10%) percentage of total ICV_d. The autocorrelation of $ICV_d(1)$ is much weaker than that of $ICV_d(2)$, which implies that forecasting $ICV_d(1)$ is much more difficult. Finally, $ICV_d(2)$ is highly correlated with ICV_d and ICV_d^{SPY}, while for $ICV_d(1)$ the correlations are much weaker.

Concluding from above, we argue that the volatility components $ICV_d(1)$ and $ICV_d(2)$ measure different aspect of risks for the individual stocks. $ICV_d(2)$ mainly captures market-wide risk factors which are highly-persistent and predictable, whereas $ICV_d(1)$ is largely influenced by liquidity shocks possibly caused by firm-specific news arrivals that are more likely to be transitory and idiosyncratic. This decomposition provides unique insights in understanding the driving forces of price volatility, and can be applied by practitioners to evaluate their exposure to different sources of risk.

Figure 8: Daily ICV decomposition for AIG and VZ



Note: The ICV decomposition is defined in Eq. (5.4) based on monthly-estimated MS(2)-ACI(1,1) model.

Table 6: Descriptive statistics of the ICV components for 9 individual stocks

	Regime	Mean	ACF(1)	ACF(5)	ACF(22)	$\rho(ICV_d)$	$\rho(ICV_d^{SPY})$
AIG	$ICV_d(1)$	0.9579	0.1049	0.0077	0.0247	0.3852***	0.1286**
	$ICV_d(2)$	8.6119	0.7317***	0.6123***	0.3595***	0.9824***	0.8974***
CVX	$ICV_d(1)$	1.1635	0.4468***	0.2508***	0.1649***	0.6623***	0.5588***
	$ICV_d(2)$	15.0555	0.9005***	0.8382***	0.3151***	0.9975***	0.8938***
GM	$ICV_d(1)$	1.758	0.0722	0.1218*	0.0156	0.5195***	0.4275***
	$ICV_d(2)$	15.4703	0.7671***	0.6521***	0.2902***	0.9898***	0.8679***
INTC	$ICV_d(1)$	0.8732	0.0786	0.0366	-0.0215	0.4246***	0.2047***
	$ICV_d(2)$	10.2234	0.6091***	0.5089***	0.3409***	0.9904***	0.8496***
JPM	$ICV_d(1)$	0.7604	0.1892***	0.2121***	-0.0768	0.3091***	0.2406***
	$ICV_d(2)$	11.3937	0.7249***	0.5643***	0.3027***	0.9965***	0.8832***
PFE	$ICV_d(1)$	0.7361	0.0276	0.0428	0.0019	0.2993***	0.0937
	$ICV_d(2)$	9.2922	0.553***	0.3211***	0.0936	0.995***	0.6537***
T	$ICV_d(1)$	0.7042	0.1365**	0.1072*	0.1107*	0.5104***	0.1525**
	$ICV_d(2)$	6.2082	0.5512***	0.3263***	0.1083*	0.9909***	0.5581***
VZ	$ICV_d(1)$	0.7313	0.1309**	0.0801	0.0025	0.3235***	0.1026*
	$ICV_d(2)$	6.2059	0.6442***	0.5297***	0.1064*	0.9832***	0.7950***
WMT	$ICV_d(1)$	0.9968	0.2179***	0.1817***	0.0419	0.5081***	0.2630***
	$ICV_d(2)$	6.8686	0.6659***	0.5146***	0.2383***	0.9889***	0.7301***

Note: ***, ** and * represent significance at 1%, 5% and 10% respectively. The mean of the ICV estimates are scaled by 10^5. ACF(x) represents the x-th lag autocorrelation. $\rho(ICV_d)$ (resp. $\rho(ICV_d^{SPY})$) is the contemporaneous correlation between $ICV_d(l)$ and ICV_d (resp. ICV_d^{SPY}, the ICV estimates for SPY) for the corresponding regime.

6 Concluding Remarks

This paper develops the Markov-Switching ACI (MS-ACI) model by extending the Markov-switching structure to the original ACI model (Russell, 1999). The stationarity and moment conditions for the ACI model are derived, which augments the stationarity condition provided by Hautsch (2012). To overcome the path dependency problem in the estimation of Markov-switching autoregressive models, we propose to estimate the model via the SAEM algorithm combined with Bauwens et al.'s (2010) single move sampler. By introducing the concept of Significance of Regimes (SoR), we demonstrate in our simulation that our SAEM estimation scheme is capable of providing reliable parameter estimates for the MS-ACI model. Our simulation analysis subsequently shows that the SoR statistic can serve as a diagnostic tool in assessing the performance of a regime-switching structure, which has a crucial impact on the accuracy of parameter estimates for the MS-ACI model.

In our empirical study, we discover two distinct volume-duration regimes with high SoR for 9 individual stocks, whereas the differences between the two identified regimes are much less pronounced for SPY. These findings are further corroborated by the empirical results on 24 additional stocks presented in the supplementary material. We connect the dominant regime with the risk transfer between market participants and the minor regime with firm-specific information arrivals, and decompose daily volatility into two components that correspond to the two detected regimes. We demonstrate that the two components have very different statistical properties, which are likely to carry different sources of risks within a trading day.

Our study has several limitations that provide ample room for future research. First, our empirical study is based on a relatively small sample due to the computationally intensive estimation of the model. It is therefore important to verify our hypothesis of volume-duration invariance on a more comprehensive dataset. Second, the specification of the MS-ACI model in our empirical analysis can be further generalized to study additional determinants of the regimes, e.g. asymmetry in the regime identification conditioning on the sign of the past return. Lastly, the regime-based volatility decomposition can be useful in various empirical applications, such as volatility forecasting and factor analysis, and is worth further empirical investigations.

Acknowledgement

We are grateful to two anonymous referees and the editor for their invaluable comments, which greatly improved the quality of the paper. We would like to acknowledge Sebastian Bayer, Winfried Pohlmeier and the Department of Economics, University of Konstanz for providing us with computational resources for the simulation. We are grateful to all participants at the 2nd and 3rd Konstanz-Lancaster Workshop for Financial Econometrics, the 1st Lancaster-Warwick Finance Workshop, the Financial Econometrics and Empirical Asset Pricing Conference at Lancaster University, the Vienna-Copenhagen

Conference on Financial Econometrics and the 4th IAAE annual conference for their helpful discussions. We would like to thank Torben Andersen, Luc Bauwens, Asger Lunde, Peter Hansen, Nikolaus Hautsch and Olan Henry for their insightful suggestions and comments. Support from the ESRC-FWF grant 'Bilateral Austria: Order Book Foundations of Price Risks and Liquidity: An Integrated Equity and Derivatives Markets Perspective' (ES/N014588/1) is gratefully acknowledged. The usual disclaimer applies.

Supplementary Material

The supplementary material of this paper contains further empirical results and robustness checks, which is available from the online version of this article at the publisher's website.

References

Allassonnière, S., Kuhn, E., and Trouvé, A. (2010). Construction of Bayesian Deformable Models via Stochastic Approximation Algorithm: A Convergence Study. *Bernouilli*, 16(3):641–678.

Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2-3):115–158.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association*, 96(453):42–55.

Augustyniak, M. (2014). Maximum likelihood estimation of the Markov-switching GARCH model. *Computational Statistics & Data Analysis*, 76:61–75.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *Econometrics Journal*, 12(3):C1–C32.

Bauwens, L., Dufays, A., and Rombouts, J. V. K. (2014). Marginal likelihood for Markov-switching and change-point GARCH models. *Journal of Econometrics*, 178:508–522.

Bauwens, L., Preminger, A., and Rombouts, J. V. K. (2010). Theory and inference for a Markov switching GARCH model. *Econometrics Journal*, 13(2):218–244.

Billio, M., Casarin, R., and Osuntuyi, A. (2014). Efficient Gibbs sampling for Markov switching GARCH models. *Computational Statistics & Data Analysis*, 100:37–57.

Bowsher, C. G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141:876–912.

Celeux, G. and Diebolt, J. (1992). A stochastic approximation type EM algorithm for the mixture problem. *Stochastics and Stochastic Reports*, 41(1-2):119–134.

Christensen, K., Oomen, R. C., and Podolskij, M. (2014). Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics*, 114(3):576–599.

Daley, D. J. and Vere-Jones, D. (2003). *An introduction to the theory of point processes.* Springer Science & Business Media, Berlin, Germany.

De Luca, G. and Zuccolotto, P. (2006). Regime-switching Pareto distributions for ACD models. *Computational Statistics & Data Analysis*, 51(4):2179–2191.

Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a Stochastic Approximation Version of the EM Algorithm. *The Annals of Statistics*, 27(1):94–128.

Dueker, M. (1997). Markov Switching in GARCH Processes and Mean- Reverting Stock-Market Volatility. *Journal of Business & Economics Statistics*, 15(1):26–34.

Easley, D., Kiefer, N., O'Hara, M., and Paperman, J. (1996). Liquidity, information, and infrequently traded stocks. *Journal of Finance*, 51:1405–1436.

Engle, R. F. and Russell, J. R. (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica*, 66:1127–1162.

Filardo, A. J. (1994). Business-cycie phases and their transitional dynamics. *Journal of Business and Economic Statistics*, 76(2):271–292.

Francq, C. and Zakoïan, J.-M. (2001). Stationarity of multivariate Markov-switching ARMA models. *Journal of Econometrics*, 102:339–364.

Gallo, G. M. and Otranto, E. (2012). The Markov switching asymmetric multiplicative error model. *CRENoS Working Paper*.

Gerhard, F. and Hautsch, N. (2002). Volatility estimation on the basis of price intensities. *Journal of Empirical Finance*, 9:57–89.

Glosten, L. R. and Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14:71–100.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1):27–62.

Haas, M., Mittnik, S., and Paolella, M. S. (2004). A New Approach to Markov-Switching GARCH Models. *Journal of Financial Econometrics*, 2(4):493–530.

Hautsch, N. (2012). *Econometrics of Financial High-Frequency Data.* Springer Berlin Heidelberg, Berlin, Heidelberg.

Holden, C. W. and Jacobsen, S. (2014). Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. *Journal of Finance*, 69(4):1747–1785.

Hong, S. Y., Nolte, I., Taylor, S., and Zhao, X. (2020). Volatility Estimation and Forecasts Based on Price Durations. *Working Paper, Lancaster University Management School*.

Hujer, R., Vuletic, S., and Kokot, S. (2002). The Markov Switching ACD Model. *Working Paper Series: Finance and Accounting, Johann Wolfgang Goethe-Universitat Frankfurt a.A.*, No.90.

Jank, W. (2006). Implementing and Diagnosing the Stochastic Approximation EM Algorithm. *Journal of Computational and Graphical Statistics*, 15(4):803–829.

Jiang, G. J., Lo, I., and Verdelhan, A. (2010). Information shocks, liquidity shocks, jumps, and price discovery: Evidence from the U.S. Treasury market. *Journal of Financial and Quantitative Analysis*, 46:527–551.

Karr, A. (1991). *Point Processes and Their Statistical Inferences*. CRC press.

Kim, C.-J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60(1-2):1–22.

Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics*, 27(2):363–394.

Kuhn, E. and Lavielle, M. (2004). Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM: Probability and Statistics*, 8(August):115–131.

Kyle, A. and Obizhaeva, A. (2016). Market Microstructure Invariance: Empirical Hypotheses. *Econometrica*, 84(4):1345–1404.

Lee, C. M. C. and Ready, M. J. (1991). Inferring Trade Direction from Intraday Data. *The Journal of Finance*, 46:733–746.

Li, Y., Nolte, I., and Nolte, S. (2019). Renewal Based Volatility Estimation. *Lancaster University Management School Working Paper*.

Russell, J. R. (1999). Econometric Modeling of Multivariate Irregularly-Spaced High-Frequency Data. *Working Paper, University of Chicago*.

Stelzer, R. (2009). On Markov-Switching ARMA Processes-Stationarity, Existence of Moments, and Geometric Ergodicity. *Econometric Theory*, 25(1):43–62.

Tse, Y.-k. and Yang, T. T. (2012). Estimation of High-Frequency Volatility: An Autoregressive Conditional Duration Approach. *Journal of Business & Economic Statistics*, 30(4):533–545.

Wei, G. C. and Tanner, M. A. (1990). A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms. *Journal of the American Statistical Association*, 85(411):699–704.

A Proof to Theorem 2.1

Since we do not include any time-varying covariates, we can consider $\lambda(t|\mathcal{F}_t)$ as a time-deformed model evaluated at every t_i. For simplicity, let us define $\lambda_i \equiv \lambda(t_i|\mathcal{F}_{t_i})$, we can write:

$$\lambda_i = e^{wa+\tilde{\Phi}_i} \cdot a x_i^{a-1}. \tag{A.1}$$

The integrated intensity between two events is available in closed form:

$$\Lambda_i \equiv \int_{t_{i-1}}^{t_i} \lambda(s|\mathcal{F}_s)ds = e^{wa+\tilde{\Phi}_i} \cdot x_i^a = \frac{x_i \lambda_i}{a} \sim i.i.d.\ exp(1). \tag{A.2}$$

We now use the MA(∞) representation of $\tilde{\Phi}_i$:

$$\tilde{\Phi}_i = \psi(L)\tilde{\varepsilon}_{i-1} = -\gamma \sum_{j=1}^{\infty} \psi_j - \sum_{j=1}^{\infty} \ln \Lambda_{i-j}^{\psi_j}. \tag{A.3}$$

Substituting the above into Eq. (A.1) and using Eq. (A.2), we can write λ_i in a compact form:

$$\lambda_i = C \Lambda_i^{1-\frac{1}{a}} \prod_{j=1}^{\infty} \Lambda_{i-j}^{-\frac{\psi_j}{a}}, \tag{A.4}$$

where $C = a e^{w-\gamma \sum_{j=1}^{\infty} \psi_j}$. From Eq. (A.2) we can also derive:

$$x_i = \frac{a}{C} \Lambda_i^{\frac{1}{a}} \prod_{j=1}^{\infty} \Lambda_{i-j}^{\frac{\psi_j}{a}}. \tag{A.5}$$

From the i.i.d.-ness of Λ_i, the k-th moments of λ_i and x_i can be written as products of moments of Weibull random variables:

$$\begin{aligned} \mathrm{E}[\lambda_i^k] &= C^k \, \mathrm{E}[\Lambda_i^{k-\frac{k}{a}}] \prod_{j=1}^{\infty} \mathrm{E}[\Lambda_{i-j}^{-\frac{k\psi_j}{a}}], \\ \mathrm{E}[x_i^k] &= \frac{a^k}{C^k} \, \mathrm{E}[\Lambda_i^{\frac{k}{a}}] \prod_{j=1}^{\infty} \mathrm{E}[\Lambda_{i-j}^{\frac{k\psi_j}{a}}]. \end{aligned} \tag{A.6}$$

It is clear that $\mathrm{E}[\lambda_i^k]$ are $\mathrm{E}[x_i^k]$ are finite if (1) $C \in (0, \infty)$; (2) the infinite products in Eq. (A.6) are finite. We now show that conditions 1 and 2 of Theorem 2.1 imply the two conditions above for $k = 2$.

Condition 1 of Theorem 2.1 implies absolute summability of ψ_i which further implies that $C \in (0, \infty)$. For the second argument, we use the following lemma:

Lemma A.1. *Let $\Lambda_i \sim exp(1)$, then $\mathrm{E}[\Lambda_i^k] < \infty$ iff $k > -1$. When $k > -1$, $\mathrm{E}[\Lambda_i^k] = \Gamma(k+1)$, where $\Gamma(x)$ is the Gamma function.*

Proof. Notice that by definition of the Gamma function $\mathrm{E}[\Lambda_i^k] = \int_0^\infty x^k e^{-x} dx = \Gamma(k+1)$. The Euler integral converges iff $k > -1$ as a property of the Gamma function. This completes the proof. □

To ensure that all expectations in Eq. (A.6) are finite, the following constraints for the parameters of the ACI model can be derived from Eq. (A.6) using Lemma A.1:

$$a > \frac{k}{k+1}, \quad \sup|\psi_i| < \frac{a}{k}. \tag{A.7}$$

Under this condition, we can write:

$$\begin{aligned} \mathrm{E}[\lambda_i^k] &= C^k \Gamma(k - k/a + 1) \prod_{j=1}^{\infty} \Gamma(1 - k\psi_j/a), \\ \mathrm{E}[x_i^k] &= \frac{a^k}{C^k} \Gamma(k/a + 1) \prod_{j=1}^{\infty} \Gamma(1 + k\psi_j/a). \end{aligned} \tag{A.8}$$

It is therefore clear that condition 2 implies that all expectations in Eq. (A.6) are finite. We are left to show that the infinite products in Eq. (A.8) are finite under conditions 1 and 2. Convergence of the infinite products in Eq. (A.8) is equivalent to convergence of the sum $\sum_{j=1}^{\infty} \ln \Gamma(1 \pm k\psi_j/a)$. We now show that absolute summability of ψ_j implies absolute summability of $\ln \Gamma(1 \pm k\psi_j/a)$. Condition 1 implies that $\lim_{j\to\infty} |\psi_j| \to 0$. We note that the following result holds by first order Taylor expansion at $\psi_j = 0$:

$$\lim_{j\to\infty} \frac{|\ln \Gamma(1 \pm k\psi_j/a)|}{|\psi_j|} \to k\gamma/a > 0, \tag{A.9}$$

where γ is the Euler-Mascheroni constant. By limit comparison test, the absolute summability of ψ_j and Eq. (A.9) imply the absolute summability of $\ln \Gamma(1 \pm k\psi_j/a)$, and the proof is complete. □

B Analytical Gradient of the MS-ACI Model

We provide analytical gradient of the MS(M)-ACI(1,1) model with Weibull baseline conditioning on the state vector which we denote by $\frac{\partial \ln \mathcal{L}(\theta;\mathbb{Y}|\mathbb{S})}{\partial \theta}$, where

$$\theta = \{\{\alpha_1(m)\}_{m\in\mathcal{M}}, \{\beta_1(m)\}_{m\in\mathcal{M}}, \{w(m)\}_{m\in\mathcal{M}}, \{a(m)\}_{m\in\mathcal{M}}, \{\eta(m)\}_{m\in\mathcal{M}}\}'$$

is a $5M \times 1$ parameter vector. Note that in the above parameter vector, we assume that $\boldsymbol{Z}(t) = Z_{\breve{N}(t)+1}$, where $\{Z_i\}_{i=1:T}$ is a univariate time-invariant covariate, and $\eta(m)$ is the corresponding state-dependent coefficient. Using the notation $\ln \mathcal{L}(\theta; x_i|\mathbb{S}) = -\ln \Lambda_i + \lambda_i$ and according to Eq. (A.2), we can write:

$$\frac{\partial \ln \mathcal{L}(\theta; x_i|\mathbb{S})}{\partial \theta} = \frac{\partial \ln \lambda_i}{\partial \theta}\Big(1 - \frac{x_i}{a(s_i)}\lambda_i\Big) - \frac{\partial \frac{x_i}{a(s_i)}}{\partial \theta}\lambda_i. \tag{B.1}$$

Therefore the analytical gradient can be computed based on the knowledge of $\frac{\partial \ln \lambda_i}{\partial \theta}$. Note that $\ln \lambda_i$ has an MS-ARMA-type representation by taking log on both sides of a regime switching version of

Eq. (A.1). Consequently, we can derive that:

$$
\begin{aligned}
\frac{\partial \ln \lambda_i}{\alpha_1(m)} &= \mathbb{1}_{\{s_i=m\}}(-\gamma - \ln x_{i-1} + a(s_{i-1}) - \ln \lambda_{i-1}) + (\beta_1(s_i) - \alpha_1(s_i))\frac{\partial \ln \lambda_{i-1}}{\alpha_1(m)}, \\
\frac{\partial \ln \lambda_i}{\beta_1(m)} &= \mathbb{1}_{\{s_i=m\}}(\ln \lambda_{i-1} - w(s_{i-1})a(s_{i-1}) - \eta(s_{i-1})Z_{i-1} - \ln a(s_{i-1}) - (a(s_{i-1}) - 1)\ln x_{i-1}) \\
&\quad + (\beta_1(s_i) - \alpha_1(s_i))\frac{\partial \ln \lambda_{i-1}}{\beta_1(m)}, \\
\frac{\partial \ln \lambda_i}{w(m)} &= \mathbb{1}_{\{s_i=m\}}a(s_i) - \mathbb{1}_{\{s_{i-1}=m\}}\beta_i(s_i)a(s_{i-1}) + (\beta_1(s_i) - \alpha_1(s_i))\frac{\partial \ln \lambda_{i-1}}{w(m)}, \\
\frac{\partial \ln \lambda_i}{a(m)} &= \mathbb{1}_{\{s_i=m\}}(w(s_i) + a(s_i)^{-1} + \ln x_i) + (\beta_1(s_i) - \alpha_1(s_i))\frac{\partial \ln \lambda_{i-1}}{a(m)} \\
&\quad + \mathbb{1}_{\{s_{i-1}=m\}}(-\beta_1(s_i)w(s_{i-1}) + (\beta_1(s_i) - \alpha_1(s_i))a(s_{i-1})^{-1} - \beta_1(s_i)\ln x_{i-1}), \\
\frac{\partial \ln \lambda_i}{\eta(m)} &= \mathbb{1}_{\{s_i=m\}}Z_i - \mathbb{1}_{\{s_{i-1}=m\}}\beta_1(s_i)Z_{i-1} + (\beta_1(s_i) - \alpha_1(s_i))\frac{\partial \ln \lambda_{i-1}}{\eta(m)}.
\end{aligned}
\tag{B.2}
$$

For a given θ, we can therefore calculate $\frac{\partial \ln \lambda_i}{\partial \theta}$ recursively using the above set of equations. Finally, note that:

$$
\frac{\partial \frac{x_i}{a(s_i)}}{\partial \theta} = \frac{x_i}{a(s_i)^2}\frac{\partial \theta}{\partial a(s_i)}. \tag{B.3}
$$

The analytical gradient can therefore be calculated by substituting Eq. (B.2) and Eq. (B.3) into Eq. (B.1).

C Estimating the Variance-Covariance Matrix of Parameters and the Most Probable State Vector

As suggested by Delyon et al. (1999) and Kuhn and Lavielle (2004), the variance-covariance matrix for parameter estimates can be estimated directly from the iterations of the SAEM algorithm. We can use the following stochastic approximation scheme to obtain an estimate of the observed information matrix:

$$
\begin{aligned}
S_n &= (1-\gamma_n)S_{n-1} + \frac{\gamma_n}{K_n}\sum_{k=1}^{K_n} s(\vartheta^{(n+1)}; \mathbb{Y}, \mathbb{S}^{(n,k)}), \\
D_n &= (1-\gamma_n)D_{n-1} \\
&\quad + \frac{\gamma_n}{K_n}\sum_{k=1}^{K_n}(\mathcal{I}(\vartheta^{(n+1)}; \mathbb{Y}, \mathbb{S}^{(n,k)}) + s(\vartheta^{(n+1)}; \mathbb{Y}, \mathbb{S}^{(n,k)})s(\vartheta^{(n+1)}; \mathbb{Y}, \mathbb{S}^{(n,k)})'), \\
H_n &= D_n - S_n S_n',
\end{aligned}
$$

where $s(\vartheta^{(n+1)}; \mathbb{Y}, \mathbb{S}^{(n,k)})$ and $\mathcal{I}(\vartheta^{(n+1)}; \mathbb{Y}, \mathbb{S}^{(n,k)})$ are the score vector and the Fisher's information matrix of the complete likelihood function $\mathcal{L}(\vartheta^{(n+1)}; \mathbb{Y}, \mathbb{S}^{(n,k)})$. Delyon et al. (1999) show that $-H_n$ converges to $-\mathcal{I}(\vartheta, \mathbb{Y})$ when ϑ converges to the limiting value and the function $\ln \mathcal{L}(\vartheta; \mathbb{Y})$ is smooth enough. Thus, the inverse of $-H_n$ obtained will be used as an estimate of the variance-covariance matrix of ϑ, which is evaluated at the termination of the SAEM algorithm.

As to the estimation of the most probable state vector, we rely on the posterior density of $\mathbb{S}$ given the data and the estimated parameter vector $\hat{\vartheta}$. Due to the dimensionality of $\mathbb{S}$, it is difficult to directly maximize the likelihood of $\mathbb{S}$ conditioning on $\hat{\vartheta}$ and $\mathbb{Y}$, and standard algorithms (e.g. Viterbi's algorithm) for the Hidden Markov Model are not applicable in this case because of the path dependency issue. Bauwens et al. (2010) use the smoothed posterior probabilities for each state to obtain an estimate of this state sequence. We hereby propose a direct Monte Carlo search method to obtain $\hat{\mathbb{S}}$ by exploiting the following relationship:

$$p(\mathbb{S}|\hat{\vartheta}, \mathbb{Y}) \propto f(\mathbb{Y}|\mathbb{S}, \hat{\vartheta})p(\mathbb{S}) = f(\mathbb{Y}, \mathbb{S}|\hat{\vartheta}). \tag{C.1}$$

Theoretically, one can use an arbitrary multivariate multinomial density $g(\mathbb{S})$[6] that corresponds to the dimension of the distribution of $\mathbb{S}$ to simulate a set of N_s trial state sequences denoted as $\{\tilde{\mathbb{S}}_1, \cdots, \tilde{\mathbb{S}}_{N_s}\}$, and a natural estimate is obtained by:

$$\hat{\mathbb{S}} = \arg\max_{n_s} f(\tilde{\mathbb{S}}_{n_s}, \mathbb{Y}|\vartheta), \quad n_s = 1 : N_s. \tag{C.2}$$

Intuitively, as $N_s \to \infty$ and $g(\tilde{\mathbb{S}}) > 0$ for all $\tilde{\mathbb{S}}$, this algorithm will exploit all possible realizations of $\tilde{\mathbb{S}}$ and obtain the one that maximizes the joint likelihood. An obvious problem is that this algorithm is extremely inefficient if $g(\mathbb{S})$ is not chosen properly, and a good choice of $g(\mathbb{S})$ should account for the information in the observed data to improve the efficiency of the simulation. Our solution is to continue performing the S-step N_s times after the convergence of the SAEM algorithm with parameters fixed to the converged value $\hat{\vartheta}$, so that $g(\mathbb{S})$ in our case becomes the single move sampler. The underlying reason is that the conditional posterior probabilities used in the single move sampler include information of the observed data, therefore is much more efficient than an uninformed sampler such as a plain Markov chain. In our empirical application we choose $N_s = 1000$.

[6] A possible choice of $g(\mathbb{S})$ can be, for example, a plain M-state Markov chain with the transition parameters $\hat{\Pi}$.

Supplement to 'High-Frequency Volatility Modelling: A Markov-Switching Autoregressive Conditional Intensity Model'

Yifan Li* Ingmar Nolte† Sandra Nolte‡

University of Manchester Lancaster University Lancaster University

1 Sensitivity Analysis on the Choice of δ

The aim of this section is to show that the empirical findings in our paper are robust to different choices of δ, the threshold used to construct price durations. Recall that in our paper, we choose a daily δ which makes the daily average price durations to be as close to 5 minutes as possible. We will refer to this choice as the base δ (δ_{base}) with a 5-minute average sampling interval. In this section, we consider alternative daily δs constructed from the 10 sampled assets with sampling intervals of 2.5 minutes and 10 minutes, which are referred to as the low δ (δ_{low}) and high δ (δ_{high}), respectively. We firstly present the daily choices of these δs in Figure 1. We can see that the base δ is always between the low and the high δ as expected. Heuristically, these daily δs are daily estimators of the square root of the average integrated variance over the corresponding sampling intervals, which satisfy the following approximated relation for each day:

$$\delta_{low} \approx \sqrt{2}\delta_{base} \approx 2\delta_{high}. \tag{1.1}$$

Intuitively, a larger (resp. smaller) choice of δ samples the high-frequency price process less (resp. more) frequently, which reduces (resp. increases) the amount of sampled observations but also weakens (resp. strengthens) the impact of market microstructure noises to the sampling procedure (Hong et al., 2020). As the different choices of δ directly change the number of price

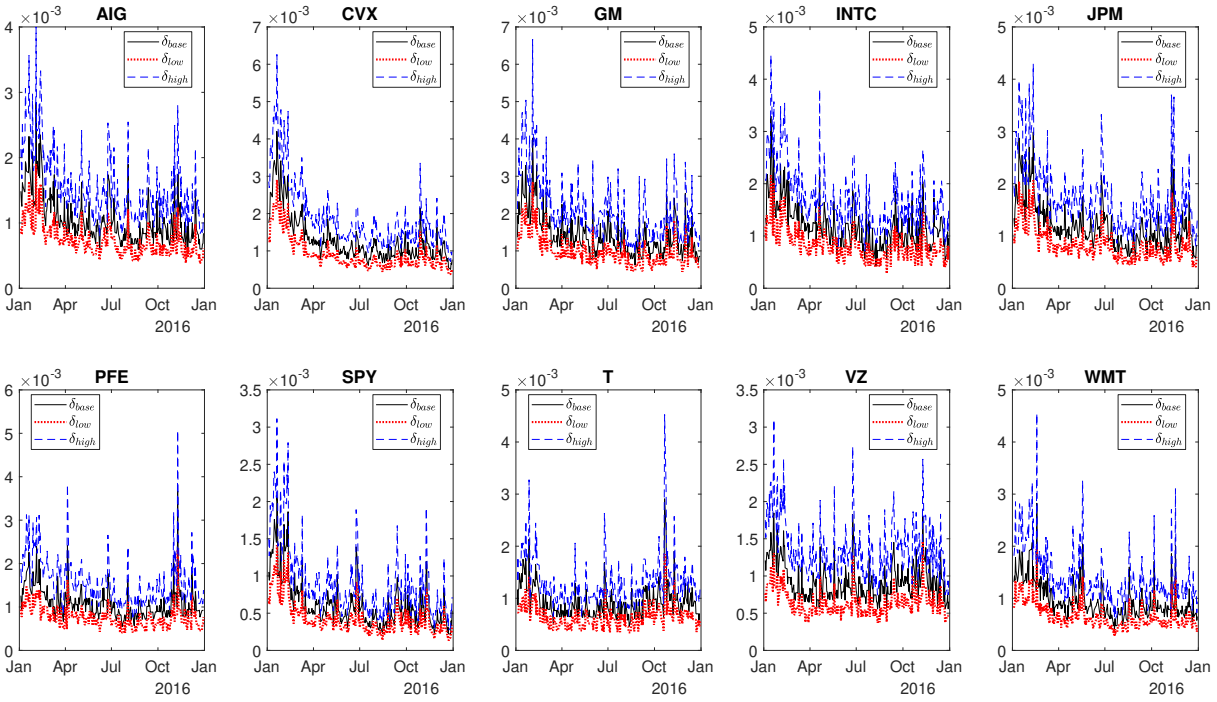
*Corresponding author: Alliance Manchester Business School, Booth Street W, Manchester, M15 6PB, UK. e-mail: yifan.li@manchester.ac.uk.

†Lancaster University Management School Bailrigg, Lancaster LA1 4YX, UK. Phone +44 15245 92644, email: I.Nolte@lancaster.ac.uk

‡Lancaster University Management School, Bailrigg, Lancaster, LA1 4YX, UK. Phone +44 15245 93634, e-mail: S.Nolte@lancaster.ac.uk.

durations and hence the sampling frequency, results in this section serve as a robustness check to show that the regime identification in our paper is insensitive to the choice of δ.

Figure 1: Daily choices of δ_{base}, δ_{low} and δ_{high}



Note: The figure plots daily price change threshold δ for the 10 sampled assets. For each asset, the base (resp. low, high) δ is calculated as the maximum threshold that produces a mean price duration which is closest to a 5-minute (resp. 2.5-minute, 10-minute) interval.

After constructing price durations and the associated cumulative volume measures using the three choices of δ, we follow the same deseasonalization procedure as in Eq. (5.1) of the paper. We present descriptive statistics of the deseasonalized datasets constructed with the low and high δs in Tables 1 and 2. Comparing these two tables with Table 3 of the paper, we find that the average price durations with δ_{low} and δ_{high} correspond to roughly 2.5 and 10 minutes as expected. The quantiles of the price duration distribution with δ_{low} and δ_{high} also appear to be proportional to those based on δ_{base}. A corresponding decrease (resp. increase) can be observed from $\ln \widetilde{Vol}_i$ with δ_{low} (resp. δ_{high}) due to the shorter (resp. longer) price durations on average. Behaviours of BAS_i are qualitatively similar across different choices of δ.

We proceed to estimate the MS(2)-ACI(1,1) model monthly for the 10 sampled assets with data constructed from two additional choices of δs. To present our findings in a concise manner and demonstrate that the two regimes we detected are insensitive to the choice of δ, we summarizes the monthly estimates of $\hat{\eta}(l)$, $\hat{\gamma}_{ll}$ and $\widehat{SoR}(l)$ for each stock and choice of δ in Tables 3 to 5, which describe the main features of the two regimes detected in our paper.

Table 1: Descriptive statistics of $\tilde{x}_i^{(\delta)}$, $\ln \widetilde{Vol}_i$ and BAS_i for the sampled assets with δ_{low}

Ticker	Variable	#Obs.	Mean	Std. Dev.	Skew.	Kurt.	Min	Q(25%)	Median	Q(75%)	Max
	$\tilde{x}_i^{(\delta)}$	38004	152.557	169.261	3.622	38.078	0.000	46.364	101.085	199.109	4963.715
AIG	$\ln \widetilde{Vol}_i$	38004	9.667	1.202	-0.304	4.907	0.000	8.926	9.682	10.438	16.059
	BAS_i	38004	0.016	0.025	9.332	117.126	0.010	0.010	0.010	0.010	0.490
	$\tilde{x}_i^{(\delta)}$	39106	150.293	158.168	2.869	18.737	0.000	48.372	101.570	197.710	2801.094
CVX	$\ln \widetilde{Vol}_i$	39106	9.955	1.116	-0.368	5.047	0.000	9.264	9.971	10.686	16.553
	BAS_i	39106	0.025	0.037	5.923	49.594	0.010	0.010	0.020	0.030	0.570
	$\tilde{x}_i^{(\delta)}$	37212	155.517	189.116	4.506	52.896	0.000	42.739	97.463	198.698	4924.328
GM	$\ln \widetilde{Vol}_i$	37212	10.341	1.189	-0.321	4.692	0.000	9.605	10.368	11.119	16.349
	BAS_i	37212	0.012	0.011	13.163	252.718	0.010	0.010	0.010	0.010	0.390
	$\tilde{x}_i^{(\delta)}$	36214	156.949	194.012	3.560	25.032	0.000	40.091	95.504	200.953	3305.375
INTC	$\ln \widetilde{Vol}_i$	36214	10.870	1.239	-0.237	5.769	0.000	10.144	10.892	11.631	18.904
	BAS_i	36214	0.012	0.011	10.233	141.814	0.010	0.010	0.010	0.010	0.310
	$\tilde{x}_i^{(\delta)}$	38421	151.455	167.341	3.290	23.343	0.000	47.235	99.821	195.761	3161.599
JPM	$\ln \widetilde{Vol}_i$	38421	10.682	1.099	-0.261	4.525	0.000	9.990	10.706	11.395	16.888
	BAS_i	38421	0.013	0.016	11.579	201.876	0.010	0.010	0.010	0.010	0.500
	$\tilde{x}_i^{(\delta)}$	36563	157.709	193.509	5.010	94.644	0.000	41.297	97.168	203.299	7795.834
PFE	$\ln \widetilde{Vol}_i$	36563	11.153	1.295	-0.364	5.035	1.263	10.374	11.172	11.977	18.120
	BAS_i	36563	0.011	0.009	15.413	392.933	0.010	0.010	0.010	0.010	0.470
	$\tilde{x}_i^{(\delta)}$	38291	151.029	180.980	3.744	36.272	0.000	40.707	94.122	194.172	5126.813
T	$\ln \widetilde{Vol}_i$	38291	9.961	1.192	-0.260	5.048	0.000	9.221	9.981	10.715	17.195
	BAS_i	38291	0.017	0.026	8.446	101.198	0.010	0.010	0.010	0.020	0.490
	$\tilde{x}_i^{(\delta)}$	36296	158.229	197.045	4.848	70.305	0.000	41.752	97.304	202.911	6981.422
VZ	$\ln \widetilde{Vol}_i$	36296	10.932	1.213	-0.412	6.134	0.000	10.213	10.955	11.694	17.850
	BAS_i	36296	0.011	0.008	11.020	177.537	0.010	0.010	0.010	0.010	0.240
	$\tilde{x}_i^{(\delta)}$	37759	152.163	182.225	3.760	32.813	0.000	41.800	94.727	196.742	4487.068
WMT	$\ln \widetilde{Vol}_i$	37759	10.369	1.199	-0.232	5.432	0.000	9.642	10.382	11.124	17.465
	BAS_i	37759	0.013	0.014	12.462	223.483	0.010	0.010	0.010	0.010	0.460
	$\tilde{x}_i^{(\delta)}$	38857	151.443	160.879	3.513	38.134	0.000	47.494	102.205	200.436	4540.227
SPY	$\ln \widetilde{Vol}_i$	38857	12.560	0.968	-0.278	3.617	3.918	11.928	12.591	13.233	15.917
	BAS_i	38857	0.011	0.005	41.398	2935.277	0.010	0.010	0.010	0.010	0.430

Note: The statistics in the table are computed from $\tilde{x}_i^{(\delta)}$, $\ln \widetilde{Vol}_i$ and BAS_i for the 10 sampled assets constructed from all trading days in 2016. #Obs. is the number of observations in the sample. Std. Dev., Skew. and Kurt. represent sample standard deviation, sample skewness and sample kurtosis, respectively.

Table 2: Descriptive statistics of $\tilde{x}_i^{(\delta)}$, $\ln \widetilde{Vol}_i$ and BAS_i for the sampled assets with δ_{high}

Ticker	Variable	#Obs.	Mean	Std. Dev.	Skew.	Kurt.	Min	Q(25%)	Median	Q(75%)	Max
AIG	$\tilde{x}_i^{(\delta)}$	9670	583.841	622.822	3.253	24.222	0.000	195.747	395.806	751.716	10611.492
	$\ln \widetilde{Vol}_i$	9670	11.062	1.191	-0.633	6.208	0.786	10.344	11.081	11.827	15.934
	BAS_i	9670	0.020	0.040	6.880	58.469	0.010	0.010	0.010	0.010	0.490
CVX	$\tilde{x}_i^{(\delta)}$	9746	588.018	584.963	2.548	14.101	0.000	201.942	408.542	768.275	7037.280
	$\ln \widetilde{Vol}_i$	9746	11.372	1.076	-0.482	4.815	3.898	10.715	11.387	12.092	16.070
	BAS_i	9746	0.030	0.050	5.210	35.448	0.010	0.010	0.020	0.030	0.570
GM	$\tilde{x}_i^{(\delta)}$	9572	579.369	645.276	3.585	27.661	0.002	183.499	384.624	748.542	10136.990
	$\ln \widetilde{Vol}_i$	9572	11.735	1.138	-0.534	5.118	2.300	11.041	11.780	12.475	15.943
	BAS_i	9572	0.014	0.018	8.831	104.595	0.010	0.010	0.010	0.010	0.360
INTC	$\tilde{x}_i^{(\delta)}$	9452	586.035	679.577	3.643	28.908	0.001	171.576	378.618	751.911	12006.922
	$\ln \widetilde{Vol}_i$	9452	12.301	1.164	-0.406	5.494	2.277	11.597	12.324	13.031	18.007
	BAS_i	9452	0.013	0.017	8.311	87.289	0.010	0.010	0.010	0.010	0.310
JPM	$\tilde{x}_i^{(\delta)}$	9722	577.341	620.490	3.660	30.491	0.000	189.206	391.902	751.928	9562.011
	$\ln \widetilde{Vol}_i$	9722	12.083	1.057	-0.415	4.808	5.053	11.427	12.109	12.779	16.512
	BAS_i	9722	0.015	0.024	8.918	113.904	0.010	0.010	0.010	0.010	0.500
PFE	$\tilde{x}_i^{(\delta)}$	9474	598.237	692.027	4.372	45.552	0.000	179.776	393.970	779.212	14334.969
	$\ln \widetilde{Vol}_i$	9474	12.556	1.266	-0.711	6.494	1.246	11.819	12.587	13.368	17.395
	BAS_i	9474	0.013	0.014	10.879	190.730	0.010	0.010	0.010	0.010	0.470
T	$\tilde{x}_i^{(\delta)}$	9678	579.861	650.027	3.382	26.917	0.000	172.555	372.896	760.210	10764.022
	$\ln \widetilde{Vol}_i$	9678	11.348	1.190	-0.612	6.615	0.000	10.633	11.361	12.117	16.929
	BAS_i	9678	0.022	0.041	6.462	53.405	0.010	0.010	0.010	0.020	0.490
VZ	$\tilde{x}_i^{(\delta)}$	9552	584.303	642.976	3.380	25.376	0.002	183.433	384.774	756.509	10736.764
	$\ln \widetilde{Vol}_i$	9552	12.326	1.166	-0.785	7.216	0.000	11.651	12.373	13.056	17.250
	BAS_i	9552	0.012	0.012	8.327	96.507	0.010	0.010	0.010	0.010	0.240
WMT	$\tilde{x}_i^{(\delta)}$	9598	580.314	647.157	3.433	26.681	0.001	176.988	377.673	751.826	11044.532
	$\ln \widetilde{Vol}_i$	9598	11.782	1.198	-0.581	5.502	3.902	11.075	11.808	12.536	16.900
	BAS_i	9598	0.015	0.024	8.545	96.202	0.010	0.010	0.010	0.010	0.460
SPY	$\tilde{x}_i^{(\delta)}$	9697	596.224	580.501	3.477	34.416	0.060	210.046	434.779	804.290	10366.776
	$\ln \widetilde{Vol}_i$	9697	13.975	0.914	-0.294	3.324	6.875	13.374	14.026	14.618	16.943
	BAS_i	9697	0.011	0.005	43.836	2892.766	0.010	0.010	0.010	0.010	0.390

Note: The statistics in the table are computed from $\tilde{x}_i^{(\delta)}$, $\ln \widetilde{Vol}_i$ and BAS_i for the 10 sampled assets constructed from all trading days in 2016. #Obs. is the number of observations in the sample. Std. Dev., Skew. and Kurt. represent sample standard deviation, sample skewness and sample kurtosis, respectively.

Table 3: Descriptive statistics of monthly estimated $\hat{\eta}(l)$, $\hat{\gamma}_{ll}$ and $\widehat{SoR}(l)$ based on δ_{base}

		Regime 1 ($l=1$)						Regime 2 ($l=2$)					
Ticker		Mean	Std. Dev.	Q(25%)	Median	Q(75%)	#5% Sig.	Mean	Std. Dev.	Q(25%)	Median	Q(75%)	#5% Sig.
	$\hat{\eta}(l)$	-0.495	0.094	-0.574	-0.494	-0.422	12	-2.337	0.238	-2.481	-2.317	-2.157	12
AIG	$\hat{\gamma}_{ll}$	0.496	0.498	0.188	0.447	0.719	6	-2.717	0.499	-3.143	-2.684	-2.307	12
	$\widehat{SoR}(l)$	0.987	0.006	0.984	0.988	0.993		0.982	0.007	0.976	0.984	0.989	
	$\hat{\eta}(l)$	-0.600	0.168	-0.661	-0.634	-0.530	12	-2.733	0.222	-2.892	-2.670	-2.551	12
CVX	$\hat{\gamma}_{ll}$	0.314	0.330	0.115	0.223	0.393	4	-2.243	0.486	-2.431	-2.107	-2.014	12
	$\widehat{SoR}(l)$	0.984	0.004	0.982	0.984	0.985		0.978	0.005	0.975	0.978	0.981	
	$\hat{\eta}(l)$	-0.502	0.223	-0.562	-0.436	-0.404	12	-2.178	0.171	-2.329	-2.206	-2.020	12
GM	$\hat{\gamma}_{ll}$	1.645	2.864	0.471	0.829	1.168	4	-3.737	1.002	-4.441	-3.435	-3.028	12
	$\widehat{SoR}(l)$	0.986	0.007	0.982	0.987	0.991		0.981	0.009	0.978	0.983	0.988	
	$\hat{\eta}(l)$	-0.301	0.105	-0.401	-0.278	-0.278	10	-2.326	0.233	-2.414	-2.303	-2.204	12
INTC	$\hat{\gamma}_{ll}$	5.960	7.504	0.672	1.101	15.504	9	-4.204	0.619	-4.566	-4.287	-3.775	12
	$\widehat{SoR}(l)$	0.990	0.003	0.988	0.991	0.993		0.988	0.004	0.986	0.988	0.990	
	$\hat{\eta}(l)$	-0.687	0.199	-0.817	-0.708	-0.632	12	-2.731	0.262	-2.878	-2.724	-2.560	12
JPM	$\hat{\gamma}_{ll}$	0.730	0.660	0.182	0.694	0.998	6	-3.356	0.892	-3.949	-3.274	-2.830	12
	$\widehat{SoR}(l)$	0.987	0.008	0.982	0.987	0.994		0.983	0.009	0.978	0.983	0.991	
	$\hat{\eta}(l)$	-0.498	0.164	-0.617	-0.458	-0.368	12	-2.422	0.208	-2.552	-2.391	-2.250	12
PFE	$\hat{\gamma}_{ll}$	11.946	7.055	8.222	15.211	16.419	11	-3.899	0.657	-4.358	-3.944	-3.584	12
	$\widehat{SoR}(l)$	0.988	0.006	0.986	0.990	0.992		0.985	0.008	0.983	0.987	0.989	
	$\hat{\eta}(l)$	-0.524	0.184	-0.619	-0.489	-0.388	12	-2.573	0.227	-2.674	-2.590	-2.389	12
T	$\hat{\gamma}_{ll}$	6.988	7.052	1.187	2.354	15.145	10	-4.206	0.806	-4.650	-4.024	-3.612	12
	$\widehat{SoR}(l)$	0.985	0.005	0.981	0.986	0.988		0.981	0.006	0.976	0.983	0.986	
	$\hat{\eta}(l)$	-0.466	0.123	-0.543	-0.453	-0.399	12	-2.533	0.229	-2.626	-2.515	-2.362	12
VZ	$\hat{\gamma}_{ll}$	3.367	6.138	0.444	0.879	1.876	6	-3.508	0.455	-3.736	-3.476	-3.346	12
	$\widehat{SoR}(l)$	0.985	0.004	0.983	0.986	0.988		0.981	0.006	0.979	0.982	0.984	
	$\hat{\eta}(l)$	-0.530	0.148	-0.622	-0.543	-0.487	12	-2.622	0.215	-2.809	-2.661	-2.433	12
WMT	$\hat{\gamma}_{ll}$	0.503	0.318	0.239	0.472	0.669	6	-2.276	0.199	-2.440	-2.258	-2.166	12
	$\widehat{SoR}(l)$	0.980	0.004	0.977	0.980	0.984		0.973	0.006	0.969	0.973	0.977	
	$\hat{\eta}(l)$	-3.325	0.990	-3.888	-3.339	-2.963	12	-3.146	0.375	-3.326	-3.101	-2.857	12
SPY	$\hat{\gamma}_{ll}$	2.126	8.541	-0.931	0.588	7.481	5	1.510	5.524	-1.174	-0.645	0.380	4
	$\widehat{SoR}(l)$	0.927	0.034	0.921	0.934	0.942		0.917	0.037	0.916	0.923	0.931	

Note: The table reports descriptive statistics of monthly estimated $\hat{\eta}(l)$, $\hat{\gamma}_{ll}$ and $\widehat{SoR}(l)$ from the MS(2)-ACI(1,1) model with price duration data associated with δ_{base}. The column #5% Sig. is the number of of significant parameter estimates out of the 12 monthly estimates.

Table 4: Descriptive statistics of monthly estimated $\hat{\eta}(l)$, $\hat{\gamma}_{ll}$ and $\widehat{SoR}(l)$ based on δ_{low}

		Regime 1 ($l=1$)						Regime 2 ($l=2$)					
Ticker		Mean	Std. Dev.	Q(25%)	Median	Q(75%)	#5% Sig.	Mean	Std. Dev.	Q(25%)	Median	Q(75%)	#5% Sig.
	$\hat{\eta}(l)$	-0.403	0.104	-0.467	-0.391	-0.331	12	-1.976	0.183	-2.142	-1.923	-1.812	12
AIG	$\hat{\gamma}_{ll}$	0.902	0.407	0.576	0.862	1.027	11	-2.581	0.309	-2.765	-2.491	-2.375	12
	$\widehat{SoR}(l)$	0.986	0.004	0.982	0.987	0.988		0.980	0.006	0.975	0.982	0.984	
	$\hat{\eta}(l)$	-0.487	0.162	-0.590	-0.448	-0.344	12	-2.304	0.272	-2.402	-2.306	-2.107	12
CVX	$\hat{\gamma}_{ll}$	0.469	0.205	0.371	0.437	0.629	9	-1.828	0.290	-1.928	-1.766	-1.660	12
	$\widehat{SoR}(l)$	0.982	0.006	0.980	0.984	0.987		0.976	0.007	0.972	0.979	0.981	
	$\hat{\eta}(l)$	-0.443	0.127	-0.545	-0.406	-0.371	12	-1.821	0.141	-1.950	-1.806	-1.682	12
GM	$\hat{\gamma}_{ll}$	4.918	6.415	1.039	1.574	8.850	10	-3.089	0.467	-3.478	-3.031	-2.894	12
	$\widehat{SoR}(l)$	0.985	0.006	0.981	0.988	0.989		0.981	0.007	0.976	0.983	0.986	
	$\hat{\eta}(l)$	-0.336	0.089	-0.363	-0.340	-0.280	12	-1.916	0.140	-1.946	-1.876	-1.851	12
INTC	$\hat{\gamma}_{ll}$	2.902	4.395	1.087	1.478	2.449	11	-3.750	0.502	-3.961	-3.725	-3.369	12
	$\widehat{SoR}(l)$	0.989	0.004	0.987	0.989	0.991		0.985	0.005	0.983	0.986	0.988	
	$\hat{\eta}(l)$	-0.610	0.152	-0.737	-0.626	-0.528	11	-2.380	0.212	-2.542	-2.453	-2.207	12
JPM	$\hat{\gamma}_{ll}$	0.961	0.756	0.393	0.832	1.327	8	-2.756	0.477	-2.953	-2.809	-2.556	12
	$\widehat{SoR}(l)$	0.983	0.006	0.979	0.984	0.987		0.978	0.007	0.973	0.980	0.984	
	$\hat{\eta}(l)$	-0.461	0.134	-0.522	-0.424	-0.370	12	-2.047	0.162	-2.165	-2.058	-1.892	12
PFE	$\hat{\gamma}_{ll}$	9.491	7.075	1.794	15.018	15.222	12	-3.546	0.455	-3.856	-3.548	-3.187	12
	$\widehat{SoR}(l)$	0.983	0.006	0.980	0.984	0.988		0.979	0.007	0.976	0.979	0.984	
	$\hat{\eta}(l)$	-0.367	0.119	-0.486	-0.361	-0.275	12	-2.115	0.218	-2.337	-2.096	-2.001	12
T	$\hat{\gamma}_{ll}$	11.844	6.109	6.614	15.290	15.931	12	-3.684	0.669	-4.204	-3.733	-3.344	12
	$\widehat{SoR}(l)$	0.985	0.006	0.979	0.988	0.990		0.982	0.007	0.974	0.985	0.988	
	$\hat{\eta}(l)$	-0.388	0.090	-0.453	-0.382	-0.331	12	-2.158	0.170	-2.248	-2.127	-2.056	12
VZ	$\hat{\gamma}_{ll}$	5.284	6.699	1.335	1.661	9.459	12	-3.297	0.365	-3.456	-3.246	-3.130	12
	$\widehat{SoR}(l)$	0.985	0.004	0.982	0.985	0.989		0.980	0.005	0.977	0.981	0.984	
	$\hat{\eta}(l)$	-0.454	0.055	-0.508	-0.450	-0.420	12	-2.289	0.187	-2.414	-2.291	-2.162	12
WMT	$\hat{\gamma}_{ll}$	0.746	0.436	0.432	0.693	1.110	10	-2.235	0.461	-2.598	-2.062	-1.868	12
	$\widehat{SoR}(l)$	0.978	0.006	0.974	0.978	0.982		0.972	0.006	0.967	0.970	0.976	
	$\hat{\eta}(l)$	-2.934	0.880	-3.395	-2.959	-2.719	12	-2.737	0.360	-2.917	-2.620	-2.480	12
SPY	$\hat{\gamma}_{ll}$	4.295	6.529	0.234	0.600	9.031	4	-0.869	0.501	-1.297	-0.988	-0.441	5
	$\widehat{SoR}(l)$	0.913	0.032	0.897	0.913	0.925		0.900	0.035	0.885	0.895	0.914	

Note: The table reports descriptive statistics of monthly estimated $\hat{\eta}(l)$, $\hat{\gamma}_{ll}$ and $\widehat{SoR}(l)$ from the MS(2)-ACI(1,1) model with price duration data associated with δ_{low}. The column #5% Sig. is the number of of significant parameter estimates out of the 12 monthly estimates.

Table 5: Descriptive statistics of monthly estimated $\hat{\eta}(l)$, $\hat{\gamma}_{ll}$ and $\widehat{SoR}(l)$ based on δ_{high}

		Regime 1 ($l=1$)						Regime 2 ($l=2$)					
Ticker		Mean	Std. Dev.	Q(25%)	Median	Q(75%)	#5% Sig.	Mean	Std. Dev.	Q(25%)	Median	Q(75%)	#5% Sig.
	$\hat{\eta}(l)$	-0.648	0.262	-0.721	-0.623	-0.487	12	-2.577	0.342	-2.841	-2.542	-2.304	12
AIG	$\hat{\gamma}_{ll}$	0.586	0.616	0.263	0.817	0.930	5	-7.176	14.059	-4.207	-3.008	-2.296	12
	$\widehat{SoR}(l)$	0.988	0.005	0.984	0.990	0.992		0.984	0.007	0.980	0.985	0.989	
	$\hat{\eta}(l)$	-1.016	0.387	-1.432	-0.863	-0.688	12	-3.227	0.365	-3.445	-3.396	-2.949	12
CVX	$\hat{\gamma}_{ll}$	0.647	1.650	-0.029	0.280	0.518	2	-3.035	1.042	-4.019	-2.713	-2.253	12
	$\widehat{SoR}(l)$	0.976	0.018	0.970	0.981	0.986		0.969	0.021	0.960	0.976	0.981	
	$\hat{\eta}(l)$	-0.844	0.559	-1.076	-0.740	-0.484	11	-2.550	0.358	-2.625	-2.468	-2.347	12
GM	$\hat{\gamma}_{ll}$	0.943	0.720	0.291	0.903	1.299	7	-3.300	1.356	-3.627	-3.168	-2.571	12
	$\widehat{SoR}(l)$	0.956	0.062	0.957	0.981	0.990		0.952	0.061	0.950	0.975	0.987	
	$\hat{\eta}(l)$	-0.283	0.128	-0.376	-0.304	-0.176	10	-2.583	0.269	-2.666	-2.588	-2.379	12
INTC	$\hat{\gamma}_{ll}$	4.884	11.188	0.164	1.263	9.498	8	-4.520	1.013	-5.236	-4.318	-3.960	12
	$\widehat{SoR}(l)$	0.990	0.004	0.989	0.991	0.992		0.988	0.004	0.987	0.988	0.991	
	$\hat{\eta}(l)$	-0.740	0.248	-0.915	-0.816	-0.552	12	-3.099	0.297	-3.215	-3.103	-2.858	12
JPM	$\hat{\gamma}_{ll}$	12.236	25.902	0.301	1.111	14.269	9	-3.358	0.286	-3.531	-3.390	-3.263	12
	$\widehat{SoR}(l)$	0.989	0.005	0.986	0.989	0.994		0.986	0.006	0.984	0.985	0.992	
	$\hat{\eta}(l)$	-0.664	0.339	-0.884	-0.537	-0.427	12	-2.825	0.452	-3.117	-2.696	-2.480	12
PFE	$\hat{\gamma}_{ll}$	8.857	8.082	0.842	9.227	16.528	7	-6.537	6.579	-5.786	-4.823	-3.939	12
	$\widehat{SoR}(l)$	0.984	0.012	0.984	0.986	0.989		0.980	0.013	0.980	0.981	0.984	
	$\hat{\eta}(l)$	-0.575	0.165	-0.714	-0.563	-0.454	12	-2.812	0.406	-3.083	-2.926	-2.498	12
T	$\hat{\gamma}_{ll}$	7.641	7.985	0.776	4.178	16.240	7	-6.242	7.040	-4.999	-4.155	-3.675	10
	$\widehat{SoR}(l)$	0.988	0.005	0.984	0.988	0.992		0.984	0.007	0.979	0.985	0.989	
	$\hat{\eta}(l)$	-0.577	0.227	-0.635	-0.551	-0.392	12	-2.825	0.294	-3.003	-2.907	-2.554	12
VZ	$\hat{\gamma}_{ll}$	3.775	6.865	0.391	1.099	1.685	6	-6.431	8.598	-4.470	-4.225	-3.813	12
	$\widehat{SoR}(l)$	0.986	0.010	0.985	0.988	0.992		0.982	0.011	0.982	0.985	0.987	
	$\hat{\eta}(l)$	-0.655	0.468	-0.720	-0.541	-0.375	12	-2.919	0.263	-3.082	-2.988	-2.771	12
WMT	$\hat{\gamma}_{ll}$	0.363	0.686	-0.287	0.570	0.679	6	-2.731	1.495	-2.809	-2.333	-2.124	12
	$\widehat{SoR}(l)$	0.978	0.013	0.976	0.980	0.985		0.971	0.015	0.966	0.975	0.980	
	$\hat{\eta}(l)$	-4.164	0.868	-4.446	-3.976	-3.577	12	-3.641	0.481	-3.976	-3.630	-3.379	12
SPY	$\hat{\gamma}_{ll}$	4.254	12.776	-6.715	7.546	15.528	7	2.050	6.789	-2.102	-0.826	6.213	5
	$\widehat{SoR}(l)$	0.902	0.049	0.894	0.902	0.928		0.895	0.041	0.890	0.894	0.917	

Note: The table reports descriptive statistics of monthly estimated $\hat{\eta}(l)$, $\hat{\gamma}_{ll}$ and $\widehat{SoR}(l)$ from the MS(2)-ACI(1,1) model with price duration data associated with δ_{high}. The column #5% Sig. is the number of of significant parameter estimates out of the 12 monthly estimates.

Comparing the results in Tables 3 to 5, we see the following common features across assets and choices of δ: (1) All estimates of $\hat{\eta}(l)$ are significant at 5%. For individual stocks, $\hat{\eta}(1)$ is much smaller in magnitude than $\hat{\eta}(2)$, indicating that volumes are much less associated with the conditional intensity process in regime 1 relative to regime 2. This effect clearly cannot be seen on SPY, whose $\hat{\eta}(l)$ estimates are similar in both regimes for all choices of δs. (2) For individual stocks, the estimates $\hat{\gamma}_{22}$ are negative and strongly significant for all individual stocks across different δs. This shows that observations that stay in regime 2 are associated with a low bid-ask spread. For $\hat{\gamma}_{11}$ it is overall positive but less significant that $\hat{\gamma}_{11}$. This effect is also not visible for SPY, as both $\hat{\gamma}_{11}$ and $\hat{\gamma}_2$ appear positive and only weakly significant. (3) All estimated $\widehat{SpR}(l)$ are larger than 0.95 for individual stocks with different choices of δs. However, for SPY, this is only around 0.9. This shows that the regime identification is on average stronger for the individual stocks than SPY. This is clear evidence supporting the argument that the parameter estimates and regime identification of the MS(2)-ACI(1,1) model are robust to the choice of δ.

In our paper, we interpret regime 1 as a regime driven by the news arrivals, while regime 2 represents the risk transfer activity of market participants. We show that this interpretation is also applicable when a different choice of δ is used. To this end, we firstly present the regime-specific annual averages of $\hat{s}_i$, $\tilde{x}_i^{(\delta)}$, $\ln \widetilde{Vol}_i$ and BAS_i for each asset under different choices of δ in Table 6, where $\hat{s}_i$ is the estimated regime for the i-th observation. The table convincingly shows that, for individual stocks, regime 2 is the dominant regime for all individual stocks which accounts for more than 80% of data with a substantially longer mean duration, larger mean log-cumulative volume and a much smaller bid-ask spread. For SPY, we can still observe some differences in terms of $\hat{s}_i$, $\tilde{x}_i^{(\delta)}$, $\ln \widetilde{Vol}_i$, but the distinction is relatively weaker in comparison to the individual stocks. More importantly, there does not seem to be a large difference for the BAS_i of SPY in the two regimes. As these results hold for all three choices of δ, this table provides some preliminary evidence to show that the two regimes are identified by the data characteristics that are not largely affected by the choice of δ.

To further verify our findings in the paper, we reproduce Table 5 of our paper and examine the intraday volume-duration invariance for regime 2 using data based on δ_{low} and δ_{high} in Table 7. Our findings from both Table 7 and Table 5 of our paper are largely similar. We see that $\hat{b}_1(1)$ and $R^2(1)$ are in general smaller than $\hat{b}_1(2)$ and $R^2(2)$ for individual stocks, while the difference in R^2s cannot be observed for SPY. The high $R^2(2)$s for individual stocks suggest that a power law relation between $\tilde{x}_i^{(\delta)}$ and $\widetilde{Vol}_i$ also holds for observations in regime 2 when we use a smaller or larger δ, which is in line with our hypothesis of volume-duration invariance. Interestingly,

Table 6: Annual averages of $\hat{s}_i$, $\tilde{x}_i^{(\delta)}$, $\ln \widetilde{Vol}_i$ and BAS_i for regimes 1 and 2 based on different choices of δ

	Regime 1				Regime 2			
Ticker	$\%(\hat{s}_i = 1)$	$\tilde{x}_i^{(\delta)}$	$\ln \widetilde{Vol}_i$	BAS_i	$\%(\hat{s}_i = 2)$	$\tilde{x}_i^{(\delta)}$	$\ln \widetilde{Vol}_i$	BAS_i
Panel 1: δ_{low}								
AIG	0.1345	74.6762	9.1655	0.0423	0.8655	164.6581	9.7447	0.0119
CVX	0.0948	69.0142	9.3847	0.0781	0.9052	158.8072	10.0143	0.0199
GM	0.1298	68.0120	9.8253	0.0212	0.8702	168.5656	10.4184	0.0106
INTC	0.1097	44.1300	10.2828	0.0255	0.8903	170.8548	10.9427	0.0103
JPM	0.1079	72.5408	10.1766	0.0293	0.8921	160.9982	10.7429	0.0113
PFE	0.1281	62.9131	10.4076	0.0188	0.8719	171.6310	11.2624	0.0104
T	0.1232	64.8092	10.3074	0.0184	0.8768	171.3529	11.0202	0.0103
VZ	0.1423	68.8633	9.8865	0.0251	0.8577	165.9863	10.4492	0.0107
WMT	0.1726	79.6712	9.5328	0.0404	0.8274	165.9174	10.0506	0.0126
SPY	0.2872	121.1903	12.5384	0.0112	0.7128	163.6315	12.5693	0.0105
Panel 2: δ_{base}								
AIG	0.1143	130.5569	9.8497	0.0559	0.8857	319.0416	10.4314	0.0125
CVX	0.0850	133.7953	10.1669	0.0955	0.9150	313.8968	10.7078	0.0208
GM	0.1160	118.0308	10.5587	0.0263	0.8840	322.2345	11.0875	0.0108
INTC	0.0927	74.9235	11.0813	0.0322	0.9073	324.9734	11.6372	0.0104
JPM	0.0841	125.3751	10.8545	0.0418	0.9159	312.8143	11.4173	0.0116
PFE	0.0913	97.2756	11.0988	0.0255	0.9087	325.8391	11.9515	0.0105
T	0.1157	119.8705	11.0834	0.0216	0.8843	329.2462	11.7086	0.0103
VZ	0.1208	123.4531	10.5768	0.0331	0.8792	321.1928	11.1385	0.0109
WMT	0.1452	141.7026	10.1616	0.0523	0.8548	321.0501	10.7405	0.0133
SPY	0.2278	237.3538	13.1662	0.0110	0.7722	320.3829	13.3029	0.0106
Panel 3: δ_{high}								
AIG	0.0941	261.2665	10.5671	0.0847	0.9059	617.3509	11.1135	0.0135
CVX	0.1298	365.2893	11.0602	0.0920	0.8702	621.2395	11.4181	0.0209
GM	0.1319	320.0325	11.4618	0.0310	0.8681	618.7894	11.7764	0.0111
INTC	0.0704	171.5377	12.1823	0.0465	0.9296	617.4036	12.3104	0.0106
JPM	0.0586	204.2368	11.7830	0.0717	0.9414	600.5786	12.1012	0.0118
PFE	0.0978	225.3298	11.9930	0.0321	0.9022	638.6826	12.6167	0.0105
T	0.0876	238.9562	11.8558	0.0310	0.9124	617.4709	12.3714	0.0106
VZ	0.0969	243.4376	11.3806	0.0514	0.9031	616.4577	11.8249	0.0111
WMT	0.1448	295.7031	10.9367	0.0662	0.8552	627.9593	11.4175	0.0144
SPY	0.2702	483.1841	13.8605	0.0112	0.7298	638.0725	14.0178	0.0106

Note: in this table we present the sample mean of regime-specific s_i, $\tilde{x}_i^{(\delta)}$, $\ln \widetilde{Vol}_i$ and BAS_i for the entire sampling period, where the regimes are estimated monthly from the MS(2)-ACI(1,1) model as in Tables 3 to 5. The column $\%(\hat{s}_i = l)$ computes the annual percentage of data being classified into regime l.

we find that $\hat{b}_1(2)$ is smaller (resp. larger) to the proposed exponent 4/5 as in Hypothesis 1 of the paper when we choose a smaller (resp. larger) δ relative to the base δ. This provides ample room for future research on the volume-duration relationship under different aggregation levels, which is beyond the scope of the paper.

Table 7: Descriptive statistics of $\hat{b}_1(l)$ and $R^2(l)$ for regimes 1 and 2 based on δ_{low} and δ_{high}

Panel 1: All Individual Stocks										
	Results based on δ_{low}					Results based on δ_{high}				
	Mean	Std. Dev.	Q(25%)	Median	Q(75%)	Mean	Std. Dev.	Q(25%)	Median	Q(75%)
$\hat{b}_1(1)$	0.3853	0.0724	0.3323	0.3786	0.4429	0.5569	0.1798	0.4589	0.5718	0.6818
$\hat{b}_1(2)$	0.7638	0.0473	0.7370	0.7678	0.7935	0.8250	0.0543	0.7919	0.8239	0.8655
$R^2(1)$	0.2669	0.0858	0.1981	0.2585	0.3301	0.3830	0.1799	0.2592	0.3956	0.5234
$R^2(2)$	0.6272	0.0631	0.5895	0.6363	0.6803	0.7027	0.0776	0.6673	0.7204	0.7547
Panel 2: SPY Only										
	Results based on δ_{low}					Results based on δ_{high}				
	Mean	Std. Dev.	Q(25%)	Median	Q(75%)	Mean	Std. Dev.	Q(25%)	Median	Q(75%)
$\hat{b}_1(1)$	0.5686	0.0786	0.5538	0.5689	0.6039	0.6225	0.0550	0.5851	0.6078	0.6575
$\hat{b}_1(2)$	0.7879	0.0511	0.7438	0.7927	0.8206	0.8236	0.0494	0.7711	0.8380	0.8683
$R^2(1)$	0.6507	0.1711	0.6535	0.6801	0.7408	0.7786	0.0601	0.7364	0.7855	0.7973
$R^2(2)$	0.6525	0.0536	0.6050	0.6535	0.6965	0.7113	0.0694	0.6535	0.7132	0.7665

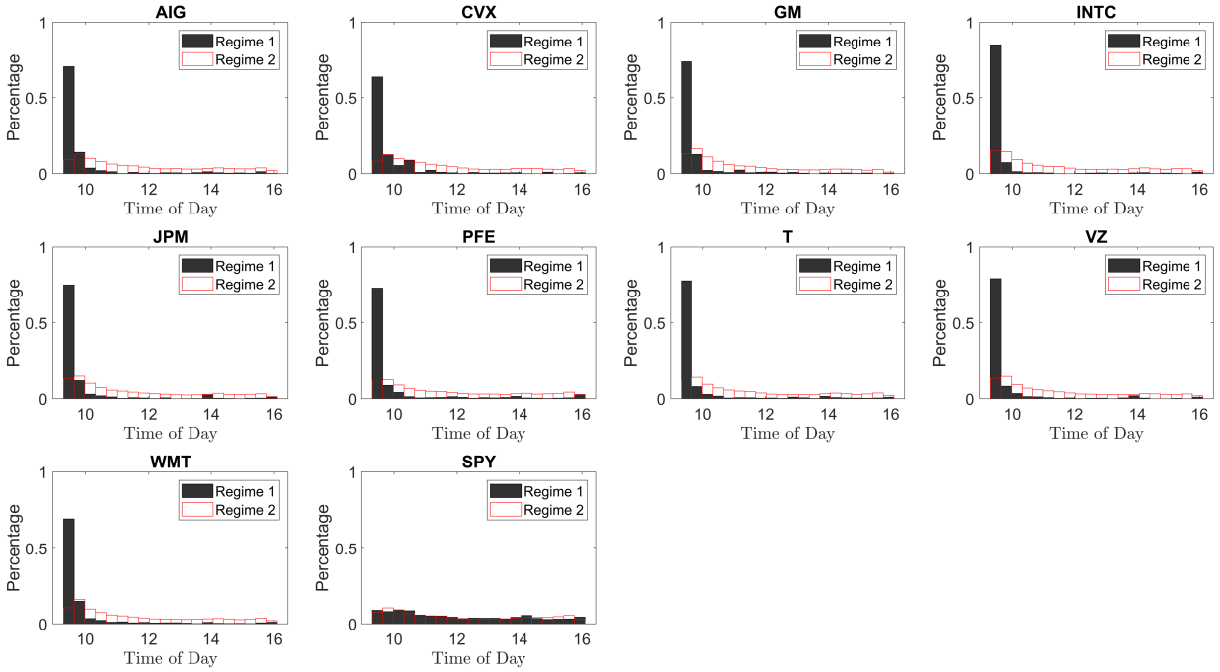
Note: the table presents descriptive statistics of estimated $\hat{b}_1(l)$ and $R^2(l)$ from $\ln \widetilde{Vol}_i = b_0(l) + b_1(l) \ln \bar{x}_i^{(\delta)} + u_i$ with $l = 1, 2$ for all 120 stock-months in our sample constructed using either δ_{low} or δ_{high}. The regime classification is estimated monthly by the MS(2)-ACI(1,1) model as in Tables 4 and 5. Results for SPY are excluded in the 'All Individual Stocks' panel and are presented separately in the 'SPY Only' panel. $Q(x\%)$ is the $x\%$ quantile.

Finally, for each asset we plot the time-of-day distribution of the observations classified into regimes 1 and 2 under different choices of δ in Figures 2 to 4. The figures show that, for individual assets, regime 1 observations largely concentrate at the beginning of a trading day which is possibly driven by the accumulation of firm-specific information overnight, while regime 2 observations spread more evenly throughout the day. This discrepancy cannot be observed from SPY as observations in both regimes have very similar time-of-day distribution. This result is again insensitive to the choice of δ considered, which demonstrates the robustness of our regime identification and interpretation.

2 Empirical Results on an Extended Sample

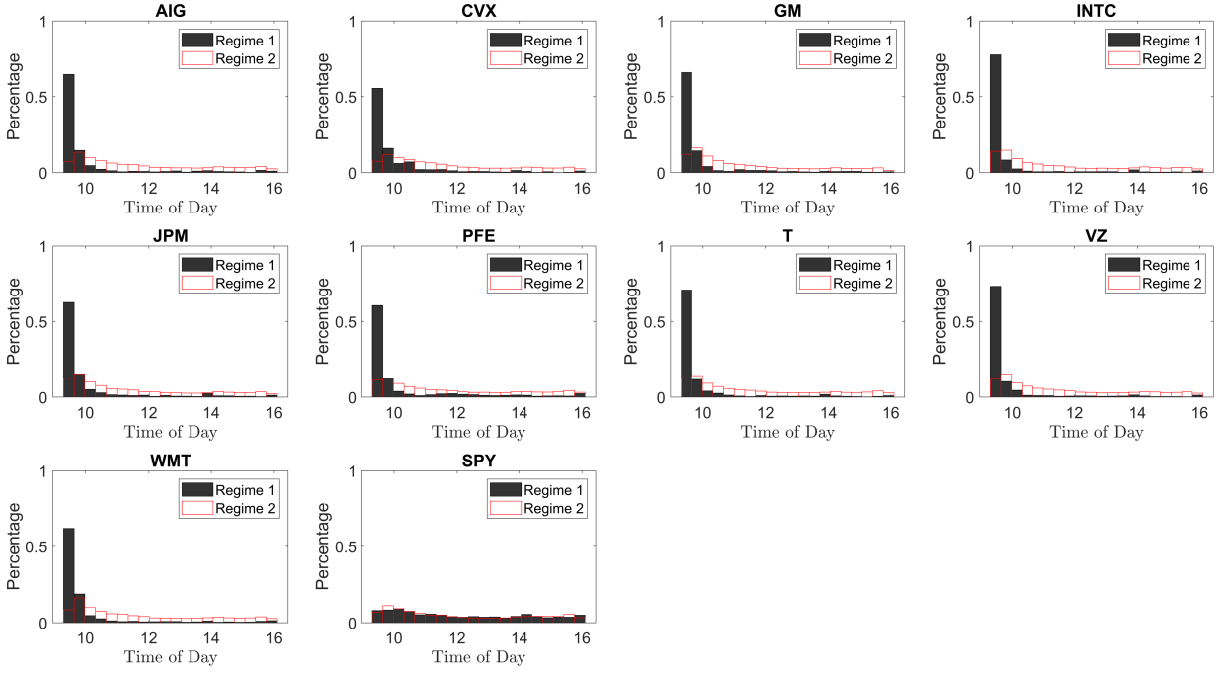
In this section, we show that our findings based on the 10 sampled assets (the main sample) in our paper can also be observed from a larger collection of stocks, which demonstrates the

Figure 2: Time-of-day distribution of regimes under δ_{base}



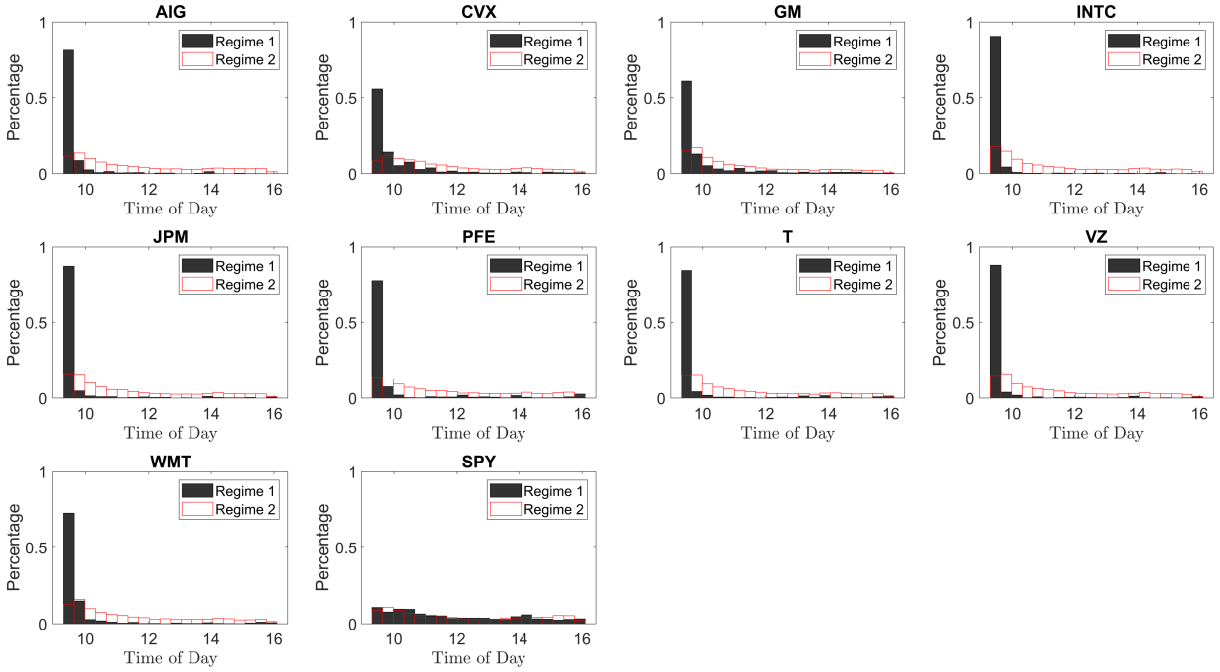
Note: The figure plots the histogram of the time of day associated with observations in regimes 1 and 2, where the regimes are estimated monthly based on the MS(2)-ACI(1,1) model under δ_{base}. The y-axis is the percentage of regime-specific observations that fall in one of the 20 equidistant time intervals.

Figure 3: Time-of-day distribution of regimes under δ_{low}



Note: The figure plots the histogram of the time of day associated with observations in regimes 1 and 2, where the regimes are estimated monthly based on the MS(2)-ACI(1,1) model under δ_{low}. The y-axis is the percentage of regime-specific observations that fall in one of the 20 equidistant time intervals.

Figure 4: Time-of-day distribution of regimes under δ_{high}



Note: The figure plots the histogram of the time of day associated with observations in regimes 1 and 2, where the regimes are estimated monthly based on the MS(2)-ACI(1,1) model under δ_{high}. The y-axis is the percentage of regime-specific observations that fall in one of the 20 equidistant time intervals.

generality of the results in our paper. Our extended sample consists of 24 additional individual stocks that cover all Dow Jones 30 constituents in 2016 (6 stocks are already included in our original sample). We present a description of the stocks and their tickers in Table 8.

Following the construction of price durations for the main sample in our paper, for each asset in the extended sample, we choose a daily δ which generates a daily average price duration that is as close to 5 minutes as possible. We deseasonalize the price durations and the log-cumulative volume in the same manner as the main sample, and estimate the MS(2)-ACI(1,1) model monthly to obtain parameter estimates and the fitted most probable state vector for each stock-month. For brevity, we do not present the values of δ and the descriptive statistics of all 24 stocks, which are available upon request.

To demonstrate that the findings on the regime identification of the MS(2)-ACI(1,1) model estimated on the main sample also hold true for the extended sample, we summarize the monthly estimated $\hat{\eta}(l)$, $\hat{\gamma}_{ll}$ and $\widehat{SoR}(l)$ for the extended sample in Table 9. The table clearly shows that for all assets in the extended sample, $\hat{\eta}(l)$ are mostly significant, with $\hat{\eta}(2)$ much larger than $\hat{\eta}(1)$ in magnitude. All estimated $\hat{\gamma}_{22}$s are negative and significant, while $\hat{\gamma}_{11}$s all appear positive with a relatively much weaker statistical significance. Finally, all average $\widehat{SoR}(l)$ estimates are

Table 8: Description of the stocks and their tickers in the extended sample

Ticker	Corporate Name	Ticker	Corporate Name
AAPL	Apple Inc.	V	Visa Inc.
AXP	American Express Company	HD	The Home Depot, Inc.
BA	The Boeing Company	IBM	International Business Machines Corporation
CAT	Caterpillar Inc.	JNJ	Johnson & Johnson
CSCO	Cisco Systems, Inc.	KO	The Coca-Cola Company
DD	E.I. du Pont de Nemours & Company	MCD	McDonald's Corporation
DIS	The Walt Disney Company	MMM	3M Company
GE	General Electric Company	MRK	Merck & Co., Inc.
GS	The Goldman Sachs Group, Inc.	MSFT	Microsoft Corporation
NKE	Nike, Inc.	PG	The Procter & Gamble Company
TRV	The Travelers Companies, Inc.	UTX	United Technologies Corporation
UNH	UnitedHealth Group Inc.	XOM	Exxon Mobil Corporation

Note: all stocks are the constituents of the Dow Jones Industrial Average index in 2016.

larger than 0.96, which suggests that the regime identification is very strong for all assets in the extended sample. This result is largely in line with our findings for the main sample in Table 3.

Following the structure of the previous section, we also show that the interpretation of the two regimes is applicable to the results of the extended sample. To this end, we firstly replicate Table 6 using the extended sample, which is presented in Table 10. The table convincingly shows that the regime-specific averages of all variables for each stock in the extended sample follow the same pattern as those for the individual stocks in Table 6.

We also replicate Table 5 of our paper again using the extended sample, and present our results in Table 11. The results in Table 11 is largely in line with Table 5 of our paper. In fact the results in Table 11 further strengthen our volume-duration invariance hypothesis as $\hat{b}_1(2) \approx 4/5$ is also observed for the extended sample. Combining the results for both samples does not alter our main conclusion.

Lastly, we compare the time-of-day distributions of observations from regimes 1 and 2 for each asset in the extended sample, which is shown in Figure 5. Comparing Figure 5 with Figure 2, it is evident that the time-of-day distributions of regimes 1 and 2 for all assets in the extended sample have the same pattern as those for the individual assets in the main sample. This corroborates our interpretation in our paper that regime 1 observations are information driven due to the aggregation of overnight firm-specific news arrivals.

Table 9: Descriptive statistics of monthly estimated $\hat{\eta}(l)$, $\hat{\gamma}_{ll}$ and $\widehat{SoR}(l)$ from the extended sample

Ticker		Mean	Std. Dev.	Q(25%)	Median	Q(75%)	#5% Sig.	Mean	Std. Dev.	Q(25%)	Median	Q(75%)	#5% Sig.
AAPL	$\hat{\eta}(l)$	-0.546	0.174	-0.613	-0.530	-0.481	11	-3.376	0.248	-3.470	-3.331	-3.198	12
	$\hat{\gamma}_{ll}$	0.510	0.651	0.358	0.524	0.916	6	-2.626	0.646	-2.836	-2.525	-2.187	12
	$\widehat{SoR}(l)$	0.987	0.006	0.982	0.988	0.990		0.983	0.007	0.977	0.983	0.987	
AXP	$\hat{\eta}(l)$	-0.543	0.186	-0.666	-0.540	-0.450	11	-2.435	0.220	-2.592	-2.466	-2.329	11
	$\hat{\gamma}_{ll}$	0.378	0.222	0.251	0.322	0.543	2	-2.436	0.422	-2.705	-2.544	-2.083	11
	$\widehat{SoR}(l)$	0.983	0.005	0.979	0.984	0.986		0.976	0.007	0.970	0.977	0.980	
BA	$\hat{\eta}(l)$	-0.509	0.205	-0.586	-0.430	-0.420	12	-2.424	0.324	-2.537	-2.434	-2.183	12
	$\hat{\gamma}_{ll}$	0.242	0.144	0.112	0.209	0.400	3	-2.256	0.420	-2.406	-2.115	-1.969	12
	$\widehat{SoR}(l)$	0.977	0.009	0.970	0.976	0.984		0.968	0.012	0.959	0.967	0.977	
CAT	$\hat{\eta}(l)$	-0.663	0.216	-0.726	-0.669	-0.508	12	-2.357	0.196	-2.550	-2.373	-2.195	12
	$\hat{\gamma}_{ll}$	0.297	0.465	0.058	0.290	0.665	5	-2.011	0.482	-2.133	-1.971	-1.792	11
	$\widehat{SoR}(l)$	0.973	0.018	0.967	0.978	0.981		0.965	0.021	0.956	0.970	0.975	
CSCO	$\hat{\eta}(l)$	-0.343	0.143	-0.404	-0.321	-0.242	11	-2.277	0.168	-2.451	-2.244	-2.130	12
	$\hat{\gamma}_{ll}$	6.749	6.700	1.438	2.194	14.563	10	-5.072	0.845	-5.576	-4.983	-4.416	12
	$\widehat{SoR}(l)$	0.990	0.007	0.990	0.992	0.995		0.987	0.009	0.986	0.990	0.992	
DD	$\hat{\eta}(l)$	-0.395	0.109	-0.493	-0.375	-0.323	12	-1.859	0.307	-1.912	-1.773	-1.709	12
	$\hat{\gamma}_{ll}$	0.405	0.247	0.269	0.340	0.625	6	-2.021	0.551	-2.463	-1.997	-1.507	12
	$\widehat{SoR}(l)$	0.979	0.006	0.975	0.978	0.984		0.971	0.007	0.967	0.971	0.976	
DIS	$\hat{\eta}(l)$	-0.561	0.179	-0.684	-0.582	-0.425	12	-2.705	0.221	-2.864	-2.729	-2.558	12
	$\hat{\gamma}_{ll}$	0.477	0.369	0.280	0.539	0.669	8	-2.303	0.277	-2.447	-2.284	-2.071	12
	$\widehat{SoR}(l)$	0.982	0.005	0.977	0.982	0.986		0.975	0.007	0.969	0.975	0.981	
GE	$\hat{\eta}(l)$	-0.711	0.272	-0.919	-0.705	-0.541	12	-2.508	0.240	-2.639	-2.506	-2.367	12
	$\hat{\gamma}_{ll}$	9.261	7.470	1.239	14.433	15.480	9	-5.708	3.903	-5.908	-4.590	-3.776	11
	$\widehat{SoR}(l)$	0.975	0.011	0.965	0.975	0.985		0.970	0.012	0.957	0.972	0.982	
GS	$\hat{\eta}(l)$	-0.746	0.507	-0.721	-0.588	-0.443	12	-2.303	0.195	-2.340	-2.275	-2.180	12
	$\hat{\gamma}_{ll}$	0.143	0.610	-0.197	0.010	0.210	3	-2.214	0.634	-2.615	-2.257	-1.759	11
	$\widehat{SoR}(l)$	0.972	0.018	0.972	0.975	0.982		0.965	0.020	0.964	0.967	0.976	
NKE	$\hat{\eta}(l)$	-0.651	0.170	-0.796	-0.631	-0.511	12	-2.601	0.434	-2.700	-2.612	-2.537	12
	$\hat{\gamma}_{ll}$	0.693	0.585	0.234	0.512	1.276	6	-2.751	0.846	-3.093	-2.444	-2.140	12
	$\widehat{SoR}(l)$	0.979	0.008	0.975	0.980	0.984		0.971	0.009	0.965	0.972	0.976	
TRV	$\hat{\eta}(l)$	-0.373	0.082	-0.436	-0.360	-0.313	12	-2.188	0.218	-2.249	-2.176	-2.037	12
	$\hat{\gamma}_{ll}$	0.426	0.221	0.261	0.398	0.529	6	-2.446	0.366	-2.768	-2.452	-2.158	12
	$\widehat{SoR}(l)$	0.984	0.004	0.981	0.984	0.986		0.977	0.006	0.974	0.976	0.980	
UNH	$\hat{\eta}(l)$	-0.536	0.155	-0.640	-0.496	-0.433	12	-2.424	0.257	-2.617	-2.401	-2.232	12
	$\hat{\gamma}_{ll}$	0.300	0.176	0.175	0.274	0.459	5	-2.369	0.462	-2.726	-2.441	-1.998	12
	$\widehat{SoR}(l)$	0.979	0.008	0.973	0.980	0.985		0.971	0.009	0.964	0.973	0.977	
V	$\hat{\eta}(l)$	-0.564	0.198	-0.729	-0.534	-0.401	12	-2.733	0.229	-2.906	-2.782	-2.521	12
	$\hat{\gamma}_{ll}$	0.699	0.312	0.480	0.678	0.852	9	-2.299	0.303	-2.539	-2.224	-2.103	12
	$\widehat{SoR}(l)$	0.987	0.002	0.986	0.987	0.988		0.981	0.002	0.980	0.981	0.983	
HD	$\hat{\eta}(l)$	-0.474	0.095	-0.539	-0.468	-0.426	12	-2.674	0.197	-2.831	-2.685	-2.530	12
	$\hat{\gamma}_{ll}$	0.406	0.207	0.275	0.368	0.510	6	-2.338	0.489	-2.738	-2.360	-2.034	12
	$\widehat{SoR}(l)$	0.983	0.005	0.979	0.984	0.987		0.976	0.007	0.971	0.978	0.982	
IBM	$\hat{\eta}(l)$	-0.458	0.161	-0.562	-0.430	-0.350	12	-2.506	0.175	-2.599	-2.530	-2.418	12
	$\hat{\gamma}_{ll}$	0.169	0.267	0.044	0.168	0.354	2	-2.461	0.540	-2.628	-2.428	-2.108	12
	$\widehat{SoR}(l)$	0.984	0.005	0.980	0.985	0.988		0.978	0.006	0.972	0.977	0.983	
JNJ	$\hat{\eta}(l)$	-0.534	0.196	-0.709	-0.550	-0.333	12	-2.795	0.200	-2.889	-2.774	-2.662	12
	$\hat{\gamma}_{ll}$	0.369	0.244	0.167	0.363	0.544	7	-1.825	0.195	-2.001	-1.848	-1.725	12
	$\widehat{SoR}(l)$	0.981	0.005	0.977	0.980	0.985		0.975	0.006	0.972	0.975	0.980	
KO	$\hat{\eta}(l)$	-0.474	0.194	-0.622	-0.440	-0.313	12	-2.252	0.241	-2.336	-2.241	-2.060	12
	$\hat{\gamma}_{ll}$	4.027	5.644	0.993	1.651	3.165	10	-3.686	0.562	-3.976	-3.575	-3.324	12
	$\widehat{SoR}(l)$	0.988	0.003	0.986	0.989	0.990		0.985	0.004	0.982	0.985	0.987	
MCD	$\hat{\eta}(l)$	-0.636	0.206	-0.680	-0.614	-0.531	12	-2.879	0.256	-3.106	-2.892	-2.684	12
	$\hat{\gamma}_{ll}$	0.397	0.264	0.228	0.436	0.545	7	-2.221	0.628	-2.272	-2.205	-1.846	12
	$\widehat{SoR}(l)$	0.979	0.007	0.978	0.979	0.984		0.971	0.009	0.968	0.971	0.978	
MMM	$\hat{\eta}(l)$	-0.392	0.171	-0.443	-0.381	-0.272	12	-2.359	0.364	-2.582	-2.192	-2.172	12
	$\hat{\gamma}_{ll}$	0.261	0.147	0.146	0.299	0.367	3	-2.549	0.403	-2.699	-2.510	-2.195	12
	$\widehat{SoR}(l)$	0.983	0.005	0.979	0.985	0.987		0.976	0.007	0.970	0.978	0.982	
MRK	$\hat{\eta}(l)$	-0.424	0.174	-0.572	-0.441	-0.282	11	-2.465	0.200	-2.653	-2.428	-2.321	12
	$\hat{\gamma}_{ll}$	0.628	0.382	0.331	0.646	0.933	6	-2.630	0.597	-2.987	-2.662	-2.166	12
	$\widehat{SoR}(l)$	0.986	0.006	0.981	0.986	0.992		0.980	0.008	0.974	0.982	0.987	
MSFT	$\hat{\eta}(l)$	-0.401	0.168	-0.513	-0.406	-0.283	11	-2.755	0.264	-2.874	-2.747	-2.556	12
	$\hat{\gamma}_{ll}$	3.255	5.683	0.548	0.828	1.583	7	-4.150	0.909	-4.891	-3.923	-3.436	12
	$\widehat{SoR}(l)$	0.992	0.003	0.990	0.991	0.993		0.990	0.003	0.987	0.989	0.991	
PG	$\hat{\eta}(l)$	-0.435	0.114	-0.526	-0.424	-0.338	12	-2.687	0.375	-2.971	-2.821	-2.320	12
	$\hat{\gamma}_{ll}$	0.666	0.382	0.420	0.801	0.904	9	-2.483	0.427	-2.732	-2.465	-2.380	12
	$\widehat{SoR}(l)$	0.983	0.008	0.976	0.985	0.989		0.978	0.009	0.969	0.979	0.983	
UTX	$\hat{\eta}(l)$	-0.568	0.293	-0.783	-0.568	-0.295	12	-2.342	0.334	-2.444	-2.362	-2.109	12
	$\hat{\gamma}_{ll}$	0.272	0.390	-0.024	0.323	0.553	7	-2.089	0.292	-2.271	-2.113	-1.865	12
	$\widehat{SoR}(l)$	0.985	0.005	0.981	0.986	0.988		0.980	0.007	0.975	0.978	0.983	
XOM	$\hat{\eta}(l)$	-0.656	0.274	-0.734	-0.669	-0.580	12	-2.962	0.326	-3.177	-2.895	-2.776	12
	$\hat{\gamma}_{ll}$	0.574	0.483	0.344	0.457	0.676	5	-2.444	0.368	-2.750	-2.421	-2.265	12
	$\widehat{SoR}(l)$	0.983	0.005	0.981	0.984	0.987		0.977	0.006	0.974	0.978	0.981	

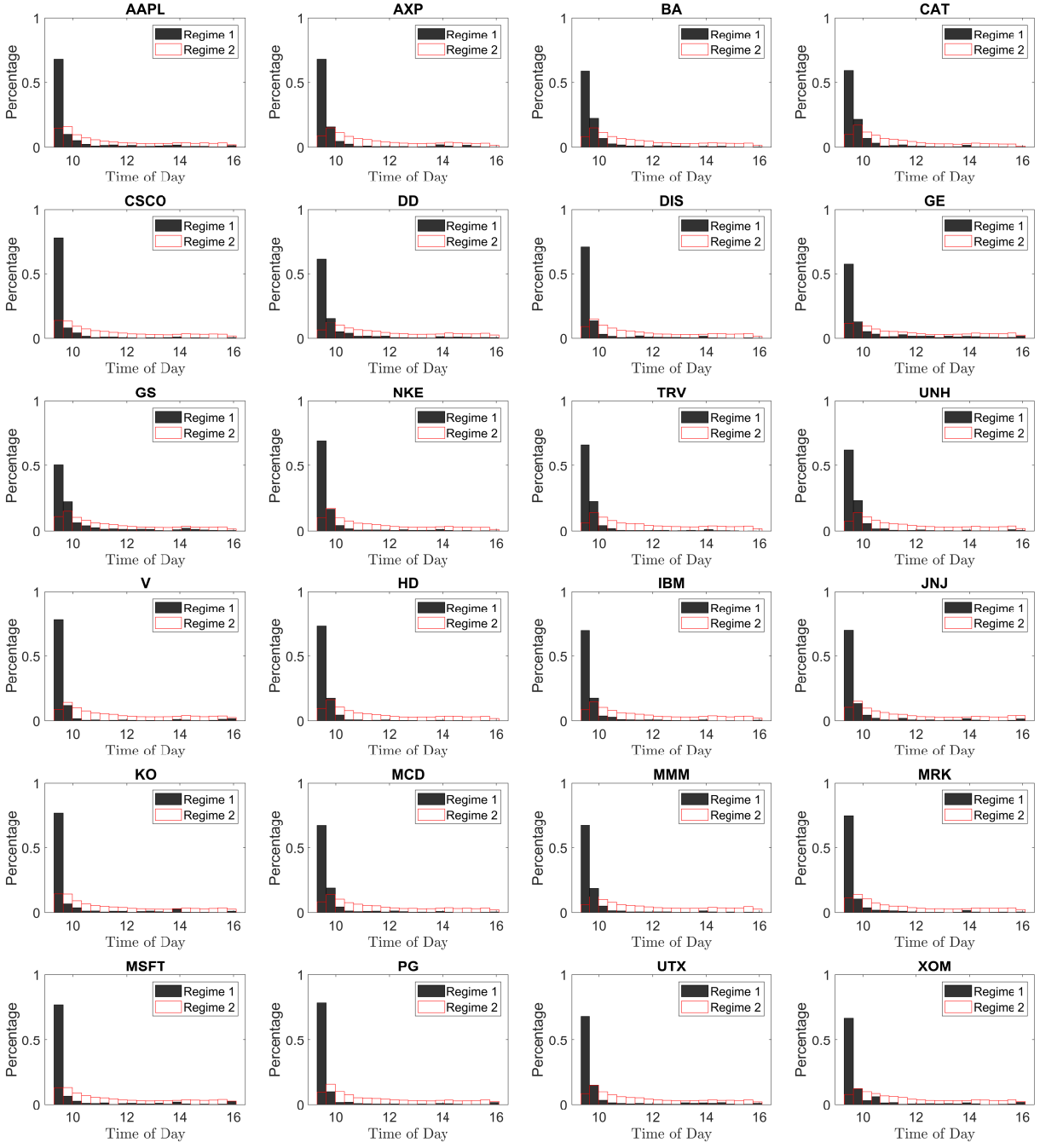
Note: The table reports descriptive statistics of monthly estimated $\hat{\eta}(l)$, $\hat{\gamma}_{ll}$ and $\widehat{SoR}(l)$ from the MS(2)-ACI(1,1) model estimated on the extended sample. The column #5% Sig. is the number of of significant parameter estimates out of the 12 monthly estimates.

Table 10: Annual averages of $\hat{s}_i$, $\tilde{x}_i^{(\delta)}$, $\ln \widetilde{Vol}_i$ and BAS_i for regimes 1 and 2 for the extended sample

	Regime 1				Regime 2			
Ticker	$\%(\hat{s}_i = 1)$	$\tilde{x}_i^{(\delta)}$	$\ln \widetilde{Vol}_i$	BAS_i	$\%(\hat{s}_i = 2)$	$\tilde{x}_i^{(\delta)}$	$\ln \widetilde{Vol}_i$	BAS_i
AAPL	0.0551	91.8817	12.0031	0.0370	0.9449	306.6248	12.3896	0.0119
AXP	0.1510	165.8976	9.6395	0.0622	0.8490	317.0809	10.1671	0.0151
BA	0.1493	177.2054	9.5152	0.1570	0.8507	315.4958	10.0664	0.0461
CAT	0.1297	179.4191	9.7900	0.0766	0.8703	309.9714	10.2893	0.0231
CSCO	0.0949	89.7579	11.1721	0.0264	0.9051	330.0185	11.7716	0.0102
DD	0.1352	153.4883	9.1055	0.0924	0.8648	320.3704	9.6237	0.0210
DIS	0.1352	151.6551	10.2023	0.0689	0.8648	315.9966	10.6810	0.0175
GE	0.1243	154.0692	11.7078	0.0144	0.8757	340.0005	12.1774	0.0102
GS	0.1165	173.5606	9.4959	0.2211	0.8835	314.3397	10.0019	0.0863
NKE	0.1533	164.3781	10.4660	0.0409	0.8467	314.1945	10.8026	0.0137
TRV	0.1445	180.1966	8.5116	0.1891	0.8555	315.7850	9.0409	0.0384
UNH	0.1592	181.8590	9.3980	0.1899	0.8408	317.1635	9.7408	0.0483
V	0.1186	132.5547	10.2186	0.0755	0.8814	318.7315	10.7278	0.0150
HD	0.1371	171.6699	9.7693	0.1321	0.8629	317.6330	10.2090	0.0330
IBM	0.1101	146.5353	9.3814	0.1916	0.8899	314.3452	9.9763	0.0479
JNJ	0.1393	134.8873	9.8100	0.0924	0.8607	325.6155	10.4627	0.0192
KO	0.1095	93.1248	10.3803	0.0305	0.8905	331.1631	11.0841	0.0106
MCD	0.1636	166.7272	9.8791	0.0973	0.8364	319.8651	10.2370	0.0238
MMM	0.1666	159.9102	8.8182	0.2284	0.8334	323.3118	9.2726	0.0523
MRK	0.1119	120.1513	10.1628	0.0617	0.8881	321.2483	10.7851	0.0126
MSFT	0.0694	69.3481	11.4449	0.0393	0.9306	317.6993	12.0264	0.0105
PG	0.1379	137.6501	10.2305	0.0667	0.8621	325.4344	10.7287	0.0136
UTX	0.1337	158.3380	9.3806	0.1224	0.8663	318.1102	9.9181	0.0277
XOM	0.1228	156.3153	10.6496	0.0570	0.8772	319.5497	11.0959	0.0138

Note: in this table we present the sample mean of regime-specific s_i, $\tilde{x}_i^{(\delta)}$, $\ln \widetilde{Vol}_i$ and BAS_i for the entire sampling period, where the regimes are estimated monthly from the MS(2)-ACI(1,1) model as in Tables 9. The column $\%(\hat{s}_i = l)$ computes the annual percentage of data being classified into regime l.

Figure 5: Time-of-day distribution of regimes for the extended sample



Note: The figure plots the histogram of the time of day associated with observations in regimes 1 and 2, where the regimes are estimated monthly based on the MS(2)-ACI(1,1) model for the extended sample. The y-axis is the percentage of regime-specific observations that fall in one of the 20 equidistant time intervals.

Table 11: Descriptive statistics of $\hat{b}_1(l)$ and $R^2(l)$ for regimes 1 and 2

	Mean	Std. Dev.	Q(25%)	Median	Q(75%)
Extended Sample Only					
$\hat{b}_1(1)$	0.4163	0.1174	0.3501	0.4109	0.4838
$\hat{b}_1(2)$	0.8042	0.0503	0.7719	0.8045	0.8410
$R^2(1)$	0.3118	0.1231	0.2308	0.3154	0.3831
$R^2(2)$	0.6745	0.0740	0.6271	0.6844	0.7311
Extended Sample + Individual Stocks of Main Sample					
$\hat{b}_1(1)$	0.4304	0.1198	0.3599	0.4240	0.4996
$\hat{b}_1(2)$	0.8030	0.0495	0.7724	0.8045	0.8381
$R^2(1)$	0.3107	0.1212	0.2325	0.3076	0.3914
$R^2(2)$	0.6735	0.0718	0.6310	0.6822	0.7256

Note: the table presents descriptive statistics of estimated $\hat{b}_1(l)$ and $R^2(l)$ from $\ln \widetilde{Vol}_i = b_0(l) + b_1(l) \ln \tilde{x}_i^{(\delta)} + u_i$ with $l = 1, 2$ for each stock-month. The 'extended sample only' panel presents parameter estimates based on the 24 assets in the extended sample, while the 'extended sample + individual stocks of main sample' also include the 9 individual stocks in the main sample. The regime classification is estimated monthly by the MS(2)-ACI(1,1) model. Results for SPY are excluded in the 'All Individual Stocks' panel and are presented separately in the 'SPY Only' panel. $Q(x\%)$ is the $x\%$ quantile.

References

Hong, S. Y., Nolte, I., Taylor, S., and Zhao, X. (2020). Volatility Estimation and Forecasts Based on Price Durations. *Working Paper, Lancaster University Management School.*