Granger Centre Discussion Paper Series

Dynamic discrete mixtures for high frequency prices

by

Leopoldo Catania, Roberto Di Mari and
Paolo Santucci de Magistris

Granger Centre Discussion Paper No. 19/05

Dynamic Discrete Mixtures for High Frequency Prices*

Leopoldo Catania† Roberto Di Mari ‡ Paolo Santucci de Magistris §

March 8, 2019

Abstract

The tick structure of the financial markets entails that price changes observed at very high frequency are discrete. Departing from this empirical evidence we develop a new model to describe the dynamic properties of multivariate time-series of high frequency price changes, including the high probability of observing no variations (*price staleness*). We assume the existence of two independent latent/hidden Markov processes determining the dynamic properties of the price changes and the excess probability of the occurrence of zeros. We study the probabilistic properties of the model that generates a zero-inflated mixture of Skellam distributions and we develop an EM estimation procedure with closed-form M step. In the empirical application, we study the joint distribution of the price changes of four assets traded on NYSE. Particular focus is dedicated to the precision of the univariate and multivariate density forecasts, to the quality of the predictions of quantities like the volatility and correlations across assets, and to the possibility of disentangling the different sources of zero price variation as generated by absence of news, microstructural frictions or by the offsetting positions taken by the traders.

Keywords: Dynamic Mixtures; Skellam Distribution; Zero-inflated series; EM Algorithm; High frequency prices; Volatility.

JEL Classification: C38; C60; G13.

*We are grateful to Umberto Cherubini, Christian Hennig, Dimitris Karlis, Alessandra Luati, André Lucas, Francesco Ravazzolo, Roberto Renò, and Jean-Michel Zakoïan for useful comments and suggestions that improved the quality of this work. We also thank Kim Christensen for providing us with the data used in the empirical analysis. The authors would also like to thank the participants of the Frontiers in High Frequency Financial Econometrics Workshop (HFFE) in Pisa 2018, the DEDA conference 2018 in Xiamen, the MBC^2 conference in Catania, the CFE 2018 conference in Pisa, and the ICEEE conference in Lecce as well as seminar participants at University of Bologna and CREST for helpful comments. Finally, we also thank Annia Hoffmeyer for a careful proofreading of the paper. Paolo Santucci de Magistris and Leopoldo Catania acknowledge the research support of the Project 2 grant of the Danish Council for Independent Research (IRFD), Social Sciences, number 8019-00015A. Paolo Santucci de Magistris and Leopoldo Catania also acknowledge the research support of AUFF and CREATES, funded by the Danish National Research Foundation (DNRF78).

†Department of Economics and Business Economics, Aarhus University and CREATES, Fuglesangs Alle 4, 8210, Aarhus V, Denmark. lcatania@econ.au.dk.

‡Department of Economics and Business, University of Catania, Italy, roberto.dimari@unict.it.

§Department of Economics and Finance, LUISS "Guido Carli" University. Viale Romania 32, 00197 Roma, Italy. CREATES, Aarhus University, Fuglesangs Alle 4, 8210 Aarhus V, Denmark, sdemagistris@luiss.it.

1 Introduction

In the last twenty years, we have witnessed a boost in the intradaily trading activity on the financial markets and, subsequently, an enormous increase in the availability of stock prices observed at high frequency. On the one hand, the availability of stock prices sampled at very high frequency has driven the empirical analysis of financial markets towards the use of ex-post measurements of (integrated) variance over fixed horizons (e.g. day), see the discussion in Andersen et al. (2010). On the other hand, prices sampled at very high-frequencies are characterized by a number of micro-structural features, which challenge the adequacy of the standard specifications typically adopted for the intradaily price changes. This opens the door to alternative model specifications for the high frequency price moves.

We contribute to this strand of literature by providing a new statistical framework for the analysis of high frequency prices, which goes beyond the standard setups. In the classic framework, the prices of financial assets are typically assumed to originate from a continuous distribution with time-varying parameters, e.g. with stochastic volatility, see Shephard (2005). The reason for the widely adopted assumption of a continuous underlying price process is made to increase model tractability. However, financial markets regulations make stock price changes intrinsically discrete due to the minimum allowed tick size (also known as decimalization effect). This discreteness becomes more evident when the sampling frequency increases. For instance, at the frequency of one second the discreteness of the price changes is the dominating feature, see also the discussion in Rossi and Santucci de Magistris (2018) on the impact of price discreteness on the inference on stochastic volatility parameters. The statistical analysis of discrete processes in $\mathbb{Z}$ poses substantial difficulties from a methodological viewpoint, greatly complicating the underlying theory and model interpretation, see among others the recent contributions of Koopman et al. (2017) for a discrete-time model and Shephard and Yang (2017) for a model built in continuous-time. Along with their intrinsic discreteness, other stylized facts of high frequency price changes are: i) the strong presence of time-varying volatility, see Koopman et al. (2017); ii) the large number of zero price variations, which is reflected in a constant price over a short time interval, a feature known as *price staleness*, see Bandi et al. (2017); iii) the presence of extreme observations

(fat tailed conditional distribution). When the goal is to carry out inference on volatility and correlations across stocks, all these features must be accounted for.

In this paper, we develop a flexible multivariate integer valued model for the analysis of price changes observed at high frequency. The goal is to answer the natural question on how prices of different assets observed at high frequency interrelate, when we incorporate in the model all the features discussed above. The model builds upon a simple mechanism for the generation of the transaction price changes that result from the realization of the difference between two unobserved random variables accounting for positive and negative moves. Since the price changes can only take values on a discrete grid, these two random variables must adhere to this constraint. The Skellam distribution of Irwin (1937) and Skellam (1946), which arises from the difference between two independent Poisson random variables, provides the natural baseline framework to model discrete price changes, see also Barndorff-Nielsen et al. (2012) and Koopman et al. (2017). In particular, Koopman et al. (2017) assume that the price changes of the individual assets traded on NYSE are conditionally distributed according to a Skellam distribution with stochastic volatility. The resulting model can be cast in the class of nonlinear non-Gaussian state space models for which the likelihood is not analytically available. This results in complicated inference and non-standard estimation procedures; Koopman et al. (2017) use simulated maximum likelihood relying on the numerically accelerated importance sampling (NAIS) method, see Koopman et al. (2015). An extension to the multivariate context within the framework of Koopman et al. (2017) is unfeasible. Indeed, like the multivariate Poisson case, the resulting multivariate Skellam distribution (see Bulla et al. (2015) and Akpoue and Angers (2017) for the iid case) is remarkably difficult to handle. Furthermore, multivariate models of this kind suffer from the "curse of dimensionality" and have been rarely applied to the case of more than two variables. A notable application of the Skellam in the context of multivariate prices changes based on a copula function and a generalized autoregressive score (GAS) specification is provided in Koopman et al. (2018).

Differently from the previous approaches, our modeling framework builds upon the idea that the observed price changes are independent conditionally to the realization of unobserved discrete-valued random variables describing the multivariate dynamic properties of

the data. In other words, our model belongs to the class of latent/hidden Markov models (HMM), see among others Vermunt et al. (1999), Bartolucci and Farcomeni (2009), Bartolucci et al. (2012) and Zucchini et al. (2017). The Markov structure is made up of two independent Markov chains responsible to capture the dynamics of the price changes and their mutual association as well as the excess probability of zeros. Conditional on the latent structure, each individual asset is assumed to be Skellam distributed and independent from other assets. The distribution of the observables, after marginalization of the latent variables, typically exhibits time-varying volatility, fat tails, excess of zeros, and time-varying probability of zero observations.

The large fraction of zero observations displaying significant autocorrelation has a substantial micro-structural interpretation; Bandi et al. (2017) measure the extent of staleness in high frequency prices by developing a novel financial indicator: the excess idle time (EXIT). Their results show that the probability of observing zero price changes in financial data varies over time and it is an indicator of price illiquidity. Similar studies addressing the time-varying probability of observing zero variations for integer valued time-series (mostly in $\mathbb{N}$) can be found in Zeger and Brookmeyer (1986), Rydberg and Shephard (2003), Bien et al. (2011), Hautsch et al. (2013), Kömm and Küsters (2015), Grønneberg and Sucarrat (2017) and Harvey and Ito (2017). However, none of these (univariate) models contemporaneously incorporates all the stylized facts characterizing the financial price changes observed at high frequency. From both a methodological and an applied perspective, allowing for multivariate dependencies between the zeros of several equities can bring substantial insights on illiquidity features within and between assets, which is essential to study commonalities during illiquid episodes.

Through the paper, we show that our latent structure has an alternative representation in terms of a single hidden Markov chain with suitable constraints, and the re-parametrization is one-to-one. This allows us to prove that the model is identifiable thus resorting to maximum likelihood estimation with no exceptional effort by means of an expectation-maximization (EM) algorithm with steps available in closed form. This is an extension of the EM algorithm proposed in Catania and Di Mari (2018) for ML estimation of a hierarchical Markov-switching model for multivariate count data (for a similar hierarchical structure for univari-

ate data, see also Bartolucci and Farcomeni (2015), Geweke and Amisano (2011), Maruotti (2011), and Maruotti and Rydén (2009), for an application in the context of longitudinal data). In addition, the hidden Markov representation of our model allows us to analytically derive the predictive, filtered, and smoothed distributions of the latent variables as well as the joint predictive distribution of price changes.

Our empirical results can be summarized as follows: the proposed modelling framework well adapts to match the univariate and multivariate empirical distributions of high frequency price changes and their associated moments. This holds true in both the normal and Lehman market periods under investigation; the latter being characterized by abnormal price variations especially at the opening of the trading day. The model well accounts for all the empirical features displayed by the high frequency data, including the large proportion of zero price variations (staleness), which often occur simultaneously on multiple assets. Furthermore, we show that disentangling the conditional probability of zeros into could be employed to predict and interpret the reduced trading activity on the markets as measured by absence of volume of transactions in certain phases of the trading day.

The paper is organized as follows: Section 2 presents the model and its stochastic properties. Section 3 discusses inference via the expectation-maximization algorithm. Section 4 outlines the empirical application, and Section 5 concludes. Finally, a document with supplementary material reports additional results concerning the empirical application.

2 Model

Let $Y_{n,t} \in \mathbb{Z}$ be the random variable representing the price change of asset n at time t, and let $y_{n,t}$ be its realization. We collect the price changes of N assets in the $N \times 1$ vector $\mathbf{Y}_t = (Y_{n,t}, n = 1, \ldots, N)' \in \mathbb{Z}^N$, with analogous notation for $\mathbf{y}_t = (y_{n,t}, n = 1, \ldots, N)'$. $\mathbf{Y}_t$ is assumed to be observed for each time point[1]. Let S_t^ω and S_t^κ be two unobserved independent homogeneous stationary first order Markov chains with finite state space $S_t^\omega \in \{1, \ldots, J\}$ and $S_t^\kappa \in \{1, \ldots, L\}$. Let also $P(S_t^\omega = i | S_{t-1}^\omega = j) = \gamma_{i,j}^\omega$ for $i, j = 1, \ldots, J$ and $P(S_t^\kappa = h | S_{t-1}^\kappa = l) = \gamma_{h,l}^\kappa$ for $h, l = 1, \ldots, L$ be the transition probabilities of the two Markov chains

[1]Note that $\mathbf{Y}_t$ can also contain missing values. Accounting for missing values is straightforward given the latent structure of our model as discussed below.

S_t^ω and S_t^κ, respectively. The transition probabilities of S_t^ω are collected in the $J \times J$ matrix $\Gamma^\omega = [\gamma_{i,j}^\omega]_{i,j=1}^J$ and S_t^κ in the $L \times L$ matrix $\Gamma^\kappa = [\gamma_{i,j}^\kappa]_{i,j=1}^L$, under the usual constraints on the positiveness and summability: $\gamma_{i,j}^c > 0$; $\Gamma^c \mathbf{u} = \mathbf{u}$, for $c = \omega, \kappa$, with $\mathbf{u}$ being a vector of ones of proper dimension. The stationary distributions of the two Markov chains are indicated by $\boldsymbol{\delta}^\omega = (\delta_j^\omega, j = 1, \ldots, J)'$ and $\boldsymbol{\delta}^\kappa = (\delta_h^\kappa, h = 1, \ldots, K)'$ for S_t^ω and S_t^κ, respectively.[2]

We allow for a second hidden layer which, conditional on the realization of S_t^ω, handles time-specific dependencies across the assets and accommodates departures from the marginal distributions assumed for each asset. We label this additional layer Z_t, which is an unobserved integer-valued random variable with state space $\{1, \ldots, K\}$. The variable Z_t is assumed to be independent from S_s^κ and $\mathbf{Y}_s$, given S_t^ω for all s. Specifically, we assume that $Z_t | S_t^\omega \perp\!\!\!\perp (S_s^\kappa, \mathbf{Y}_s, S_g^\omega)$ for all s and $g \neq t$. Furthermore, $Z_t | S_t^\omega$ is assumed to be categorically distributed with $P(Z_t = k | S_t^\omega = j) = \omega_{j,k}$, with $\omega_{j,k} > 0$ and $\sum_{l=1}^K \omega_{j,l} = 1$ for all $j = 1, \ldots, J$ and $k = 1, \ldots, K$. In other words, the joint latent process $(Z_t; S_t^\omega)$ can be decomposed into an unconditional Markovian state process S_t^ω that handles serial dependence in the data, and another process, Z_t, that conditionally on S_t^ω handles cross dependencies. Analogously, we let $B_{n,t}$ be an unobserved Boolean random variable with state space $\{0, 1\}$, which inflates the probability of observing a zero for the price change of asset n at time t. All the $B_{n,t}$ are collected in the vector $\mathbf{B}_t = (B_{n,t}, n = 1, \ldots, N)' \in [0, 1]^N$. We assume that, given S_t^κ, all the $B_{n,t}$ for $n = 1, \ldots, N$, are independent between each other and from S_s^ω and Z_s for all s, that is $B_{n,t} | S_t^\kappa \perp\!\!\!\perp (B_{m,t}, S_g^\kappa, S_s^\omega, Z_s)$ for all s, $n \neq m$, and $g \neq t$. We further assume that $P(B_{n,t} = 1 | S_t^\kappa = l) = \kappa_{n,l}$, where $0 < \kappa_{n,l} < 1$ for all $n = 1, \ldots, N$ and $l = 1, \ldots, L$. The aforementioned dependence structure is reported in Figure 1.

The observed random variables $Y_{n,t}$ are assumed to be independently Skellam distributed given Z_t and $B_{n,t} = 0$: $Y_{n,t} | (Z_t, B_{n,t} = 0) \perp\!\!\!\perp Y_{m,t} | Z_t, B_{m,t=0}$ for all $n \neq m$. Furthermore, $\mathbf{Y}_t | Z_t, B_{n,t} = 0$ is assumed to be independent from S_s^ω, S_s^κ and B_s for all s. In the following we exploit the stochastic representation of a Skellam random variable as the difference of two independent Poisson distributed random variables. Specifically, we have that

$$Y_{n,t} | (Z_t, B_{n,t} = 0) = X_{n,t}^{(1)} | Z_t - X_{n,t}^{(2)} | Z_t, \tag{1}$$

[2]By the stationarity assumption, the initial distribution of each Markov chain is set equal to the stationary distribution.

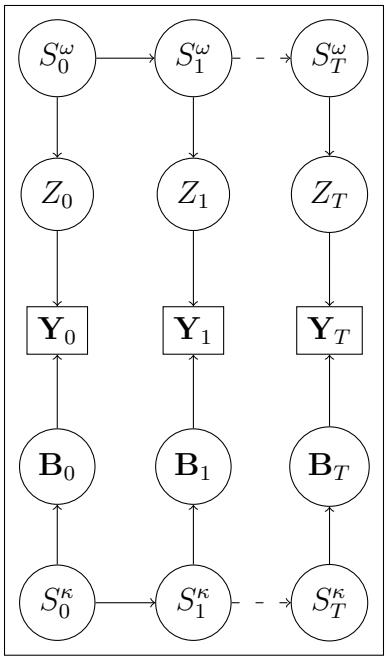


Figure 1: The model path diagram. S_t and Z_t are two integer–valued unobserved stochastic variables. S_t follows a first order Markov process with state space $\{1, \ldots, J\}$, while Z_t is independently and identically distributed given S_t, with support $\{1, \ldots, K\}$. $\mathbf{Y}_t$ is a multivariate observed integer–valued random variable, which is independently and identically distributed given S_t and Z_t.

where $X_{n,t}^{(1)}|Z_t = k$ and $X_{n,t}^{(2)}|Z_t = k$ are two independent Poisson random variables with intensity parameters $\lambda_{n,k}^{(1)}$ and $\lambda_{n,k}^{(2)}$, respectively. The probability mass function of $Y_{n,t}|(Z_t, B_{n,t} = 0)$ is given by

$$P(Y_{n,t} = y|Z_t = k, B_{n,t} = 0) = e^{-(\lambda_{n,k}^{(1)} + \lambda_{n,k}^{(2)})} \mathcal{I}_y\left(2\sqrt{\lambda_{n,k}^{(1)}\lambda_{n,k}^{(2)}}\right)\left(\frac{\lambda_{n,k}^{(1)}}{\lambda_{n,k}^{(2)}}\right)^{y/2}, \tag{2}$$

where

$$\mathcal{I}_y(a) = \left(\frac{1}{2}a\right)^y \sum_{r=0}^{\infty} \frac{\left(\frac{1}{4}a^2\right)^r}{r!\Gamma(y + r + 1)},$$

is the modified Bessel function of the first kind, and $\Gamma(\cdot)$ is the Gamma function. By conditioning on the event $B_{n,t} = 1$, we assume that $P(Y_{n,t} = y|Z_t = k, B_{n,t} = 1) = P(Y_{n,t} = y|B_{n,t} = 1) = \psi(y)$ where

$$\psi(y) = \begin{cases} 1, & \text{if } y = 0 \\ 0, & \text{otherwise,} \end{cases}$$

is a Dirac mass at 0. By removing the conditioning on $B_{n,t} = 0$, we obtain the following

stochastic representation

$$Y_{n,t}|Z_t = (1 - B_{n,t}) \left(X^{(1)}_{n,t}|Z_t - X^{(2)}_{n,t}|Z_t \right) .$$

Marginalizing out the effect of $B_{n,t}$ and conditioning on the event $S^{\kappa}_t = j$, we recover a zero inflated Skellam distribution with probability mass function

$$P(Y_{n,t} = y|Z_t = k, S^{\kappa}_t = l) = \kappa_{n,l}\psi(y) + (1 - \kappa_{n,l})\mathcal{SK}(y, \lambda^{(1)}_{n,k}, \lambda^{(2)}_{n,k}) ,$$

where $\mathcal{SK}(y, \lambda^{(1)}_{n,k}, \lambda^{(2)}_{n,k}) = P(Y_{n,t} = y|Z_t = k, B_{n,t} = 0)$ is reported in (2). Under the conditional independence assumption for the $Y_{n,t}$, the joint probability mass function of $\mathbf{Y}_t$ is given by

$$P(\mathbf{Y}_t = \mathbf{y}|Z_t = k, S^{\kappa}_t = l) = \prod_{n=1}^{N} P(Y_{n,t} = y_n|Z_t = k, S^{\kappa}_t = l).$$

The marginalization of Z_t is achieved by conditioning on S^{ω}_t as follows

$$P(\mathbf{Y}_t = \mathbf{y}|S^{\omega}_t = j, S^{\kappa}_t = l) = \sum_{k=1}^{K} \omega_{j,k} \prod_{n=1}^{N} P(Y_{n,t} = y_i|Z_t = k, S^{\kappa}_t = l).$$

Finally, after marginalization of the two Markov chains, we recover the unconditional distribution of $\mathbf{Y}_t$ as

$$P(\mathbf{Y}_t = \mathbf{y}) = \sum_{j=1}^{J}\sum_{l=1}^{L}\sum_{k=1}^{K}\prod_{n=1}^{N} \delta^{\omega}_j \delta^{\kappa}_l \omega_{j,k} \left(\kappa_{n,l}\psi(y_n) + (1 - \kappa_{n,l})\mathcal{SK}(y, \lambda^{(1)}_{n,k}, \lambda^{(2)}_{n,k}) \right) ,$$

from which we recognize a three layer mixture of conditionally independent zero inflated Skellam distributions. We label this model *Dynamic Mixture of Skellam*, DMS, henceforth. If we wish to condition to past values of $\mathbf{Y}_t$, the distribution simply reduces to

$$P(\mathbf{Y}_t = \mathbf{y}|\mathbf{Y}_{1:t-s}) = \sum_{j=1}^{J}\sum_{l=1}^{L}\sum_{k=1}^{K}\prod_{n=1}^{N} \pi^{\omega}_{j,t|s}\pi^{\kappa}_{l,t|s}\omega_{j,k} \left(\kappa_{n,l}\psi(y_n) + (1 - \kappa_{n,l})\mathcal{SK}(y_n, \lambda^{(1)}_{n,k}, \lambda^{(2)}_{n,k}) \right) , \tag{3}$$

for $s > 0$, with

$$\pi^{\omega}_{h,t|s} = \frac{\sum_{i=1}^{J}[(\Gamma^{\omega})^s]_{ij}\alpha^{\omega}_{i,t-s}}{P(\mathbf{Y}_{1:t-s} = \mathbf{y}_{1:t-s})}, \quad \pi^{\kappa}_{h,t|s} = \frac{\sum_{i=1}^{K}[(\Gamma^{\kappa})^s]_{ij}\alpha^{\kappa}_{i,t-s}}{P(\mathbf{Y}_{1:t-s} = \mathbf{y}_{1:t-s})},$$

where $\pi^{\omega}_{h,t|s} := P(S^{\omega}_t = h|\mathbf{Y}_{1:t-s} = \mathbf{y}_{1:t-s})$ and $\pi^{\kappa}_{h,t|s} := P(S^{\kappa}_t = h|\mathbf{Y}_{1:t-s} = \mathbf{y}_{1:t-s})$ represent the predictive distribution of S^c_t in state h, and $\alpha^c_{i,t} = P(S^c_t = h, \mathbf{Y}_{1:t} = \mathbf{y}_{1:t})$ for $c = \omega, \kappa$ are the forward probabilities delivered by the forward filtering backward smoothing (FFBS) algorithm; more details are provided in Section 3. The notation $[(\Gamma^c)^s]_{ij}$ indicates the ij-th element of the power s of the matrix Γ^c.

2.1 Equivalent formulations

The model formulation reported in Figure 1 is convenient to easily determine the dependence structure underlying the DMS. However, the tasks of filtering and smoothing typically required in the estimation of the model turn out to be rather involved due to the presence of the two unobserved Markov chains and the additional latent variables (Z_t and $B_{n,t}$). Therefore, we present an equivalent model representation that allows us to adapt all the statistical tools developed for general HMM with discrete support to the present framework. For instance, the equivalent model representation entails that filtering and smoothing of the latent chains can be computed by the FFBS algorithm (for technical details, see for instance Frühwirth-Schnatter (2006)).

We proceed by defining a new stationary first order homogeneous Markov chain $S^{\omega,Z}_t$ with state space $\{1, \ldots, JK\}$. The Markov chain $S^{\omega,Z}_t$ is defined by combining the Markov chain S^{ω}_t and the integer random variable Z_t. In addition, let $\mathbf{\Omega} = \omega_{k,j}$ be a $K \times J$ matrix containing the mixture probabilities and let $\mathbf{U}$ and $\mathbf{u}$ be respectively a $K \times K$ matrix and a JK-vector of ones. The transition probability matrix of $\Gamma^{\omega,Z}$ is given by

$$\Gamma^{\omega,Z} = \mathbf{u}\text{vec}(\mathbf{\Omega})' \odot (\Gamma^{\omega} \otimes \mathbf{U}), \tag{4}$$

where $\odot$ and $\otimes$ are the Hadamard and Kronecker products, respectively. By incorporating Z_t in S^{ω}_t via the new Markov chain $S^{\omega,Z}_t$, we obtain the equivalent model representation reported in Figure 2. The term $S^{\omega,Z}_t$ is still a homogeneous stationary first order Markov chain; however, its state space has been enlarged compared to that of S^{ω}_t, and its transition probability matrix has a constrained structure provided by (4). To the constrained structure imposed to the transition probability matrix corresponds a unique ordering in the conditional

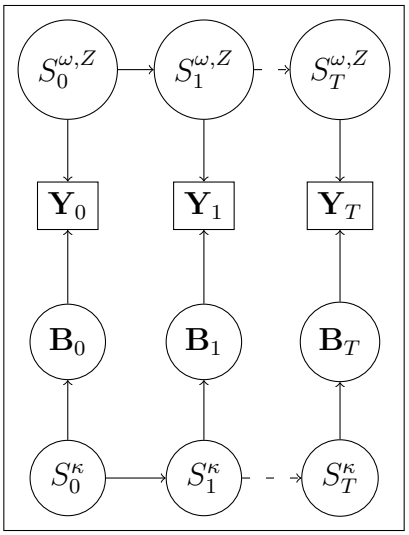


Figure 2: The model path diagram where Z_t is incorporated in S_t^{ω}, resulting in $S_t^{\omega,Z}$.

densities of $Y_t|S_t^{\omega,Z}$. Specifically, let $i^{\omega} \in \{1,\ldots,J\}$ and $j^{Z} \in \{1,\ldots,K\}$ be two indexes spanning over the state space of S_t^{ω} and Z_t, respectively. The state space of $S_t^{\omega,Z}$ can be represented as $\{1^{\omega}1^{Z},\ 1^{\omega}2^{Z},\ \ldots,\ 1^{\omega}K^{Z},\ 2^{\omega}1^{Z},\ \ldots,\ i^{\omega}j^{Z},\ \ldots,\ J^{\omega}K^{Z}\}$, i.e. the conditional densities of the first K regimes of the $S_t^{\omega,Z}$ Markov chain are the K mixture components of the first regime of the Markov chain S_t^{ω}, from $K+1$ to $2K$ of the second regime, and so on.

Furthermore, by combining the two independent Markov chains $S_t^{\omega,Z}$ and S_t^{B} in a third Markov chain $S_t^{\omega,Z,B}$ with state space $\{1,\ldots,JKB\}$ and transition probability matrix $\Gamma^{\omega,Z,B} = \Gamma^{B} \otimes \Gamma^{\omega,Z}$, we obtain a representation of the DMS in terms of a single Markov chain. As for the previous representation, the state space of $S_t^{\omega,Z,B}$ can be represented as $\{1^{\omega}1^{Z}1^{B},\ 1^{\omega}2^{Z}1^{B},\ldots,\ 1^{\omega}K^{Z}1^{B},\ 2^{\omega}1^{Z}1^{B},\ldots,\ 1^{\omega}1^{Z}2^{B},\ i^{\omega}j^{Z}h^{B},\ldots,\ J^{\omega}K^{Z}L^{B}\}$, where $h^{B} \in \{1,\ldots,L\}$ spans the state space of S_t^{B}. Let us denote this set of indexes $\mathcal{R}$ and define the subsets of indexes for which $Z_t = k$ as $\mathcal{R}(Z_t = k)$, $S_t^{\omega} = j$ as $\mathcal{R}(S_t^{\omega} = j)$, and $S_t^{B} = l$ as $\mathcal{R}(S_t^{B} = l)$. For example, let $J = 2$, $K = 3$, and $L = 4$, in this case the number of regimes in the $S_t^{\omega,Z,B}$ Markov chain is 24, and the state space of $S_t^{\omega,Z,B}$ can be represented in as

$$\begin{aligned}\mathcal{R} = \{&111, 121, 131, 211, 221, 231,\\ &112, 122, 132, 212, 222, 232,\\ &113, 123, 133, 213, 223, 233,\\ &114, 124, 134, 214, 224, 234\}.\end{aligned}$$

For instance, if $q = 6$, the corresponding label is 231, and we have $q_1 = 2$, $q_2 = 3$, and $q_3 = 1$. In this case, the subset $\mathcal{R}(Z_t = k)$ is given by the indexes of the enlarged state space for which $Z_t = k$. For example, if $k = 2$, we have that $\mathcal{R}(Z_t = 2) = \{2, 5, 8, 11, 14, 17, 20, 23\}$. The representation associated with $S_t^{\omega,Z,B}$ is reported in Figure 3. The path diagram displayed

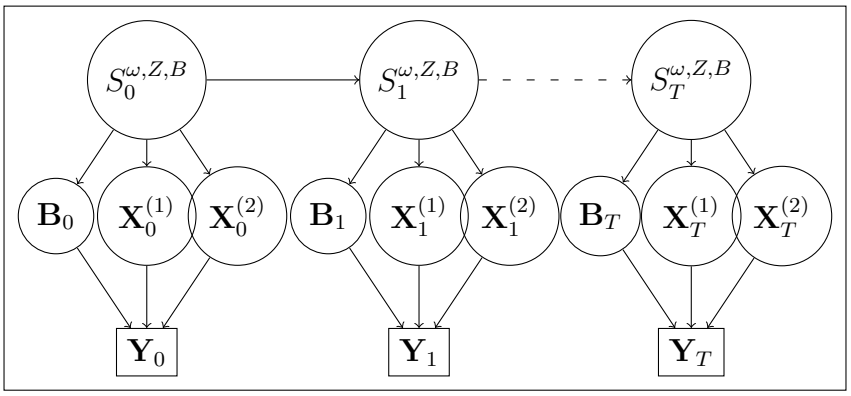


Figure 3: The model path diagram with one chain, $S_t^{\omega,Z,B}$.

in Figure 3 is that of a standard HMM for which the conditional distribution of $Y_{n,t}$ is given by a zero inflated Skellam distribution. Specifically, conditional on $S_t^{\omega,Z,B} = q$, we denote the probability mass function of $\mathbf{Y}_t$ as

$$P(\mathbf{Y}_t = \mathbf{y}|S_t^{\omega,Z,B} = q) = \prod_{n=1}^{N} \kappa_{q_3,n}\psi(y_n) + (1 - \kappa_{q_3,n})\mathcal{SK}(y, \lambda_{n,q_2}^{(1)}, \lambda_{n,q_2}^{(2)}), \tag{5}$$

which is adopted for filtering and smoothing of the latent states.

2.2 Identifiability of the DMS model

The fact that the DMS maps into a simple hidden Markov structure allows us to adopt a number of results on model identification that are standard in the hidden Markov literature. Identification is proven under the following classic set of assumptions: (A1) S_t^{ω} and S_t^{κ} are irreducible, (A2) the rows of $\mathbf{\Omega}$ are linearly independent, (A3) $\kappa_{n,l_1} \neq \kappa_{n,l_2}$ and $(\lambda_{n,j_1}^{(1)}, \lambda_{n,j_1}^{(2)}) \neq (\lambda_{n,j_2}^{(1)}, \lambda_{n,j_2}^{(2)})$ for all n, $l_1 \neq l_2$, and $j_1 \neq j_2$. The following proposition establishes the identification of the DMS model.

Proposition 2.1 (Identification)**.** *Given Assumptions (A1)–(A3) and provided that $K \geq J$ the DMS model is identified up to label swapping.*

The proof of Proposition 2.1 is made exploiting Theorem 1 and Proposition 2 of Gassiat et al. (2016). Specifically, consider a further reparametrization in which we let $\mathbf{\Gamma}^{\omega,B} = \mathbf{\Gamma}^{\omega} \otimes \mathbf{\Gamma}^{B}$ be the transition probability matrix related to the homogeneous stationary first order Markov chain $S_t^{\omega,B} = \{S_t^{\omega}, S_t^{B}\}$, with state space $\{1, \ldots, JL\}$, and state densities

$$P(\mathbf{Y}_t = \mathbf{y} | S_t^{\omega,B} = r) = \sum_{k=1}^{K} \omega_{r_1,k} \prod_{n=1}^{N} \left[\kappa_{n,r_2} \psi(y_n) + (1 - \kappa_{n,r_2}) \mathcal{SK}(y, \lambda_{n,r_1}^{(1)}, \lambda_{n,r_1}^{(2)}) \right], \quad (6)$$

where r_1 and r_2 are indexes associated with $S_t^{\omega,B}$ in the representation displayed in Figure 4.

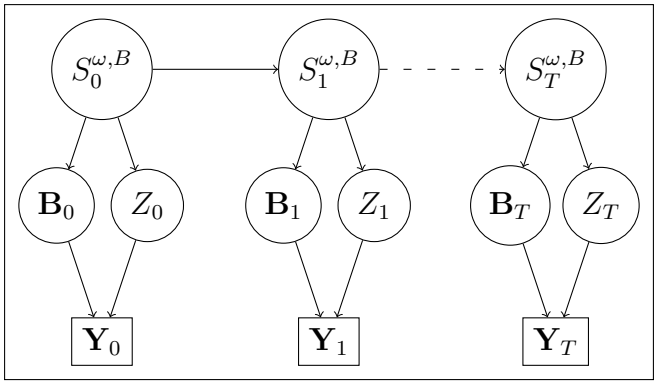


Figure 4: The model path diagram with one chain, $S_t^{\omega,B}$.

Under assumption (A1), $S_t^{\omega,B}$ is irreducible implying that the rank of $\mathbf{\Gamma}^{\omega,B}$ is full. Furthermore, under assumptions (A3) and (A4), the state densities reported in (6) are distinct. Hence, Proposition 2 of Gassiat et al. (2016) holds true, and Theorem 1 can be applied. Given that the identifiability of the DMS holds under assumptions (A1)–(A3), the maximum likelihood estimator obtained via the EM algorithm outlined in the next section is guaranteed to be the solution of a well posed problem (up to label swapping).

3 Estimation via the EM algorithm

In order to develop an EM algorithm for the maximum likelihood (ML) estimation of the parameters of the DMS model, we exploit the stochastic representation of the Skellam distribution as the difference between two independent Poisson distributions reported in (1). Consider now the joint distribution of $(Y_{n,t}, X_{n,t}^{(1)}) | (Z_t = k, B_{n,t} = 0)$; this distribution will

be useful for the derivation of the EM algorithm. Omitting for simplicity the conditioning events, we have that

$$\begin{aligned} P(Y_{n,t} = y_n, X_{n,t}^{(1)} = x_n) &= P(Y_{n,t} = y_n | X_{n,t}^{(1)} = x_n) P(X_{n,t}^{(1)} = x_n) \\ &= P(X_{n,t}^{(1)} - X_{n,t}^{(2)} = y_n | X_{n,t}^{(1)} = x_n) P(X_{n,t}^{(1)} = x_n) \\ &= P(X_{n,t}^{(2)} = x_n - y_n) P(X_{n,t}^{(1)} = x_n) \\ &= \frac{e^{-\left(\lambda_{n,k}^{(1)} + \lambda_{n,k}^{(2)}\right)} {\lambda_{n,k}^{(1)}}^{(x_n)} {\lambda_{n,k}^{(2)}}^{(x_n - y_n)}}{x_n!(x_n - y_n)!}, \end{aligned}$$

that is, the product of the two Poisson probability mass functions with intensity $\lambda_{n,k}^{(1)}$ and $\lambda_{n,k}^{(2)}$ evaluated in x_n and $x_n - y_n$, respectively.

Assume to observe a sample of T observations for N price changes collected in the vector $\mathbf{y}_{1:T} = (\mathbf{y}_1', \ldots, \mathbf{y}_T')'$. Consider also the series of unobserved random variables $\mathbf{S}_{1:T}^{\omega} = (S_1^{\omega}, \ldots, S_T^{\omega})'$, $\mathbf{S}_{1:T}^{B} = (S_1^{B}, \ldots, S_T^{B})'$, $\mathbf{Z}_{1:T} = (Z_1, \ldots, Z_T)'$, and $\mathbf{B}_{1:T} = (\mathbf{B}_1', \ldots, \mathbf{B}_T')'$ and their realizations (which are not observed) collected in the vectors $\mathbf{s}_{1:T}^{\omega}$, $\mathbf{s}_{1:T}^{B}$, $\mathbf{z}_{1:T}$, and $\mathbf{b}_{1:T}$. In order to exploit the stochastic representation of the Skellam as the difference of two Poisson distributions, consider also the random variable $\mathbf{X}_{1:T}^{(1)} = (\mathbf{X}_1^{(1)'}, \ldots, \mathbf{X}_T^{(1)'})'$, where $\mathbf{X}_t^{(1)} = (X_{n,t}^{(1)}, n = 1, \ldots, N)'$ and its (unobserved) realization $\mathbf{x}_{1:T}^{(1)}$. We collect all model parameters in the vector $\boldsymbol{\theta} = (\text{vec}(\boldsymbol{\Omega})', \text{vec}(\boldsymbol{\kappa})', \text{vec}(\Gamma^{\omega})', \text{vec}(\Gamma^{B})')$. The likelihood of observed and unobserved random variables, $\mathcal{L}(\boldsymbol{\theta} | \mathbf{S}_{1:T}^{\omega} = \mathbf{s}_{1:T}^{\omega}, \mathbf{S}_{1:T}^{B} = \mathbf{s}_{1:T}^{B}, \mathbf{Z}_{1:T} = \mathbf{z}_{1:T}, \mathbf{B}_{1:T} = \mathbf{b}_{1:T}, \mathbf{X}_{1:T}^{1} = \mathbf{x}_{1:T}^{1}, \mathbf{Y}_{1:T} = \mathbf{y}_{1:T})$, is

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta} | \cdot) = \delta_{s_1^{\omega}}^{\omega} \delta_{s_1^{B}}^{B} &\left(\prod_{t=2}^{T} \gamma_{s_t^{\omega}, s_{t-1}^{\omega}}^{\omega} \right) \left(\prod_{t=2}^{T} \gamma_{s_t^{B}, s_{t-1}^{B}}^{B} \right) \prod_{t=1}^{T} \omega_{z_t, s_t^{\omega}} \\ &\times \prod_{n=1}^{N} \psi(y_{n,t})^{b_{n,t}} \left(\frac{e^{-\left(\lambda_{n,z_t}^{(1)} + \lambda_{n,z_t}^{(2)}\right)} {\lambda_{n,z_t}^{(1)}}^{(x_{n,t})} {\lambda_{n,z_t}^{(2)}}^{(x_{n,t} - y_{n,t})}}{x_{n,t}!(x_{n,t} - y_{n,t})!} \right)^{1 - b_{n,t}} \kappa_{n,s_t^{B}}^{b_{n,t}} (1 - \kappa_{n,s_t^{B}})^{1 - b_{n,t}}. \end{aligned}$$

By taking the logarithm and removing the quantities that do not depend on model param-

eters, we obtain

$$\begin{aligned}\log \mathcal{L}(\boldsymbol{\theta}|\cdot) \propto & \log(\delta^{\omega}_{s^{\omega}_1}) + \log(\delta^{B}_{s^{B}_1}) + \sum_{t=1}^{T} \log(\omega_{z_t,s^{\omega}_t}) + \sum_{t=2}^{T} \log(\gamma^{\omega}_{s^{\omega}_t,s^{\omega}_{t-1}}) \\ & + \sum_{t=2}^{T} \log(\gamma^{B}_{s^{B}_t,s^{B}_{t-1}}) + \sum_{n=1}^{N}\sum_{t=1}^{T} b_{n,t} \log(\kappa_{n,s^{B}_t}) + \sum_{n=1}^{N}\sum_{t=1}^{T} (1-b_{n,t}) \log(1-\kappa_{n,s^{B}_t}) \\ & + \sum_{n=1}^{N}\sum_{t=1}^{T} (1-b_{n,t}) \left(-\left(\lambda^{(1)}_{n,z_t} + \lambda^{(2)}_{n,z_t}\right) + x_{n,t} \log\left(\lambda^{(1)}_{n,z_t}\right) + (x_{n,t}-y_{n,t}) \log\left(\lambda^{(2)}_{n,z_t}\right)\right).\end{aligned}$$

Unfortunately, this log-likelihood cannot be directly maximized due to the presence of latent quantities. The EM algorithm threats these unobserved terms as missing values and proceeds by the estimation of the expected value of the so-called complete data log-likelihood (CDLL). For the implementation of the EM algorithm, we introduce the following additional variables

$$u^{\omega}_{j,t} = \begin{cases} 1, & \text{if } S^{\omega}_t = j \\ 0, & \text{otherwise.} \end{cases} \qquad u^{B}_{h,t} = \begin{cases} 1, & \text{if } S^{B}_t = h \\ 0, & \text{otherwise.} \end{cases} \qquad z_{j,k,t} = \begin{cases} 1, & \text{if } Z_t = k, S^{\omega}_t = j \\ 0, & \text{otherwise.} \end{cases}$$

$$v^{\omega}_{j,l,t} = \begin{cases} 1, & \text{if } S^{\omega}_{t-1} = j, S^{\omega}_t = l, \\ 0, & \text{otherwise.} \end{cases} \qquad v^{B}_{h,l,m} = \begin{cases} 1, & \text{if } S^{B}_{t-1} = h, S^{B}_t = m, \\ 0, & \text{otherwise.} \end{cases}$$

The first two sets of variables, $u^{c}_{j,t}$ and $v^{c}_{j,l,t}$ for $c = \omega, B$, follow from the standard implementation of the algorithm for Markov-switching models, see McLachlan and Peel (2000), whereas the third set, $z_{j,k,t}$ (for $j = 1, \ldots, J$, and $k = 1, \ldots, K$), is specific to our model and is related to the additional latent variables Z_t. The new variables allow us to write the CDLL as

$$\begin{aligned}\log \mathcal{L}^{\mathsf{c}}(\boldsymbol{\theta}|\cdot) \propto & \sum_{j=1}^{J} u^{\omega}_{j,1} \log(\delta^{\omega}_j) + \sum_{l=1}^{L} u^{B}_{l,1} \log(\delta^{B}_l) + \sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{l=1}^{L} u^{B}_{l,t}(1-b_{l,n,t}) \log(1-\kappa_{n,l}) \\ & + \sum_{t=2}^{T}\sum_{j=1}^{J}\sum_{l=1}^{J} v^{\omega}_{j,l,t} \log(\gamma^{\omega}_{j,l}) + \sum_{t=2}^{T}\sum_{j=1}^{L}\sum_{l=1}^{L} v^{B}_{j,l,t} \log(\gamma^{B}_{j,l}) + \sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{l=1}^{L} u^{B}_{l,t} b_{l,n,t} \log(\kappa_{n,l}) \\ & + \sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{k=1}^{K} z_{j,k,t} \log(\omega_{k,j}) + \sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{l=1}^{L}\sum_{n=1}^{N} (1-b_{l,n,t}) z_{j,k,t} \\ & \times \left(-\left(\lambda^{(1)}_{n,k} + \lambda^{(2)}_{n,k}\right) + x^{(1)}_{k,n,t} \log\left(\lambda^{(1)}_{n,k}\right) + \left(x^{(1)}_{j,n,t} - y_{n,t}\right) \log\left(\lambda^{(2)}_{n,k}\right)\right), \end{aligned} \tag{7}$$

where $b_{l,n,t}$ indicates the realization of $B_{n,t}|S_t^B = l$, and $x_{k,n,t}^{(1)}$ indicates the realization of $X_{n,t}|Z_t = k$.

The EM algorithm iterates between the expectation-step (E-step) and maximization-step (M-Step) until convergence. Given a value of the model parameters at iteration m, $\boldsymbol{\Theta}^{(m)}$, the E-step consists in the evaluation of the so-called $\mathcal{Q}$ function defined as $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m)}) = \mathbb{E}^{\boldsymbol{\theta}^{(m)}}(\log \mathcal{L}^{\mathsf{c}}(\boldsymbol{\theta}|\cdot))$, where the expectation is taken with respect to the joint distribution of the missing variables conditional to the observed variables using parameter values at iteration m. Exploiting the formulation of the CDLL in (7), the $\mathcal{Q}$ function can be factorized as

$$\begin{aligned}
\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m)}) \propto & \sum_{j=1}^{J} \widehat{u}_{j,1}^{\omega} \log(\delta_j^{\omega}) + \sum_{l=1}^{L} \widehat{u}_{l,1}^{B} \log(\delta_l^{B}) + \sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{l=1}^{L} \widehat{u}_{l,t}^{B}(1 - b_{l,n,t}) \log(1 - \kappa_{n,l}) \\
& + \sum_{t=2}^{T}\sum_{j=1}^{J}\sum_{l=1}^{J} \widehat{v}_{j,l,t}^{\omega} \log(\gamma_{j,l}^{\omega}) + \sum_{t=2}^{T}\sum_{j=1}^{L}\sum_{l=1}^{L} \widehat{v}_{j,l,t}^{B} \log(\gamma_{j,l}^{B}) + \sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{l=1}^{L} \widehat{u}_{l,t}^{B} \widehat{b}_{l,n,t} \log(\kappa_{n,l}) \\
& + \sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{k=1}^{K} \widehat{z}_{j,k,t} \log(\omega_{k,j}) + \sum_{t=1}^{T}\sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{n=1}^{N} (1 - b_{l,n,t}) \widehat{z}_{j,k,t} \\
& \times \left(- \left(\lambda_{n,k}^{(1)} + \lambda_{n,k}^{(2)} \right) + \widehat{x}_{k,n,t}^{(1)} \log\left(\lambda_{n,k}^{(1)} \right) + \left(\widehat{x}_{k,n,t}^{(1)} - y_{n,t} \right) \log\left(\lambda_{n,k}^{(2)} \right) \right), \qquad (8)
\end{aligned}$$

where $\widehat{u}_{j,t}^{c} = P(S_t^c = j|\mathbf{Y}_{1:T} = \mathbf{y}_{1:T})$, $\widehat{v}_{j,l,t}^{c} = P(S_{t-1}^c = j, S_t^c = l|\mathbf{Y}_{1:T} = \mathbf{y}_{1:T})$ for $c = \omega, B$, $\widehat{z}_{j,k,t} = P(Z_t = k, S_t^{\omega} = j|\mathbf{Y}_{1:T} = \mathbf{y}_{1:T})$, $\widehat{b}_{l,n,t} = P(B_{n,t} = 1|S_t^B = l, \mathbf{Y}_{1:T} = \mathbf{y}_{1:T})$, and $\widehat{x}_{k,n,t} = \mathbb{E}[X_{n,t}^{(1)}|Z_t = k, \mathbf{Y}_{1:T} = \mathbf{y}_{1:T}]$. The E-step of the algorithm involves the computation of these quantities. Furthermore, let us define $\alpha_{q,t} = P(S_t^{\omega,Z,B} = q, \mathbf{Y}_{1:t} = \mathbf{y}_{1:t})$ and $\beta_{q,t} = P(\mathbf{Y}_{t+1:T} = \mathbf{y}_{t+1:T}|S_t^{\omega,Z,B} = q)$, for $\beta_{q,T} = 1$ and $\alpha_{q,0} = \delta_q^{\omega,Z,B}$ for all $q = 1, \ldots, JKL$, where $\delta_q^{\omega,Z,B} = P(S_t^{\omega,Z,B} = q)$ is the stationary distribution of $S_t^{\omega,Z,B}$ in state q. These are the forward and backward probabilities for the third model representation reported in Figure 3 and can be evaluated using the FFBS algorithm. Once these probabilities are evaluated, the following smoothed probabilities can be computed

$$\begin{aligned}
P(S_t^{\omega,Z,B} = q|\mathbf{Y}_{1:T} = \mathbf{y}_{1:T}) &= \frac{\alpha_{q,t}\beta_{q_t}}{\sum_{h=1}^{JKL} \alpha_{h,t}\beta_{h_t}} \\
P(S_t^{\omega,Z,B} = j, S_{t-1}^{\omega,Z,B} = l|\mathbf{Y}_{1:T} = \mathbf{y}_{1:T}) &= \frac{\alpha_{l,t-1}\gamma_{l,j}^{\omega,Z,B} P(\mathbf{Y}_t = \mathbf{y}_t|S_t^{\omega,Z,B} = j)\beta_{j,t}}{\sum_{h=1}^{JKL} \alpha_{h,t}\beta_{h,t}},
\end{aligned}$$

where $\gamma_{j,l}^{\omega,Z,B}$ is the (j,l)-th element of $\Gamma^{\omega,Z,B}$, and $P(\mathbf{Y}_t = \mathbf{y}_t|S_t^{\omega,Z,B} = j)$ is in equation (5).

Since $P(S_t^{\omega,Z,B}, S_{t-1}^{\omega,Z,B}|\mathbf{Y}_{1:T}) = P(S_t^{\omega}, Z_t, S_t^{B}, S_{t-1}^{\omega}, Z_{t-1}, S_{t-1}^{B}|\mathbf{Y}_{1:T})$ and $P(S_t^{\omega,Z,B}|\mathbf{Y}_{1:T}) = P(S_t^{\omega}, Z_t, S_t^{B}|\mathbf{Y}_{1:T})$ it follows that $\widehat{u}_{j,t}^{c}$, $\widehat{v}_{j,l,t}^{c}$ for $c = \omega, B$, and $\widehat{z}_{j,k,t}$ can be evaluated by simple marginalization of the relevant variables from $P(S_t^{\omega,Z,B}|\mathbf{Y}_{1:T})$ and $P(S_t^{\omega,Z,B}, S_{t-1}^{\omega,Z,B}|\mathbf{Y}_{1:T})$. It also follows that the joint probabilities $P(S_t^{\omega} = j, Z_t = k, S_t^{B} = l|\mathbf{Y}_{1:T})$ are immediately available. The remaining quantities are given by

$$\widehat{b}_{l,n,t} = \begin{cases} 0, & \text{if } y_{n,t} \neq 0 \\ \sum_{k=1}^{K}\sum_{j=1}^{J} \frac{\kappa_{n,l} P(S_t^{\omega}=j, Z_t=k, S_t^{B}=l|\mathbf{Y}_{1:T})}{P(S_t^{B}=l|\mathbf{Y}_{1:T})\left(\kappa_{n,l}+(1-\kappa_{n,l})\mathcal{SK}(y,\lambda_{n,k}^{(1)},\lambda_{n,k}^{(2)})\right)}, & \text{otherwise,} \end{cases}$$

and

$$\widehat{x}_{k,n,t} = \sum_{j=1}^{J}\sum_{l=1}^{L} \lambda_{n,k} \frac{\mathcal{SK}(y-1, \lambda_{n,k}^{(1)}, \lambda_{n,k}^{(2)})}{\mathcal{SK}(y, \lambda_{n,k}^{(1)}, \lambda_{n,k}^{(2)})} \frac{P(S_t^{\omega}=j, Z_t=k, S_t^{B}=l|\mathbf{Y}_{1:T}=\mathbf{y}_{1:t})}{P(Z_t=k|\mathbf{Y}_{1:T}=\mathbf{y}_{1:t})}.$$

In the M-step of the algorithm, the function $\mathcal{Q}$ is maximized with respect to the model parameters $\boldsymbol{\theta}$. Solving the Lagrangian associated with this (constrained) optimization leads to the following solution of the maximization problem:

$$\gamma_{j,l}^{\omega\,(m+1)} = \frac{\sum_{t=2}^{T}\widehat{v}_{j,l,t}^{\omega}}{\sum_{l=1}^{L}\sum_{t=2}^{T}\widehat{v}_{j,l,t}^{\omega}}, \quad \gamma_{j,l}^{B\,(m+1)} = \frac{\sum_{t=2}^{T}\widehat{v}_{j,l,t}^{B}}{\sum_{l=1}^{L}\sum_{t=2}^{T}\widehat{v}_{j,l,t}^{B}}, \quad \kappa_{n,c}^{(m+1)} = \frac{\sum_{t=1}^{T}\widehat{u}_{c,t}^{B}\widehat{b}_{n,t}}{\sum_{t=1}^{T}\widehat{u}_{c,t}^{B}},$$

$$\omega_{k,j}^{(m+1)} = \frac{\sum_{t=1}^{T}\widehat{u}_{j,t}^{\lambda}\widehat{z}_{k,t}}{\sum_{t=1}^{T}\sum_{l=1}^{L}\widehat{u}_{l,t}^{\lambda}\widehat{z}_{k,t}}, \quad \lambda_{n,k}^{(1)\,(m+1)} = \frac{\sum_{t=1}^{T}\sum_{j=1}^{J}\widehat{z}_{j,k,t}(1-\widehat{b}_{l,n,t})(y_{n,t}-2\widehat{x}_{n,t})}{\sum_{t=1}^{T}\widehat{z}_{k,t}(1-\widehat{b}_{n,t})},$$

$$\lambda_{n,k}^{(2)\,(m+1)} = \lambda_{n,k}^{(1)\,(m+1)} - \frac{\sum_{t=1}^{T}\sum_{j=1}^{J}\widehat{z}_{j,k,t}(1-\widehat{b}_{l,n,t})y_{n,t}}{\sum_{t=1}^{T}\widehat{z}_{k,t}(1-\widehat{b}_{n,t})}.$$

Given an initial guess $\boldsymbol{\theta}^{(0)}$, the algorithm iterates between the E- and the M-steps until convergence. Convergence to a local optimum is guaranteed since the M-step increases the likelihood value at each iteration. As for standard HMMs, the likelihood function can present several local optima and there is no guarantee that convergence to the global optimum is achieved. To this end, running the algorithm several times with different starting values is a standard procedure to better explore the likelihood surface.

3.1 Intradaily seasonality

A well known stylized fact of high frequency prices is that the variability of their changes exhibits a pervasive intradaily seasonal pattern, see among others Andersen and Bollerslev (1997) and the recent contribution of Andersen et al. (2018). For instance, at the opening of the market, the volatility is generally at its peak as a consequence of the re-balancing activity by market participants processing the overnight information. On the contrary, the volatility is typically very low during lunch time.

With the goal of incorporating this stylized fact in our modeling framework, we introduce a series of dummy variables f_{dt} for $d = 1, \ldots, D$ associated with different intradaily trading periods. We set $D = 14$ dummy variables according to the following scheme: $d = 1$ for 9:30-9:35, $d = 2$ for 9:35-10:00, $d = 3$ for 10:00-10:30, $d = 4$ for 10:30-11:00 and so on, until $d = 14$ for 16:30-17:00. To preserve the tractability of the model outlined above, we impose that the intensity parameters of the two Poisson distributions are time-varying with the following multiplicative structure $\lambda^{(h)}_{n,k,t} = \lambda^{(h)}_{n,k}\beta_{n,t}$ for $h = 1, 2$, where $\beta_{n,t} = \prod_{d=1}^{D} \beta_{n,d}^{f_{d,t}}$. The additional parameters $\beta_{n,d} > 0$ for $n = 1, \ldots, N$ and $d = 1, \ldots, D$ inflate (or deflate) the Poisson intensity parameters when $f_{d,t} = 1$. Note that different assets are allowed to react differently to each trading period and that the previously defined latent variables do not affect this feature. It follows that the $\beta_{n,d}$ coefficients can be regarded as a fixed intradaily effect on the trading intensity. The E-step previously presented remains unchanged by this modification, except for replacing $\lambda^{(h)}_{n,k}$ with $\lambda^{(h)}_{n,k,t}$. The closed-form M-step for the new set of parameters at iteration $(m + 1)$ is given by

$$\beta_{n,d}^{(m+1)} = \frac{\sum_{t=1}^{T} f_{d,t} \sum_{l=1}^{L} \sum_{k=1}^{K} (1 - b_{l,n,t})(2x_{k,n,t} - y_{n,t})}{\sum_{t=1}^{T} f_{d,t} \sum_{l=1}^{L} \sum_{k=1}^{K} (1 - b_{l,n,t})(\lambda_{n,k}^{(1)\,(m+1)} + \lambda_{n,k}^{(2)\,(m+1)})},$$

and the M-step for $\lambda^{(1)}_{n,k}$ and $\lambda^{(2)}_{n,k}$ is slightly modified to

$$\lambda_{n,k}^{(1)\,(m+1)} = \frac{\sum_{t=1}^{T} \sum_{j=1}^{J} \widehat{z}_{j,k,t}(1 - \widehat{b}_{l,n,t})(y_{n,t} - 2\widehat{x}_{n,t})}{\sum_{t=1}^{T} \widehat{z}_{k,t}(1 - \widehat{b}_{n,t})\beta_{n,t}^{(m)}}, \tag{9}$$

$$\lambda_{n,k}^{(2)\,(m+1)} = \lambda_{n,k}^{(1)\,(m+1)} - \frac{\sum_{t=1}^{T} \sum_{j=1}^{J} \widehat{z}_{j,k,t}(1 - \widehat{b}_{l,n,t}) y_{n,t}}{\sum_{t=1}^{T} \widehat{z}_{k,t}(1 - \widehat{b}_{n,t})\beta_{n,t}^{(m)}}. \tag{10}$$

Finally, as it will be evident from Figure 5 in Section 4, the frequency of zeros also exhibits a seasonal pattern over the trading day. If this pattern is not properly accounted for, the unobserved state variable S_t^B can be affected by the seasonal component, preventing a clear interpretation of possible changes in the behaviour of $\mathbf{B}_t$. In this case, we allow the state dependent Bernoulli probabilities $\kappa_{n,l}$ to depend on an additional deterministic seasonal component, $g_{d,t}$, where $g_{d,t} = 1$, if time t coincides with season d, for $d = \{1, \ldots, U\}$. Specifically, we modify the Bernoulli probabilities as $\kappa_{n,l,t} = \sum_{d=1}^{U} g_{d,t}\kappa_{n,ld}$, where $\kappa_{n,l,d}$ are static seasonal-dependent Bernoulli probabilities that need to be estimated alongside the other parameters. The E- and M-steps of all other parameters remain unchanged, while we need to substitute $\kappa_{n,l}$ with $\kappa_{n,l,t}$. The M-step for the new Bernoulli probabilities $\kappa_{n,l,d}$ at iteration $(m+1)$ is given by $\kappa_{n,l,d}^{(m+1)} = \frac{\sum_{t=1}^{T} g_{d,t}\widehat{u}_{l,t}^{B}\widehat{b}_{n,t}}{\sum_{t=1}^{T} g_{d,t}\widehat{u}_{l,t}^{B}}$, for $l = 1, \ldots, L, i = 1, \ldots, N$, and $d = 1, \ldots, U$.

4 Empirical Application

4.1 Data description and summary statistics

We consider the discrete stock price moves of four companies listed on the Dow Jones index (DJIA) in different time periods. The stocks under investigation are the same as in Koopman et al. (2017): Caterpillar (CAT), Coca Cola (KO), JP Morgan (JPM), and Walmart (WMT). We consider two sampling periods: a normal one from November 6, 2013, to November 19, 2013, and a turbulent one (labelled as "Lehman"), from September 11, 2008, to September 25, 2008, which includes the bankruptcy of Lehman Brothers Holdings Inc. Data are collected from the Trades and Quotes (TAQ) database and a preliminary cleaning of the high frequency prices is performed following the procedure of Brownlees and Gallo (2006) and Barndorff-Nielsen et al. (2009). Although the DMS can be employed with price changes observed at any sampling frequency, we have decided to focus on stock prices sampled at 15 seconds by means of the *previous-tick* method. A comparison of the results obtained with other sampling frequencies would be extremely time consuming and would add great length to the paper, thus is it left for future research.

Table 1 reports the main summary statistics of the price changes for the four stocks

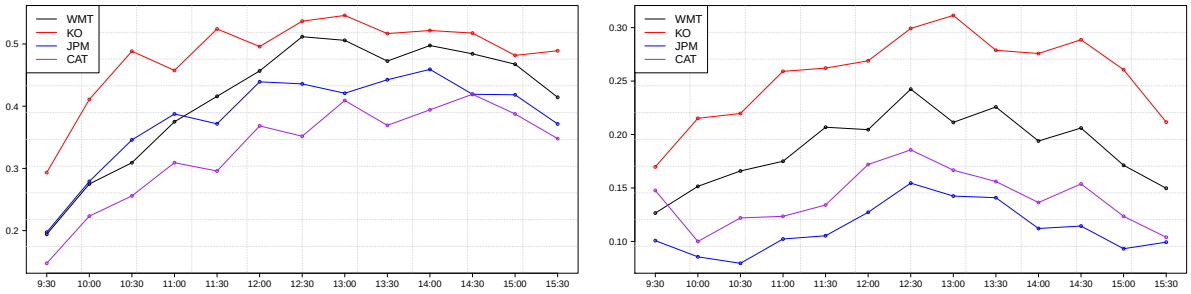
	Normal Period				*Lehman Period*			
	WMT	KO	JPM	CAT	WMT	KO	JPM	CAT
Mean	0.02	0.00	0.01	0.00	0.01	0.00	0.03	-0.04
Mode	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Median	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
% Zeros	0.42	0.50	0.39	0.34	0.20	0.28	0.12	0.16
Max	24	8	21	22	125	119	152	91
Min	-21	-12	-15	-24	-113	-163	-385	-142
Variance	2.12	0.85	2.05	3.35	13.17	11.03	50.44	29.19
Skewness	0.28	-0.19	0.42	-0.11	-0.08	-2.69	-4.41	-0.98
Kurtosis	17.00	5.13	10.34	8.77	118.57	304.70	292.45	47.53

Table 1: Summary statistics of the price changes. The table reports summary statistics for the integer-valued price changes (in cents of dollar) for two different sample periods: normal (from November 6, 2013, to November 19, 2013) and Lehman (from September 11, 2008, to September 25, 2008). The DJIA stocks considered are CAT, KO, JPM, and WMT.

considered. Notably, both the median and the mode of $Y_{i,t}$ is zero for both the normal and the Lehman periods. This provides a first evidence on the relevance of explicitly accounting for an excess probability of zeros when dealing with stock prices sampled at high frequencies. For instance, during normal period, the percentage of zeros is between 34% for CAT and 50% for KO, which is the least liquid asset. The percentage of zeros drastically reduces over the Lehman episode, as a consequence of the large amount of news arriving to the market in this period and the increased uncertainty about the fundamentals across investors. The sample average of price changes is also very close to zero and, especially for the normal period, the level of skewness is almost null, thus signaling a rather symmetric distribution. On the contrary, all series are negatively skewed during the Lehman period: this is due to the arrival of several bad news on the overall stability of the financial sector, which generated large negative price moves. As expected, in this period, both variance and kurtosis are very large, and the magnitude of the price variations might be rather extreme as testified by the maximum and minimum variations in the order of hundreds of cents. Notably, the largest price variations in both the normal and the Lehman periods take place at the opening of the trading day, thus signaling the relevance of properly accounting for this fixed effect trough seasonal dummies as illustrated in Section 3.1.

Figure 5 shows that the probability of zeros is also subject to non-negligible variability at

the intradaily level with a reverse U-shape reflecting the different amounts of trading activity within the day. This evidence is consistent across the four assets under investigation with KO being the least active stock with more than half of the trades associated with zero variations during the central business hours of the normal period.



(a) Normal

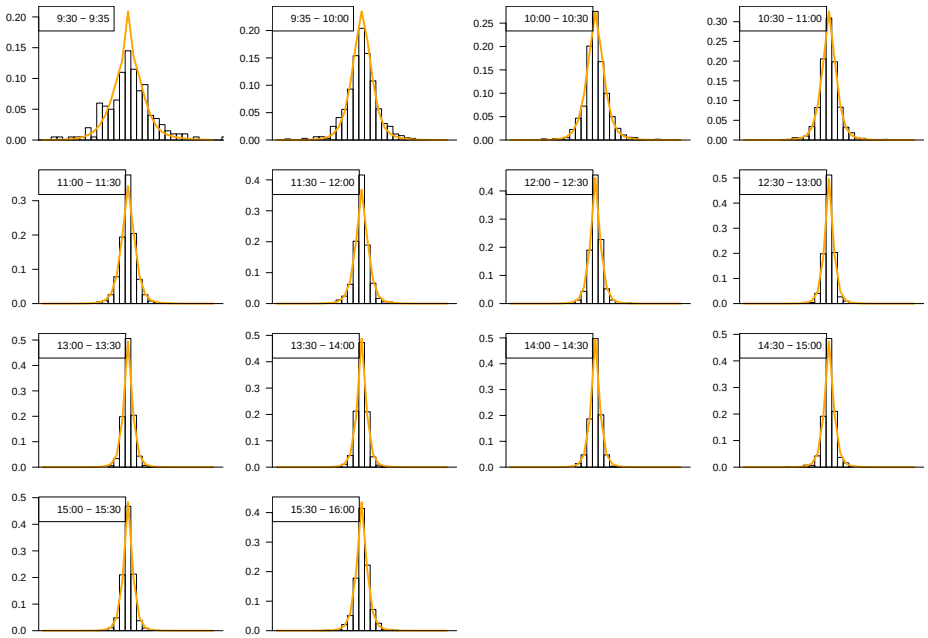
(b) Lehman

Figure 5: Empirical frequency of zeros. The figure reports the percentage of zero-price moves in different intradaily periods (30 minutes) for the two samples: normal (from November 6, 2013, to November 19, 2013) and Lehman (from September 11, 2008, to September 25, 2008). The DJIA stocks considered are CAT, KO, JPM, and WMT.

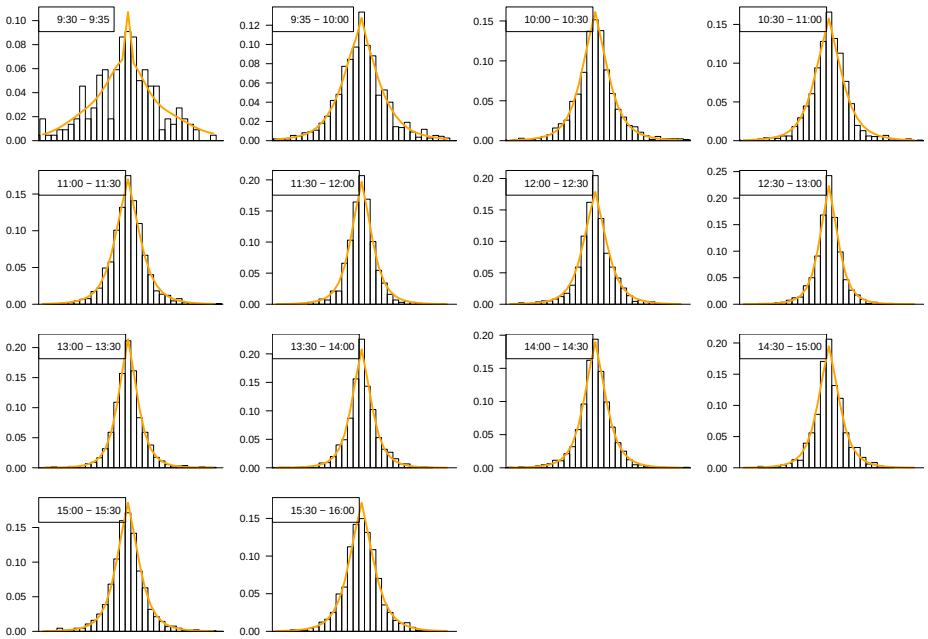
4.2 Model selection and goodness of fit

The DMS model is estimated on both the normal and the Lehman periods for all combinations of $L \in \{1, \ldots, 6\}$, $K \in \{1, \ldots, 15\}$, and $J \in \{1, \ldots, 6\}$. We use fourteen seasonal terms for both the Skellam coefficients and the Bernoulli probabilities as illustrated in Section 3.1. To capture the intensive trading activity at the opening of the market, the first season coincides with the first five minutes of the trading day from 9:30 to 9:35, the second from 9:35 to 10:00, and the remaining run 30 minutes each until the market closure at 16:30. The selection of the best model is performed via the Bayesian Information Criteria (BIC). The BIC selects $J = 5$, $K = 5$ and $L = 1$ for the normal period, and $J = 5$, $K = 12$, and $L = 2$ for the Lehman period.[3] Interestingly, the variability and erratic nature of the price moves during the Lehman episode requires not only many mixture components (K = 12), but also two states for the excess probability of zeros. On the contrary, a more parsimonious model is selected for the normal period. The goodness of fit of the univariate marginal distributions

[3]All details are available upon request to the authors.



(a) Normal



(b) Lehman

Figure 6: Comparison between the empirical and model-implied unconditional distribution of CAT the normal, Panel (a), and the Lehman, Panel (b), period. The first figure of each panel reports the distribution of the first 5 minutes of trading activity (9:30 - 9:35), the second reports the distribution for the following 25 minutes (9:35 - 10:00), figures from the third to the fifteen display the distribution computed every 30 minutes. Yellow lines represent the probability implied by the unconditional distribution of the DMS model.

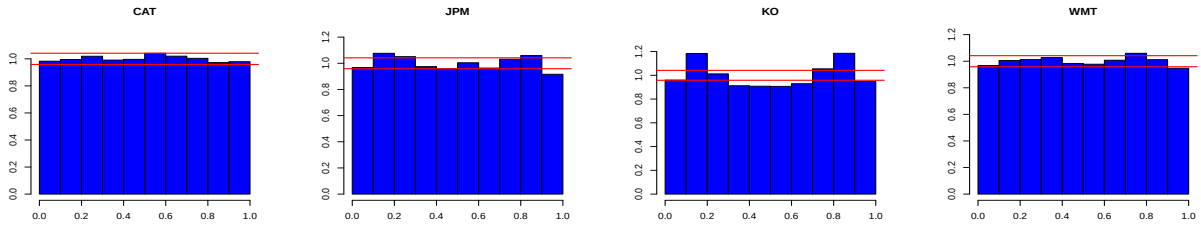
can be visually assessed by looking at Figure 6. The fit to the empirical frequencies achieved by the unconditional distribution of the DMS model is remarkable, and it signals the ability

of the dynamic mixture model to adapt to different market conditions and intensities of the trading process.[4] Indeed, the fit proves robust for all the intradaily business periods defined according to the seasonal dummies (9:30-9:35, 9:35-10:00, 10:00-10:30,...). As expected, the empirical distribution is more dispersed at the opening, i.e., from 9:30 until 9:35, thus justifying the use of a specific seasonal term, $\beta_{1,d}$, for this period. Furthermore, during the Lehman episode, the probability mass is more dispersed than in the normal period; also during the central hours of the day.

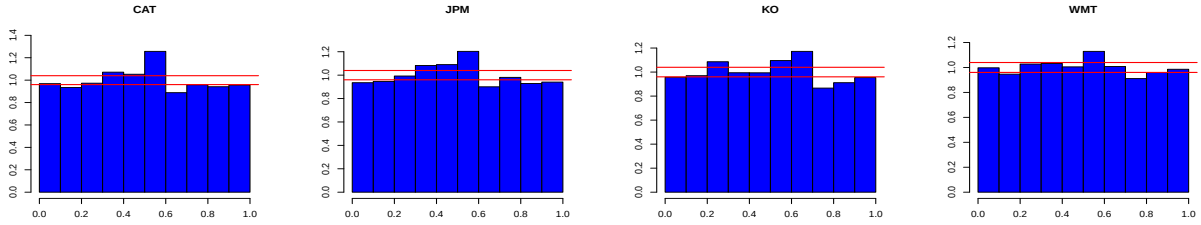
We also assess the quality of the fit of the univariate distributions by means of the test of Berkowitz (2001), which is a classic tool used to study the quality of density forecasts for financial risk management applications. The test is based on the probability integral transforms (PITs) of the data with respect to their conditional (on the past) distribution, which for the DMS is easily computed from Equation (3) by marginalization. The Berkowitz's test relies upon the previous results by Fisher (1932) and Pearson (1938) stating that, under correct model specification and when the support of the observables is continuous, PITs should be iid uniformly distributed over the $(0, 1)$ interval, and their transformation according to the Gaussian quantile function should be iid Gaussian distributed. For discrete random variables the PITs cannot be uniformly distributed, and modifications should be made to the testing procedure. To tackle this issue, we compute the randomized, yet uniform, PITs for integer valued variables derived by *continuization* of the discrete conditional pmf, see Smith (1985), Brockwell (2007), and Liesenfeld et al. (2008).

Figure 7 displays the histogram of the PITs divided in ten bins for all series. We report results for both the in-sample and the out-of-sample periods, where the latter covers the 10 trading days after the in-sample period. The plots highlight the ability of DMS to provide an overall good fit. Indeed, the PITs are approximately uniformly distributed in all cases since the vast majority of the relative frequencies (blue columns) falls within the 95% confidence bands (red line), which are very narrow due to the extremely large sample size. Table 2 reports the results of the Berkowitz's testing procedure. Columns labeled $\tau = 1\%$, $\tau = 5\%$, and $\tau = 10\%$, report the value of the Berkowitz's test statistic, when the coverage below

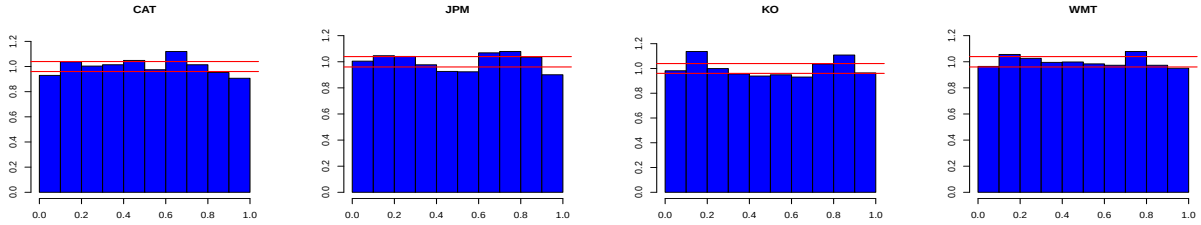
[4]Figures 4-6 in the supplementary material confirm an analogous level of goodness of fit for the other stocks.



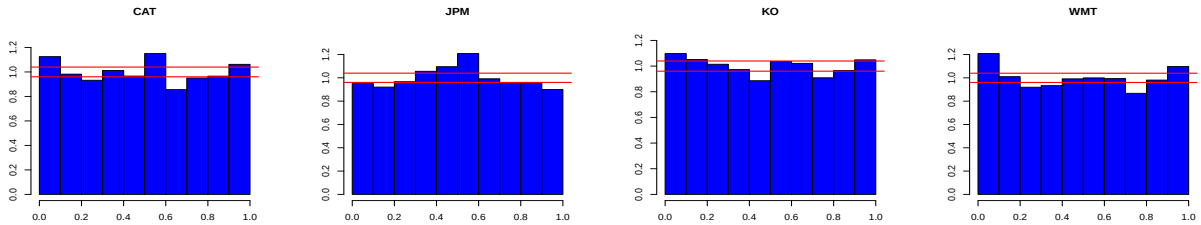
(a) Normal - In-Sample



(b) Lehman - In-Sample



(c) Normal - Out of Sample



(d) Lehman - Out of Sample

Figure 7: PITs. In-sample and out-of-sample randomized PITs as in Brockwell (2007) computed according to the one step ahead univariate conditional distribution of each asset. PITs are divided in 10 bins such that under the null hypothesis of correct model specification the area of each bin should be 10%. Confidence intervals based on the methodology of Diebold et al. (1998) and computed at the 5% level are reported around the theoretical value of the uniform density.

the τ quantile level is tested. Columns labeled "All" report the statistics, when looking at the entire distribution (i.e., $\tau = 100\%$), and columns labeled "Joint" report the value of the statistic, when independence and coverage are jointly tested. The results from Table

In-sample

	Lehman Period					*Normal Period*				
	$\tau = 1\%$	$\tau = 5\%$	$\tau = 10\%$	All	Joint	$\tau = 1\%$	$\tau = 5\%$	$\tau = 10\%$	All	Joint
WMT	1.21	4.06	1.98	4.55	90.82	0.11	1.65	0.05	3.13	9.36
KO	39.06	82.13	96.28	31.01	725.01	3.12	3.83	9.92	9.43	83.14
JPM	9.74	20.53	19.25	8.64	175.55	11.27	10.09	8.63	19.69	53.59
CAT	0.60	1.10	0.54	2.52	123.56	2.39	6.23	4.17	6.95	44.13

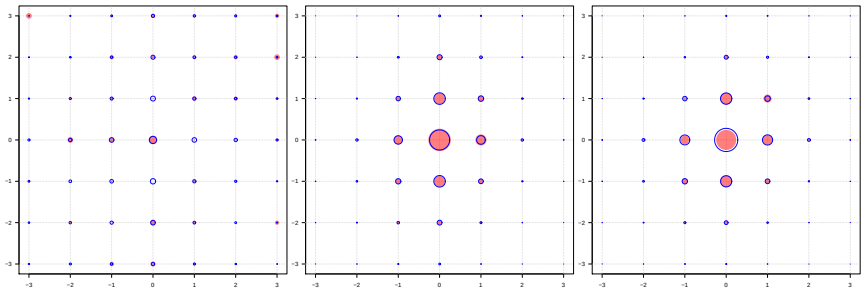
Out-of-sample

	Lehman Period					*Normal Period*				
	$\tau = 1\%$	$\tau = 5\%$	$\tau = 10\%$	All	Joint	$\tau = 1\%$	$\tau = 5\%$	$\tau = 10\%$	All	Joint
WMT	26.90	59.63	85.58	166.17	190.73	0.53	2.74	3.03	2.10	21.55
KO	46.79	58.15	56.84	79.46	113.05	43.56	38.19	40.54	13.56	629.01
JPM	21.53	19.64	22.23	11.89	27.89	4.92	0.85	0.74	4.50	104.96
CAT	6.30	20.60	30.76	48.81	54.74	5.64	6.20	17.57	16.91	78.53

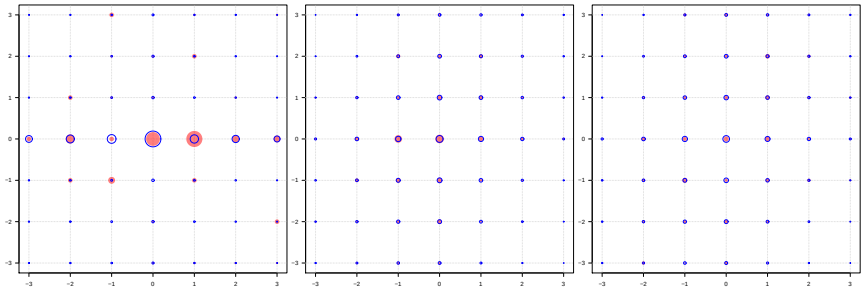
Table 2: LR test statistics of Berkowitz (2001). The tests are computed using the randomized PITs as in Brockwell (2007). We consider the coverage of the left tail below the τ% quantile level. Results are reported for the in-sample and out-of-sample periods during normal market conditions and during the Lehman episode. Columns labeled "All" correspond to unconditional coverage of the whole distribution ($\tau = 100\%$). Columns labeled "Joint" report the statistics associated with the joint test for the null of correct unconditional coverage and independence of the PITs. Gray cells indicate values below the 5% critical value associated with the asymptotic distribution of the test.

2 are mixed and can be summarized as follows: i) the conditional distribution is generally correctly specified for WMT and CAT in both in-sample periods and only for WMT and JPM in the normal out-of-sample period, ii) during the Lehman out-of-sample period, we always reject the null hypothesis, and iii) the null hypothesis of independence and correct coverage of the transformed PIT is always rejected. The rejection of the null hypothesis is somehow expected due to the very large sample size and the parameters instability following the Lehman episode. We conclude that, although the tests reported in Table 2 often reject the null hypothesis, histograms displayed in Figure 7 are encouraging and suggest that the fit of the univariate distributions achieved by DMS is reasonable in both the in-sample and the out-of-sample periods.

The goodness of fit of the bivariate distribution of CAT-WMT for different intradaily periods (opening, lunch, closing) is reported in Figure 8. The fit to the empirical frequencies (red area) by the DMS (blue line) is again remarkable for both the normal and the Lehman periods. For what concerns the normal period, Panel a) highlights that the bivariate distribution of the price variations is rather sparse at the opening, while in Panel b) and c)



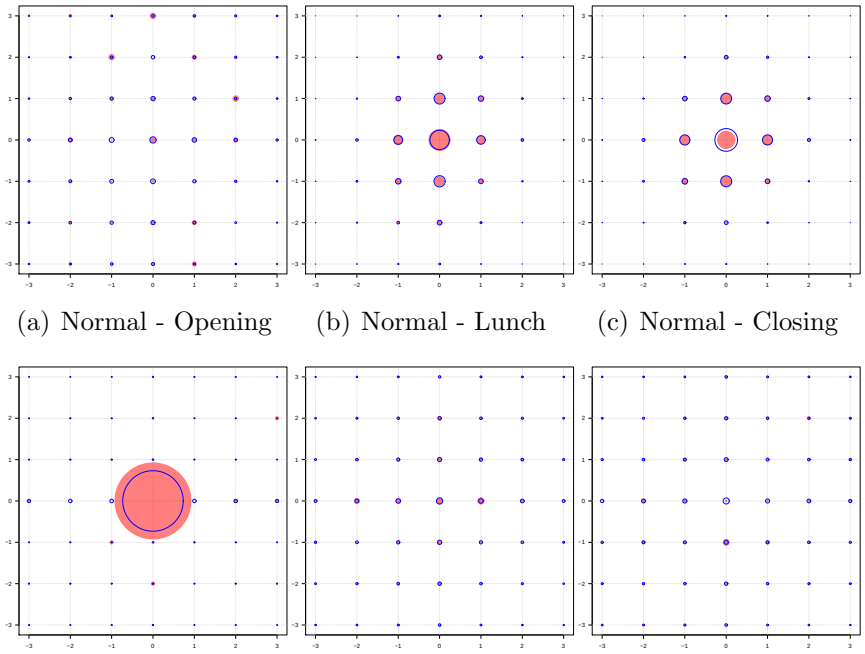
(a) Normal - Opening (b) Normal - Lunch (c) Normal - Closing



(d) Lehman - Opening (e) Lehman - Lunch (f) Lehman - Closing

Figure 8: Bivariate unconditional distributions. The figure reports the empirical and model-based unconditional distributions of CAT and WMT during the opening of the market (9:30-09:35), the lunch time (12:00 - 12:30), and the closing (15:30-16:00) for both the normal and the Lehman periods. The full red circles represent the empirical frequencies computed over the estimation period. The blue circles represent the theoretical frequencies computed according to the unconditional bivariate distribution of CAT and WMT.

most of the probability mass is associated with price variations in the range between -1 and +1 cents, with a relatively high percentage of joint zero variations. The picture drastically changes in the Lehman period. The bivariate empirical probability is dispersed in all intradaily periods (including lunch and closing hour). The fit is remarkable also in this case, suggesting that the DMS model is sufficiently flexible to account for a large number of shapes of the bivariate distribution. In particular, the probability mass on the zeros is extremely high at the opening for CAT-JPM (while not for CAT-WMT). This evidence is associated with the event of a trading halt at the opening on September 15, 2008, for several stocks traded on NYSE. Indeed, at the opening of Monday, September 15, 2008, the trading of CAT, KO, JPM stopped, resulting in a *frozen* market and a prolonged period of no price variations. In particular, Panel c) of Figure 9 displays the effect of the market freezing on the unconditional probability of joint zeros on CAT and JPM. In Section 4.3, we characterize the ability of the DMS model to predict and adapt to prolonged periods of price staleness.



(a) Normal - Opening (b) Normal - Lunch (c) Normal - Closing

(d) Lehman - Opening (e) Lehman - Lunch (f) Lehman - Closing

Figure 9: Bivariate unconditional distributions. The figure reports the empirical and model-based unconditional distributions of CAT and JPM during the opening of the market (9:30-09:35), the lunch time (12:00 - 12:30), and the closing (15:30-16:00) for both the normal and the Lehman period. The full red circles represent the empirical frequencies computed over the estimation period. The blue circles represent the theoretical frequencies computed according to the unconditional bivariate distribution of CAT and JPM.

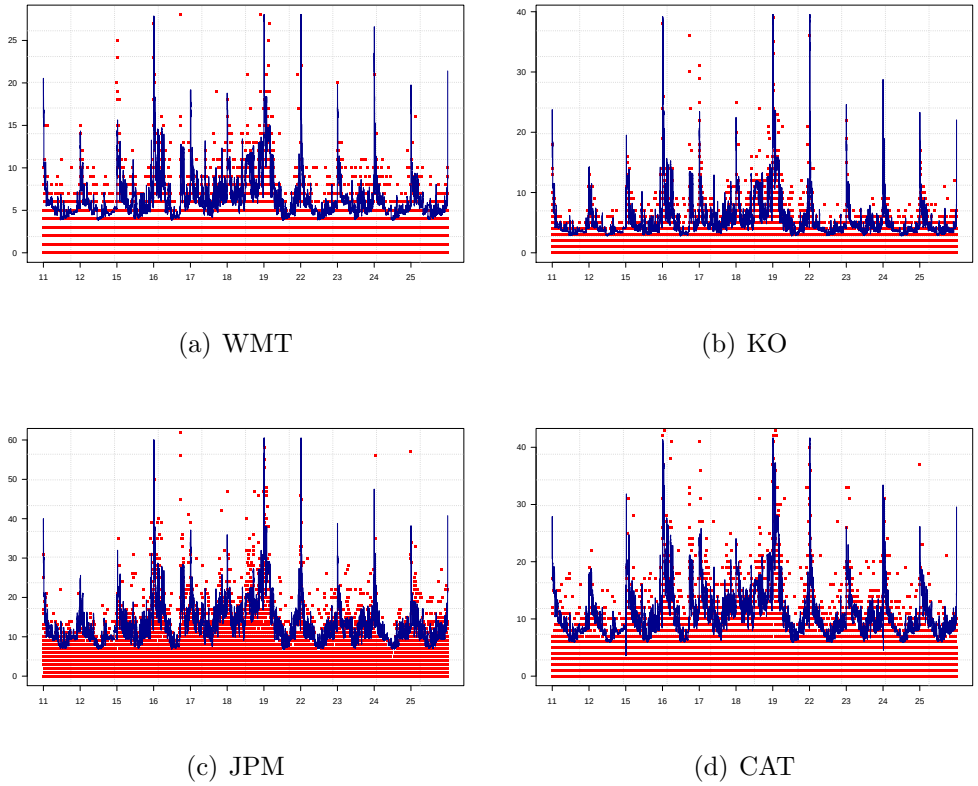
4.2.1 Filtered Variance and Correlation

The same FFBS algorithm adopted in the estimation via EM can be exploited to extrapolate the intradaily (spot) volatilities of each individual stock under consideration. Similarly to Koopman et al. (2017), Figure 10 displays the absolute value of the price changes together with the extrapolated volatilities, $\hat{\sigma}_{t|t-1,i}$.

The extrapolated volatilities are computed as the square root of the diagonal elements of the predicted covariance matrix, $\widehat{\Sigma}_{t|t-1}$, obtained as

$$\widehat{\Sigma}_{t|t-1} = \mathbf{G}_{t|t-1} - \boldsymbol{\mu}_{t|t-1}\boldsymbol{\mu}'_{t|t-1}, \tag{11}$$

where $\mathbf{G}_{t|t-1}$ and $\boldsymbol{\mu}_{t|t-1}$ are the $N \times N$ matrix and N valued vector of conditional second cross moments and mean, respectively. The typical elements $G_{i,j,t|t-1}$ and $\mu_{i,t|t-1}$ are given



(a) WMT (b) KO

(c) JPM (d) CAT

Figure 10: Absolute price changes (red squares) with in-sample predicted volatility, $\hat{\sigma}_{t|t-1,i}$, (blue solid line) during the Lehman period. Predicted volatilities are computed according to the one step ahead conditional distribution over the in-sample period.

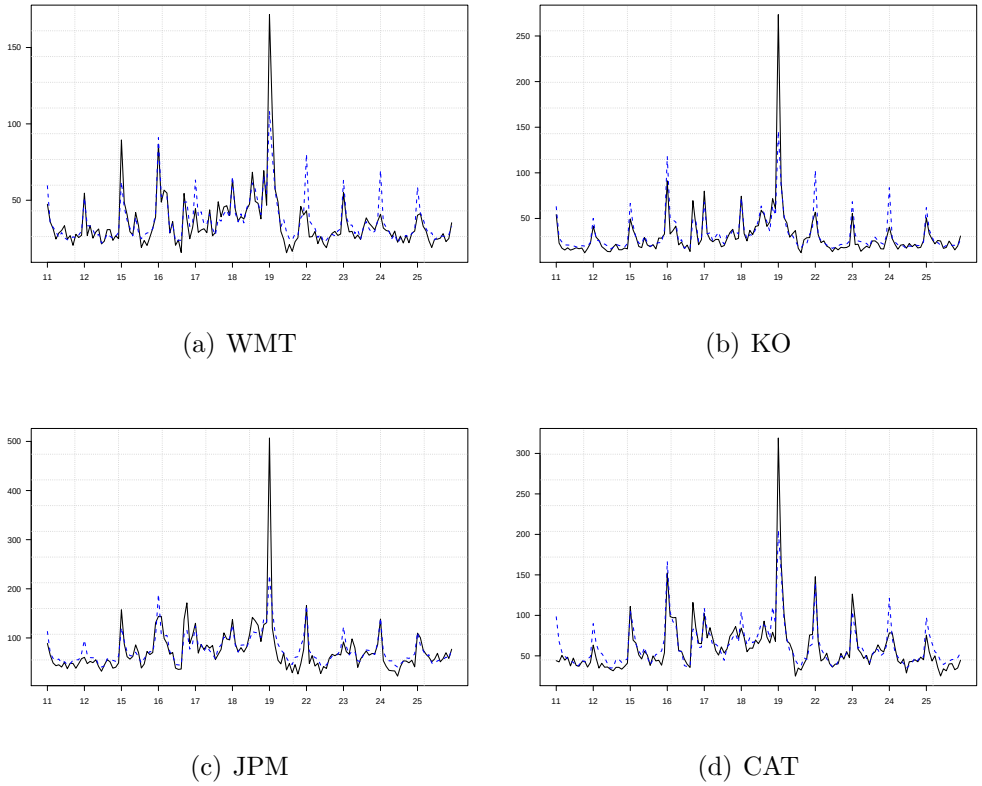
by

$$G_{i,j,t|t-1} = \sum_{j=1}^{J}\sum_{l=1}^{L}\sum_{k=1}^{K} \pi^{\omega}_{j,t|t-1}\pi^{\kappa}_{l,t|t-1}\omega_{j,k}(1-\kappa_{l,i,t})(1-\kappa_{l,j,t})\delta_{i,t}\delta_{j,t} \qquad \text{if} \quad i \neq j$$

$$G_{i,j,t|t-1} = \sum_{j=1}^{J}\sum_{l=1}^{L}\sum_{k=1}^{K} \pi^{\omega}_{j,t|t-1}\pi^{\kappa}_{l,t|t-1}\omega_{j,k}(1-\kappa_{l,i,t})(2\varpi_{i,t}+\delta^2_{i,t}) \qquad \text{if} \quad i = j$$

$$\mu_{i,t|t-1} = \sum_{j=1}^{J}\sum_{l=1}^{L}\sum_{k=1}^{K} \pi^{\omega}_{j,t|t-1}\pi^{\kappa}_{l,t|t-1}\omega_{j,k}(1-\kappa_{l,i,t})\delta_{i,t},$$

where $\delta_{i,t} = \lambda^{(1)}_{i,k,t} - \lambda^{(2)}_{i,k,t}$ and $\varpi_{i,t} = (\lambda^{(1)}_{i,k,t} + \lambda^{(2)}_{i,k,t})/2$. The intradaily patterns in the magnitude of the price variations are clearly reflected in the extrapolated volatilities, which are, by construction, smoother than the ex-post realizations. Similarly to Koopman et al. (2017), we aggregate the (spot) variances, $\hat{\sigma}^2_{t|t-1,i}$, over 30-minutes horizons and compare them with the realized (ex-post) variance based on high frequency data sampled at 1 minute using the realized kernel estimator of Barndorff-Nielsen et al. (2008). Figure 11 suggests that the



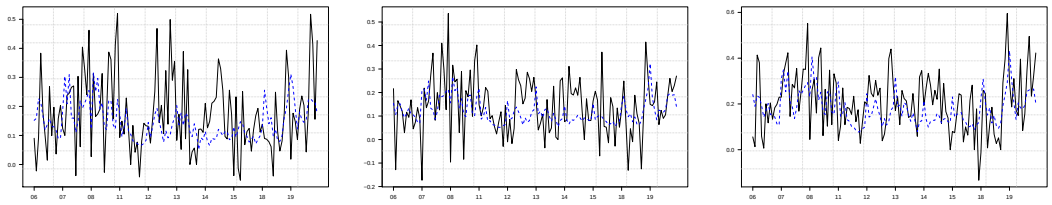
(a) WMT (b) KO

(c) JPM (d) CAT

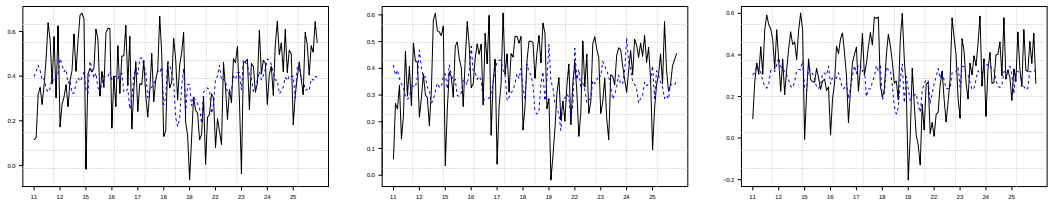
Figure 11: Aggregated predicted volatility (blue dashed line) and realized volatility (black solid line) the Lehman period over 30-minutes intervals. Realized volatilities are computed as square roots of the realized kernel estimator. The aggregated predicted volatilities are computed as the square roots of the one-step-ahead variances, $\hat{\sigma}^2_{t|t-1,i}$, aggregated over 30-minutes intervals.

correlation between the model-based and the realized variances is almost maximal. Analogously, the computation of the predicted covariance matrix in (11) via the FFBS algorithm allows us to compute the correlations aggregated over 30-minutes horizons. These are compared with the realized (ex-post) correlations computed with the realized kernel estimator of Barndorff-Nielsen et al. (2008) based on high frequency prices sampled at 1 minute. Figure 12 highlights the ability of the model-based correlations to provide an unbiased and smooth prediction of the realized ones.

We also perform an out-of-sample analysis of the DMS model to assess its ability to adapt to changing market conditions and to capture the relevant features of the high frequency price changes outside the estimation period. Table 3 presents a summary of the conditional variance forecast accuracy of the DMS compared with that of a SARIMA model estimated on squared price variations with AR, MA, and seasonal AR and MA orders selected according



(a) WMT (b) KO (c) JPM



(d) WMT (e) KO (f) JPM

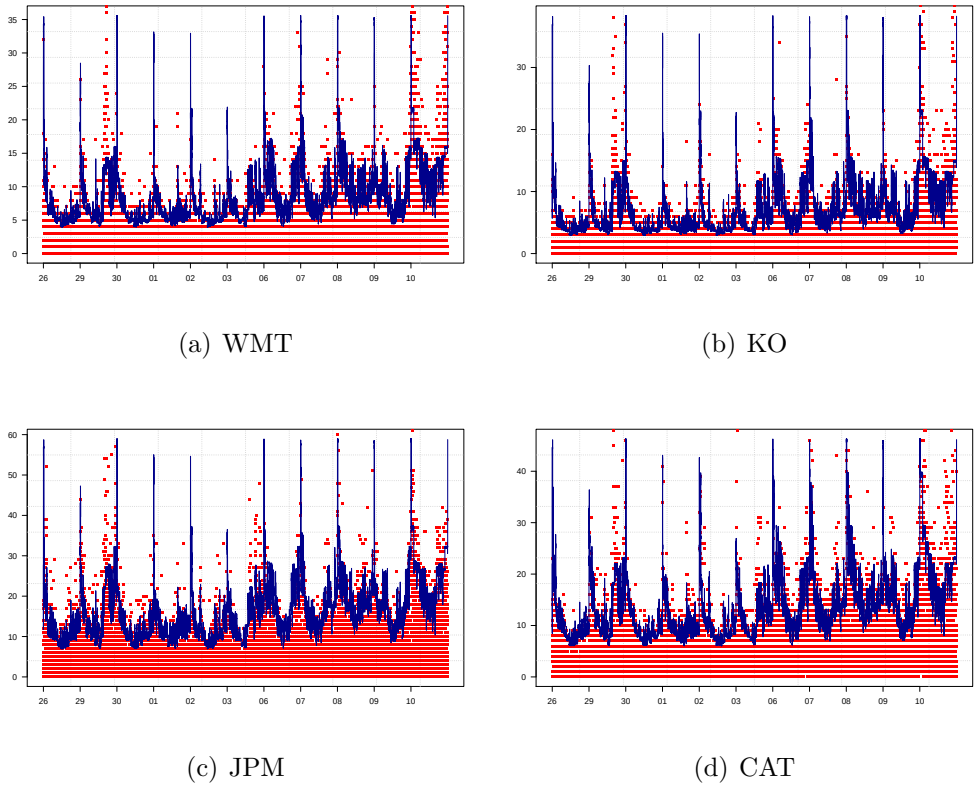
Figure 12: Ex-ante and ex-post correlations. The figures report the filtered correlation after aggregation (blue dashed line) and realized correlation (black solid line) in the Lehman (top) and normal (bottom) periods of CAT versus other assets. The correlation are constructed aggregating $\hat{\Sigma}_{t|t-1}$ in (11) over 30-minutes intervals. The realized correlations are computed through the realized kernel estimator of the covariance matrix of price changes.

	MSE	QLIKE	MSE-LOG	MSE-SD	MSE-Prop	MAE	MAE-LOG	MAE-SD	MAE-Prop
Normal Period									
WMT	1.05	1.06	1.02	1.02	1.11	1.01	0.99	0.99	1.06
KO	0.95	1.06	1.01	1.01	1.06	0.99	0.99	0.99	1.06
JPM	0.94	0.96	0.99	0.97	0.96	0.97	0.99	0.98	0.96
CAT	0.96	1.05	1.02	1.01	0.99	0.99	1.01	1.02	1.05
Lehman Period									
WMT	0.83	0.97	0.99	0.90	0.89	0.89	0.99	0.96	0.97
KO	0.97	1.03	0.95	0.97	0.77	1.01	0.96	0.96	1.03
JPM	1.02	0.77	0.91	0.96	0.83	1.01	0.93	0.94	0.77
CAT	0.74	1.13	1.06	0.84	0.97	0.90	1.02	0.98	1.13

Table 3: Volatility predictions. The table reports the comparison of the one-step-ahead volatility predictions of DMS with those of a SARIMA model for both the normal and the Lehman period. Results are reported according to the nine volatility loss functions detailed in Patton (2011). DMS losses computed over the full out-of-sample period are averaged and reported relative to those of the SARIMA. Values smaller than one indicate outperfomance of DMS with respect to SARIMA and viceversa. Gray cells indicate rejection of the bilateral null hypothesis of equal predictive ability of Diebold et al. (1998) at the 5% confidence level.

to BIC. The ex-post variances are proxied by the squared price variations. The comparison of the predictive accuracy of the two models is performed through the Diebold and Mariano (2002) test based on a number of loss functions, which are those adopted in Patton (2011). The forecasting window includes the 10 days after the in-sample interval for both the normal

and the Lehman periods. Overall, the DMS provides out-of-sample predictions of squared price variations that are statistically superior to those of the SARIMA with 5% significance level in 35 out of 72 cases. On the contrary, the SARIMA is statistically superior only in 18 cases. The forecasts of DMS prove particularly good after the Lehman period, where the forecast accuracy achieved with the DMS is higher than that of the SARIMA in 23 out of 36 cases (while SARIMA is superior in only 4 cases). This finding testifies the ability of the DMS to provide a very flexible conditional distribution of the price variations, which adapts well in mutated market conditions. This is also highlighted in Figure 13, which reports the



(a) WMT (b) KO

(c) JPM (d) CAT

Figure 13: Absolute price changes (red squares) with out-of-sample predicted volatility, $\hat{\sigma}_{t|t-1,i}$, (blue solid line) during the Lehman period. Predicted volatilities are computed according to the one step ahead conditional distribution over the out-of-sample period corresponding to the 10 days after the Lehman period.

absolute price changes (red squares) with the model-based filtered volatility ($\widehat{\sigma}_{i,t|t-1}$) in the out-of-sample period. As for the in-sample period, the volatility patterns closely follow the magnitude of the price variations also in the out-of-sample interval.

4.3 Predicting and disentangling staleness

As shown in Table 1, the unconditional probability of observing zero variations in the dataset of prices observed at 15 seconds frequency is very high and generally well above 30%. This phenomenon is well known in the high frequency literature, see the recent contribution of Bandi et al. (2017). The absence of price movements might signal the inability of a market to frequently update by incorporating relevant information into the stock price. This is possibly associated with (weak) forms of market inefficiency. For instance, price-based illiquidity measures based on the percentage of zeros on a given interval (e.g. at daily level) have been proposed in several papers. For instance, Lesmond (2005) and Bekaert et al. (2007) study the illiquidity on the emerging markets, where the full extent of the available information is not fully reflected in the observed prices. Irregular trading and price staleness have been studied in several articles such as the early works of Atchison et al. (1987) and Lo and MacKinlay (1990), and the more recent contributions of Bandi et al. (2017, 2018), with the definition of excess idle time in the univariate and multivariate context, respectively. A common trait of most of the studies on high frequency market imperfections is the assumption of a continuous underlying price process with microstructural features modeled as an additional source of randomness (like a censoring or a barrier) preventing the efficient price to be observed. Indeed, modeling the price process as a continuous random variable automatically assigns zero probability to the event of zero price variations. On the contrary, the Skellam distribution can assign positive probability to the event of zero price variation. Table 4 displays the parameter estimates of the following predictive logit regression over the out-of-sample period

$$\text{logit}\left(\Pi_{y,t}\right) = \beta_0 + \beta_1 \mathcal{P}_{t|t-1}(Y_t = 0) + W_t\gamma, \tag{12}$$

where $\Pi_{y,t}$ is the probability of a zero price variation at time t, W_t is a vector of control variables and $\mathcal{P}_{t|t-1}(Y_t = 0)$ denotes the model-based predictive probability of no price variation at time t conditional on the information set at time $t-1$, that is

$$\mathcal{P}_{t|t-1}(Y_t = 0) = \sum_{j=1}^{J}\sum_{l=1}^{L}\sum_{k=1}^{K} \pi^{\omega}_{j,t|t-1}\pi^{\kappa}_{l,t|t-1}\omega_{j,k}\left(\kappa_l + (1-\kappa_l)\mathcal{SK}(0,\lambda_k^{(1)},\lambda_k^{(2)})\right),$$

where we explicitly drop the dependence on $i = 1, \ldots, N$ for notational convenience. All figures in Table 4 signal the positive and highly significant dependence between the ex-ante (model-based) probability of zeros and the ex-post realization of price staleness, also when correcting for intradaily seasonal patterns, autocorrelation in the dependent variable, and liquidity of the market as measured by the bid-ask spread.

Intuitively, the presence of zeros in the high frequency prices can be due to several factors. First, the presence of zeros might be the consequence of frictions in the form of bid-

	Normal Period				*Lehman Period*			
	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
WMT								
$\widehat{\beta}_0$	-2.21^{***}	-2.59^{***}	-2.44^{***}	-2.00^{***}	-2.74^{***}	-2.85^{***}	-2.80^{***}	-2.17^{***}
$\widehat{\beta}_1$	4.38^{***}	5.08^{***}	4.36^{***}	4.04^{***}	6.75^{***}	6.74^{***}	6.08^{***}	4.94^{***}
Dummy		✓	✓	✓		✓	✓	✓
Lags			✓	✓			✓	✓
BA				✓				✓
KO								
$\widehat{\beta}_0$	-2.04^{***}	-2.61^{***}	-2.48^{***}	-2.23^{***}	-2.70^{***}	-2.94^{***}	-2.83^{***}	-2.56^{***}
$\widehat{\beta}_1$	3.75^{***}	4.78^{***}	4.27^{***}	4.20^{***}	6.14^{***}	6.94^{***}	5.87^{***}	5.45^{***}
Dummy		✓	✓	✓		✓	✓	✓
Lags			✓	✓			✓	✓
BA				✓				✓
JPM								
$\widehat{\beta}_0$	-2.20^{***}	-2.76^{***}	-2.67^{***}	-2.10^{***}	-3.17^{***}	-3.13^{***}	-3.05^{***}	-2.67^{***}
$\widehat{\beta}_1$	4.20^{***}	5.50^{***}	5.05^{***}	4.84^{***}	9.50^{***}	9.04^{***}	7.32^{***}	6.88^{***}
Dummy		✓	✓	✓		✓	✓	✓
Lags			✓	✓			✓	✓
BA				✓				✓
CAT								
$\widehat{\beta}_0$	-2.34^{***}	-2.68^{***}	-2.60^{***}	-2.20^{***}	-3.17^{***}	-3.38^{***}	-3.27^{***}	-2.89^{***}
$\widehat{\beta}_1$	4.78^{***}	5.33^{***}	4.95^{***}	4.46^{***}	9.50^{***}	10.04^{***}	8.31^{***}	7.33^{***}
Dummy		✓	✓	✓		✓	✓	✓
Lags			✓	✓			✓	✓
BA				✓				✓

Table 4: Estimated coefficients of the logistic regression in (12). The table reports the results for each asset over the normal and Lehman periods. We consider regression (12) with no control variables (a), with seasonal dummies (b), with seasonal dummies and 15 autoregressive terms of the dependent variable (c), with seasonal dummies, autoregressive terms and bid-ask spread (BA). Apexes ***, **, and * indicate statistical significance at the 1%, 5%, and 10% confidence levels, respecively. The standard errors are computed according to the Newey-West heteroscedasticity and autocorrelation consistent (HAC) standard errors.

ask spread, which are partly responsible for the observed sluggishness of the high frequency prices. Second, the absence of price variations might be the consequence of the absence of news, such that the traders do not revise their reservation prices and do not generate any trade and price movement. Third, even in presence of news, if the aggregated traders' reactions to the news are of opposite sign but with the same magnitude, then the observed transaction price does not move. In this case, we say that the market is in a *dyadic* state. Our model is able to separately identify the three sources of zero variation in the observed high frequency transaction price. Hence, we can disentangle the probability of zeros as

- **No news**: $P(Y_{i,t} = 0|B_{i,t} = 0, X_1 = 0, X_2 = 0)$.
- **Dyadic market**: $P(Y_{i,t} = 0|B_{i,t} = 0, X_1 > 0, X_1 = X_2)$.
- **Frictions**: $P(Y_{i,t} = 0|X_1 > 0, X_2 > 0, X_1 \neq X_2)$.

At this point, we look at the relation between trading activity and different sources of price staleness, and we let the mixture of distribution hypothesis of Clark (1973) and Tauchen and Pitts (1983) to provide an ideal and simple setup to interpret the empirical findings. In particular, we relate the absence of price movements to the volume of trades by assuming that the market consists of a finite number, $M \geq 2$, of active traders, who take long or short positions on a given asset. Within a given trading period of unit length (e.g. an hour, a day, a week), the market passes through a sequence of $i = 1, \dots I$ equilibria. The evolution of the equilibrium price is motivated by the arrival of new information to the market. At intra period i, the desired position of the m-th trader ($m = 1, \dots, M$) is $q_{i,m} = \xi(p^*_{i,m} - p_i)$, where $p^*_{i,m}$ is the reservation price of the m-th trader, p_i is the current market price, and the constant $\xi > 0$ measures the resilience of the market. The reservation price of each trader might reflect individual preferences, liquidity issues, asymmetries in information sets, and/or different expectations about the fundamental values. As new information arrives, the traders adjust their reservation prices, resulting in a change in the market price given by the average of the increments of the reservation prices. In absence of news, individual traders do not update their reservation prices and no trading volume is generated. On the contrary, the MDH prescribes that if the aggregated reservation prices of the traders have opposite signs, then trades would take place (and trading volume would be generated), but we would not

observe price moves. Finally, the microstructural frictions such as transaction costs in the form of bid-ask spread (BA) would set to zero the traded quantities, when $|p^*_{i,j} - p_i| < BA$. Summarizing, we expect the price staleness to be associated with absence of trading volume

	Normal Period				*Lehman Period*			
	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
WMT								
$\widehat{\beta}_0$	-5.12^{***}	-7.21^{***}	-7.06^{***}	-5.92^{***}	-8.66^{***}	-27.06^{***}	-27.02^{***}	-27.22^{***}
$\widehat{\beta}_1$	5.40^{***}	5.02^{***}	4.56^{***}	3.81^{***}	14.21^{*}	33.45^{***}	32.72^{***}	33.14^{***}
Dummy		✓	✓	✓		✓	✓	✓
Lags			✓	✓			✓	✓
BA				✓				✓
KO								
$\widehat{\beta}_0$	-8.74^{***}	-11.17^{***}	-10.95^{***}	-2.96	-5.69^{***}	-25.45^{***}	-23.03^{***}	-23.17^{***}
$\widehat{\beta}_1$	10.02^{***}	8.66^{***}	8.19^{***}	7.32^{***}	5.24^{***}	34.50^{***}	14.64^{***}	14.32^{***}
Dummy		✓	✓	✓		✓	✓	✓
Lags			✓	✓			✓	✓
BA				✓				✓
JPM								
$\widehat{\beta}_0$	-7.71^{***}	-23.36^{***}	-23.31^{***}	-17.28^{***}	-6.95^{***}	-26.32^{***}	-25.27^{***}	-25.26^{***}
$\widehat{\beta}_1$	8.81^{***}	8.49^{***}	8.34^{***}	6.78^{***}	30.43^{***}	45.01^{***}	23.85^{***}	23.85^{***}
Dummy		✓	✓	✓		✓	✓	✓
Lags			✓	✓			✓	✓
BA				✓				✓
CAT								
$\widehat{\beta}_0$	-3.93^{***}	-6.62^{***}	-6.38^{***}	-6.43^{***}	-6.04^{***}	-20.38^{***}	-20.49^{***}	-20.67^{***}
$\widehat{\beta}_1$	4.73^{***}	7.54^{***}	6.60^{***}	6.68^{***}	34.63^{***}	-4.98	-2.08	-1.79
Dummy		✓	✓	✓		✓	✓	✓
Lags			✓	✓			✓	✓
BA				✓				✓

Table 5: Estimated coefficients of the logistic regression in (13). The table reports the results for each asset over the normal and the Lehman periods. We consider regression (13) with no control variables (a), with seasonal dummies (b), with seasonal dummies and 15 autoregressive terms of the dependent variable (c), with seasonal dummies, autoregressive terms, and bid-ask spread (BA). Apexes ***, **, and * indicate statistical significance at the 1%, 5%, and 10% confidence levels, respectively. The standard errors are computed according to the Newey-West heteroscedasticity and autocorrelation consistent (HAC) standard errors.

(due to absence of news and frictions), while the trading volume are generated without price moves when the market is in a dyadic state. We study this empirical prediction by looking at the following logit regression

$$\text{logit}\left(\Pi_{v,t}\right) = \beta_0 + \beta_1 \widetilde{\mathcal{P}}_{t|t-1}(Y_t = 0) + W_t\gamma, \tag{13}$$

where $\Pi_{v,t}$ is the probability of zero trading volume at time t, $\widetilde{\mathcal{P}}_t(Y_t = 0) = P(Y_{i,t} = 0|B_{i,t} = 0, X_1 = 0, X_2 = 0) + P(Y_{i,t} = 0|X_1 > 0, X_2 > 0, X_1 \neq X_2)$, and W_t is a vector of control variables such as intradaily seasonal dummies and autoregressive terms. Furthermore, since in the MDH framework the presence of transaction costs would reduce the amount of traded securities, we also control for liquidity proxies in the form of bid-ask spread, since repeated trades on the ask or on the bid sides would result in a sequence of zero price variations associated with non-zero transaction volume. We expect the parameter β_1 to be significantly positive, since absence of news and frictions should increase the probability of observing zero trading volume. Table 5 presents the parameter estimates for all stocks under consideration. For both the normal and the Lehman periods, the predicted probabilities of absence of news and frictions are associated with a significant increase in the probability of observing zero trading volume. This finding also holds when controlling for autocorrelation, intradaily seasonality, and bid-ask spread. This confirms the ability of the DMS to disentangle the price staleness of financial prices observed at high frequencies and associate it to prediction of the reduced trading activity as measured by the absence of trading volume.

5 Conclusions

Building upon the framework of hidden/latent Markov chains, we provide a hierarchical HMM model for multivariate count data based on the Skellam distribution. We apply it to the prices of stocks traded on NYSE and observed at very high frequencies (15 seconds). Our model captures most of the features of the price variations observed at high frequencies both in-sample and out-of-sample. Furthermore, it allows to disclose new characteristics of the financial microstructure. For instance, the model is able to account for the large proportion of contemporaneous zero price variations on several assets (co-staleness), which might be associated with *frozen* market conditions and illiquidity episodes preventing the efficient transmission of news to the financial prices. Furthermore, we study the relationship between the model-implied probability of absence of price variations due to frictions and the absence of trading volume, and we find it is in line with the empirical predictions of the MDH theory coupled with the presence of microstructure noise. To conclude, we believe that

the DMS can be beneficial for several financial applications not limited to the one presented in this paper, e.g., when the goal is to investigate illiquidity spillover effects on a large scale. Furthermore, the DMS might represent a suitable modeling framework also in non-financial applications involving signal extraction in the presence of rounding errors. For instance, when measuring air pollutants to assess their effect on air quality or when predicting the risk of a given disease based on censored scores.

References

Akpoue, B. P. and Angers, J.-F. (2017). Some contributions on the multivariate Poisson-Skellam probability distribution. *Communications in Statistics-Theory and Methods*, 46(1):49–68.

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., and Diebold, F. X. (2010). Parametric and nonparametric volatility measurement. *Handbook of Financial Econometrics: Tools and Techniques*, pages 1–67.

Andersen, T. G., Thyrsgaard, M., and Todorov, V. (2018). Time-varying periodicity in intraday volatility. *Journal of the American Statistical Association*, 0(0):1–39.

Atchison, M. D., Butler, K. C., and Simonds, R. R. (1987). Nonsynchronous security trading and market index autocorrelation. *The Journal of Finance*, 42(1):111–118.

Bandi, F. M., Pirino, D., and Renò, R. (2017). Excess idle time. *Econometrica*, 85(6):1793–1846.

Bandi, F. M., Pirino, D., and Renò, R. (2018). Systematic staleness. Technical report, SRRN.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536.

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

Barndorff-Nielsen, O. E., Pollard, D. G., and Shephard, N. (2012). Integer-valued Lévy processes and low latency financial econometrics. *Quantitative Finance*, 12(4):587–605.

Bartolucci, F. and Farcomeni, A. (2009). A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *Journal of the American Statistical Association*, 104(486):816–831.

Bartolucci, F. and Farcomeni, A. (2015). A discrete time event-history approach to informative drop-out in mixed latent Markov models with covariates. *Biometrics*, 71(1):80–89.

Bartolucci, F., Farcomeni, A., and Pennoni, F. (2012). *Latent Markov Models for Longitudinal Data*. Chapman and Hall / CRC Press.

Bekaert, G., Harvey, C. R., and Lundblad, C. (2007). Liquidity and expected returns: Lessons from emerging markets. *The Review of Financial Studies*, 20(6):1783–1831.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4):465–474.

Bien, K., Nolte, I., and Pohlmeier, W. (2011). An inflated multivariate integer count hurdle model: an application to bid and ask quote dynamics. *Journal of Applied Econometrics*, 26(4):669–707.

Brockwell, A. (2007). Universal residuals: A multivariate transformation. *Statistics & Probability Letters*, 77(14):1473–1478.

Brownlees, C. T. and Gallo, G. M. (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis*, 51(4):2232–2245.

Bulla, J., Chesneau, C., and Kachour, M. (2015). On the bivariate Skellam distribution. *Communications in Statistics-Theory and Methods*, 44(21):4552–4567.

Catania, L. and Di Mari, R. (2018). Hierarchical hidden Markov models for multivariate integer-valued time-series. Technical report, SSRN.

Clark, P. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1):135–55.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39:863–883.

Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 20(1):134–144.

Fisher, R. A. (1932). *Statistical methods for research workers*. Genesis Publishing Pvt Ltd.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer Science & Business Media.

Gassiat, É., Cleynen, A., and Robin, S. (2016). Inference in finite state space non parametric hidden Markov models and applications. *Statistics and Computing*, 26(1-2):61–71.

Geweke, J. and Amisano, G. (2011). Hierarchical Markov normal mixture models with applications to financial asset returns. *Journal of Applied Econometrics*, 26(1):1–29.

Grønneberg, S. and Sucarrat, G. (2017). Risk estimation when the zero probability of financial return is time-varying. *MPRA Working Paper*.

Harvey, A. and Ito, R. (2017). Modeling time series with zero observations. Technical report, Economics Group, Nuffield College, University of Oxford.

Hautsch, N., Malec, P., and Schienle, M. (2013). Capturing the zero: a new class of zero-augmented distributions and multiplicative error processes. *Journal of Financial Econometrics*, 12(1):89–121.

Irwin, J. O. (1937). The frequency distribution of the difference between two independent variates following the same poisson distribution. *Journal of the Royal Statistical Society*, 100(3):415–416.

Kömm, H. and Küsters, U. (2015). Forecasting zero-inflated price changes with a Markov switching mixture model for autoregressive and heteroscedastic time series. *International Journal of Forecasting*, 31(3):598–608.

Koopman, S. J., Lit, R., and Lucas, A. (2017). Intraday stochastic volatility in discrete price changes: the dynamic Skellam model. *Journal of the American Statistical Association*, 112(520):1490–1503.

Koopman, S. J., Lit, R., Lucas, A., and Opschoor, A. (2018). Dynamic discrete copula models for high-frequency stock price changes. *Journal of Applied Econometrics*, 1:1–20.

Koopman, S. J., Lucas, A., and Scharth, M. (2015). Numerically accelerated importance sampling for nonlinear non-Gaussian state-space models. *Journal of Business & Economic Statistics*, 33(1):114–127.

Lesmond, D. A. (2005). Liquidity of emerging markets. *Journal of Financial Economics*, 77(2):411–452.

Liesenfeld, R., Nolte, I., and Pohlmeier, W. (2008). *Modelling financial transaction price movements: a dynamic integer count data model*, pages 167–197. Physica-Verlag HD, Heidelberg.

Lo, A. W. and MacKinlay, A. C. (1990). An econometric analysis of nonsynchronous trading. *Journal of Econometrics*, 45(1-2):181–211.

Maruotti, A. (2011). Mixed hidden Markov models for longitudinal data: an overview. *International Statistical Review*, 79(3):427–454.

Maruotti, A. and Rydén, T. (2009). A semiparametric approach to hidden markov models under longitudinal observations. *Statistics and Computing*, 19(4):381.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. John Wiley & Sons.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256.

Pearson, E. S. (1938). The probability integral transformation for testing goodness of fit and combining independent tests of significance. *Biometrika*, 30(1/2):134–148.

Rossi, E. and Santucci de Magistris, P. (2018). Indirect inference with time series observed with error. *Journal of Applied Econometrics*, 33(6):874–897.

Rydberg, T. H. and Shephard, N. (2003). Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics*, 1(1):2–25.

Shephard, N. (2005). *Stochastic volatility: selected readings*. Oxford University Press on Demand.

Shephard, N. and Yang, J. J. (2017). Continuous time analysis of fleeting discrete price moves. *Journal of the American Statistical Association*, 112(519):1090–1106.

Skellam, J. G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society*, 109(3):296–296.

Smith, J. (1985). Diagnostic checks of non-standard time series models. *Journal of Forecasting*, 4(3):283–291.

Tauchen, G. E. and Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica*, 51:485–505.

Vermunt, J. K., Langeheine, R., and Böckenholt, U. (1999). Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, 24:179–207.

Zeger, S. L. and Brookmeyer, R. (1986). Regression analysis with censored autocorrelated data. *Journal of the American Statistical Association*, 81(395):722–729.

Zucchini, W., MacDonald, I. L., and Langrock, R. (2017). *Hidden Markov models for time series: an introduction using R*. CRC press.