UNIVERSITY OF
CAMBRIDGE

Cambridge Working Papers in Economics

Spline-DCS for Forecasting Trade Volume in High-Frequency Financial Data

Ryoko Ito

CWPE 1606

Spline-DCS for Forecasting Trade Volume in High-Frequency Financial Data

Ryoko Ito*
Faculty of Economics, University of Cambridge

This version: January 24, 2016

ABSTRACT

The Spline-DCS model is developed and applied to forecasting the high-frequency trade volume of selected equity and foreign currency exchange pairs. The cubic spline model of Harvey and Koopman (1993) is applied to capture intra-day periodic patterns. The model is robust to outliers as the dynamics of scale is driven by the score. The empirical application illustrates that Spline-DCS is a practical forecasting tool that is robust to the choice of sampling frequency or sampling period. The predictive performance of the model is compared with the state-of-the-art volume forecasting model, named the component-MEM, of Brownlees et al. (2011). The model can substantially outperform the component-MEM in minimizing common forecast error functions.

KEYWORDS: robustness; score; volume prediction; VWAP; slicing strategy; price impact

JEL Classification: C22

*I would like to thank Andrew Harvey for his thoughtful comments and guiding my academic progress throughout my postgraduate studies. I would also like to thank Jamie Walton and Zhangbo Shi for giving me the opportunity to undertake this project and providing helpful guidance. I am also grateful to Philipp Andres, Michele Caivano, Adam Clements, Oliver Linton, Donald Robertson, Mark Salmon, Stephen Thiele, and the participants of the Score Workshop in 2013 at Tinbergen Institute, especially Andre Lucas, for providing thoughtful comments on a preliminary version of this paper. I would also like to thank Xiaxi Huang, Pongphat Taptagaporn, Jason Ricci, Jian Chen, and Yan Li for their useful tips on this project and helping me learn high-frequency finance. Finally, I would like to thank the International Monetary Fund, the Cambridge Trust, and the Royal Economic Society for funding my PhD, and the Keynes Fund and the Stevenson Fund (of the Faculty of Economics, Cambridge University) for funding my travels to present this paper.

1 Introduction

A key objective of execution algorithms in high-frequency trading is to minimize the price impact of a given order by slicing it into smaller transaction sizes and spreading the timing of transactions throughout the day. This reduces the risk of slippage in price, which is the difference between the expected price of a trade and its actual traded price. Accurate intra-day volume prediction can help investors optimize the size and the timing of orders in this sense since the level of market liquidity and trade intensity change throughout the day. It also helps investors achieve the execution price of transactions for the day to be near the Volume-Weighted Average Price (VWAP) benchmark.[1] It is a measure widely-used for a range of purposes, such as assessing the performance of a given trading strategy in minimizing the price impact, or as a guarantee to clients that their orders will be executed at the VWAP target.

Volume prediction in high-frequency finance is a non-trivial task due to the statistically complex features of high-frequency trade volume. Until the seminal work by Brownlees et al. (2011), there has been no well-established methodology for forecasting high-frequency trade volume. Brownlees et al. (2011) introduced the components multiplicative error model (MEM), and showed that the model can outperform some of the existing common methods in volume prediction.

In this paper, we introduce the Spline-DCS model[2] and use it to forecast the trade volume of the IBM stock traded on the New York Stock Exchange (NYSE), as well as two of the most popular currency exchange pairs, euro-dollar (EURUSD) and dollar-yen (USDJPY), traded in the foreign exchange market (FX). We show that the model captures salient empirical features such as intra-day periodic patterns, autocorrelation, and the highly non-Gaussian features of the empirical distribution of our data.

The components MEM model is a GARCH-type observation driven model. It decomposes the dynamics of the conditional moment of data into several components. The periodic component adopts the Fourier series approximation technique, which is a commonly used methodology for capturing intra-day periodic patterns. Our Spline-DCS also has a component specification, but its periodic component applies a cubic spline function. Thus, in order to clarify the distinction between the components-MEM and Spline-DCS, we refer to the

[1] For a given asset class or an order, it is the transaction prices weighted by the associated volume of each transaction.

[2] The model was initially studied by Ito (2013), but the paper has been restructured and partially merged with a later project to become this paper and the latest version of Ito (2013).

former as Fourier-MEM in this paper.

The defining features of Spline-DCS are the cubic spline component and the dynamic specification of the filter. Spline-DCS is an extension of the dynamic conditional score (DCS) model, formally defined and studied by Harvey (2013) and also independently by Creal et al. (2011, 2013). The latter authors call DCS the generalized autoregressive score (GAS) model. It is a relatively new class of observation-driven model for forecasting the conditional density of data via modeling the time-varying volatility, scale, or location. DCS is useful for modeling data with heavy-tails due to its robustness feature: the score, which drives the dynamics of the filter, weighs down the effect of extreme observations when some well-known or useful distributions are used to characterize the distribution of data. The parameters of the model are estimated easily by the method of maximum likelihood (ML).

The periodic component of Spline-DCS applies the cubic spline model of Harvey and Koopman (1993). The authors originally developed it to estimate periodic patterns in hourly electricity demand. It is a parsimonious way of modeling intra-day periodicity, and it can be estimated simultaneously with all other components of the model by ML. Aside from the computational and other empirical advantages of the spline function that we investigate extensively in this paper, an additional notable feature of their cubic spline is that the pattern of periodicity can be allowed to evolve stochastically. As such, Harvey and Koopman (1993) call it the dynamic cubic spline model. Bowsher and Meeks (2008) interpret it as a special type of "dynamic factor model", where the knots of the spline are the factors and the factor loadings are treated as given and specified according to the requirement that the model has to be a cubic spline. Their spline was employed by Harvey et al. (1997) to model a changing seasonal component of weekly money supply in the U.K., and also by Bowsher and Meeks (2008) to forecast zero-coupon yield curves. In order to simplify (and maintain our focus on) the task of illustrating the usefulness of Spline-DCS in high-frequency volume forecasting, we keep the model specification simple by fixing the pattern of periodicity to be the same every day, which is a standard assumption in the literature. Ito (2013) generalizes our proposed model by fully specifying the dynamic cubic spline version of Spline-DCS, and illustrate the empirical merit of allowing for the patterns of intra-day periodicity to evolve in high-frequency finance.

Trade volume is a measure of intensity of trading activity. There is a variety of volume measures including the number of shares traded, number of

transactions, and turnover (shares traded divided by shares outstanding). For the IBM equity data, we model the number of shares traded. For the FX data, the volume is as measured by the traded units in the left hand currency of a given pair (i.e. Euro for EURUSD and dollar for USDJPY), which is priced in the right hand currency of the same pair (i.e. dollar for EURUSD and yen for USDJPY). The sampling frequency we consider ranges between 30 seconds and 10 minutes, which is relatively high in the volume prediction literature.

Although this study deals with non-negative time series, Spline-DCS can be applied to variables with support over the entire real line, in which case the model should be viewed as an extension of Beta-t-EGARCH of Harvey and Chakravarty (2008). As such, Spline-DCS can be used to model asset returns and compared with the seminal work by Andersen and Bollerslev (1998).

The key findings of our analysis can be summarized in two parts as follows. First, we produce both density and level forecasts of volume using Spline-DCS and illustrate the model's practicality, good in-sample fit, and out-of-sample forecasting performance in the context of both equity and FX. Testing the model in these two disperate markets reveal that our model can capture any periodic shape including convex, concave, and multi-modal ones, without imposing any restrictions on the shape of the spline to achieve it. The estimation results are robust to the choice of initial parameters, the sampling frequency, or the sampling period, largely due to the robustness feature of DCS. We compare our model's predictive performance with that of Fourier-MEM, which is the state-of-the-art volume forecasting model in the literature. Our extensive backtesting[3] procedure using the FX data show that Spline-DCS can outperform the components MEM in minimizing common forecast error functions.

Second, we discuss several computational and analytical advantages of Spline-DCS that we identified to highlight its practicality in high-frequency finance. It is a field in which a typical analysis deals with a very large set of data, and thus forecasting can become computationally intensive. These advantages of Spline-DCS we highlight stem from the ML estimation procedure and the use of the spline. Our experience with estimating Fourier-MEM by the generalized method of moments (GMM) as proposed by Brownlees et al. (2011) led us to think that the computational cost of it is noticeably higher than Spline-DCS for a given sample size and convergence tolerance. With the FX data, the optimization procedure for Spline-DCS always converged in less than 5

[3]Backtest is a financial jargon that refers to the test of predictive performance of a model using historical data.

minutes, whereas the Fourier-MEM usually took hours to converge (about 4 hours or longer). We think that the main source of computational cost in Fourier-MEM is the combination of the GMM criteria equation and the use of the Fourier series as the intra-day periodic component. These computational features are compared in Section 7. The spline yields smoother and more intuitive shape of intra-day periodic patterns with relatively few parameters compared to the use of Fourier series.

The plan of this paper is as follows. Section 2 describes the characteristics of our data and motivates the construction of our model. Section 3 defines Spline-DCS. Sections 3.6 and 6.2 lay out our estimation and forecast evaluation methods. Sections 4 and 5 report the in-sample and out-of-sample results for Spline-DCS. Section 6 compares the predictive performance of the competing models using the FX data. Section 7 discusses the computational and practical aspects of the models studied in this paper. Section 8 concludes by laying out aspects of our study that can be improved for further research.

2 Data characteristics

Before proceeding to detailed modeling strategies and forecasting results, it is useful to get an overall feel for the trade volume data. This section provides an initial investigation of our data to motivate our model formally defined in Section 3.

The equity trade volume we consider is the number of shares of IBM stock traded on NYSE during the market opening hours (9.30am-4pm in the New York local time) between Monday 28 February and Friday 31 March 2000, which includes 25 trading days and no public holidays. Our raw data set is in tick-format and consists of the record of every trade in the order of occurrence. The tick-data is irregularly spaced and often has multiple transactions in one second. In order to explore the effects of marginal changes in the aggregation interval on our inference, we aggregate the tick-data by 30 seconds and 1 minute. There are 780 observations per trading day if the aggregation interval is 30 seconds, and 390 observations if 1 minute. For convenience, we refer to the aggregated series as IBM30s if the aggregation interval is 30 seconds, and IBM1m if 1 minute. IMB30s aggregates to IBM1m over any 1-minute interval.

We also consider the trade volume of EURUSD and USDJPY, which are two of the most widely-traded pairs in FX. The definition of trade volume we employ here is the quantity traded in the units of the left hand currency (e.g. euro for EURUSD), which is priced in the units of the right hand currency (e.g. the US

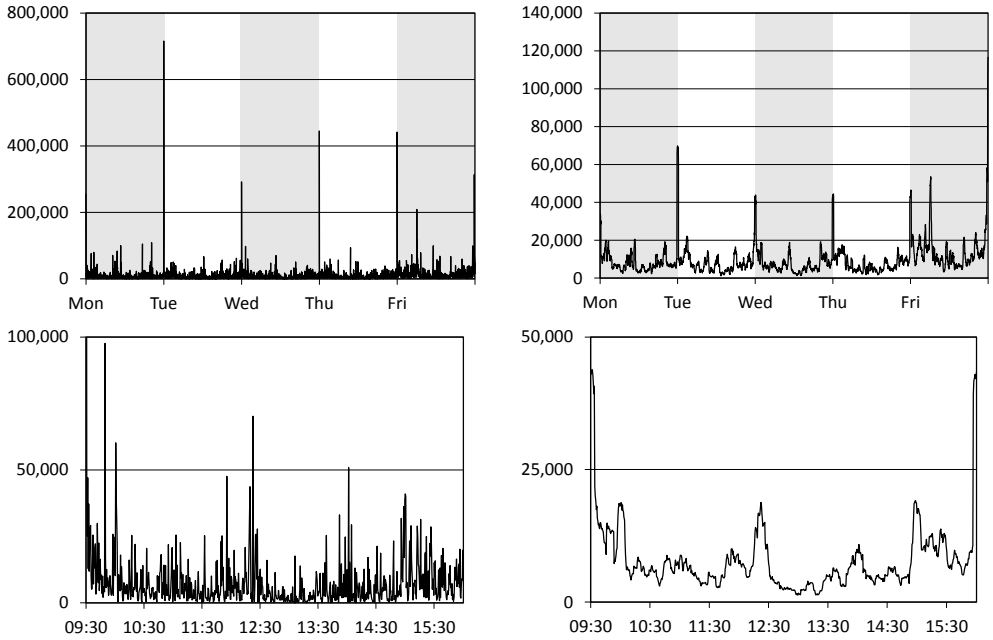


Figure 1. IBM30s (left column) and the same series smoothed by the simple moving average of nearest 20 observations (right column). Time on the x-axis. Top panel: Monday 20 - Friday 24 March 2000. Bottom panel: Wednesday 22 March 2000. Any given day covers the market opening hours between 9.30am-4pm only (in the New York local time).

dollar for EURUSD).[4] The sampling frequency we consider here is 10 minutes. Thus, we have 144 observations per trading day for the FX data. The highest sampling frequency Brownlees et al. (2011) consider is 15 minutes. For a given sampling frequency, the sample size per day for the FX data is more than 3 times larger than for the equity data. This is because trades in FX typically take place throughout the day and night during the weekdays and data is collected 24 hours every day. We use the FX data to compare Spline-DCS and Fourier-MEM in Sections 6 and 7. The sampling period is between Monday 6 January and Sunday 21 December 2014, which is split into sixteen rolling sub-sample windows for the backtesting analysis. See Table 1.

We have more recent and larger samples for the FX data compared to the equity data. This is simply due to the data availability, as we did not have access to more recent volume data for equity at the level of sampling frequency we desired to study. We study both the equity and FX data to illustrate the

[4]For confidentiality reasons, we divided the original FX trade volume series by some arbitrarily chosen constant number to hide the actual level of traded quantity. This pre-estimation transformation does not affect the dynamic structure or the shape of the empirical distribution of data. Thus, the estimated value of ω (defined in Section 3) for the FX data should be viewed in this context.

	In-sample (3 weeks)		Out-of-sample (2 weeks)	
Window #	From	To	From	To
1	Mon 06-Jan-14	Sun 26-Jan-14	Mon 27-Jan-14	Sun 09-Feb-14
2	Mon 27-Jan-14	Sun 16-Feb-14	Mon 17-Feb-14	Sun 02-Mar-14
3	Mon 17-Feb-14	Sun 09-Mar-14	Mon 10-Mar-14	Sun 23-Mar-14
4	Mon 10-Mar-14	Sun 30-Mar-14	Mon 31-Mar-14	Sun 13-Apr-14
5	Mon 31-Mar-14	Sun 20-Apr-14	Mon 21-Apr-14	Sun 04-May-14
6	Mon 21-Apr-14	Sun 11-May-14	Mon 12-May-14	Sun 25-May-14
7	Mon 12-May-14	Sun 01-Jun-14	Mon 02-Jun-14	Sun 15-Jun-14
8	Mon 02-Jun-14	Sun 22-Jun-14	Mon 23-Jun-14	Sun 06-Jul-14
9	Mon 23-Jun-14	Sun 13-Jul-14	Mon 14-Jul-14	Sun 27-Jul-14
10	Mon 14-Jul-14	Sun 03-Aug-14	Mon 04-Aug-14	Sun 17-Aug-14
11	Mon 04-Aug-14	Sun 24-Aug-14	Mon 25-Aug-14	Sun 07-Sep-14
12	Mon 25-Aug-14	Sun 14-Sep-14	Mon 15-Sep-14	Sun 28-Sep-14
13	Mon 15-Sep-14	Sun 05-Oct-14	Mon 06-Oct-14	Sun 19-Oct-14
14	Mon 06-Oct-14	Sun 26-Oct-14	Mon 27-Oct-14	Sun 09-Nov-14
15	Mon 27-Oct-14	Sun 16-Nov-14	Mon 17-Nov-14	Sun 30-Nov-14
16	Mon 17-Nov-14	Sun 07-Dec-14	Mon 08-Dec-14	Sun 21-Dec-14

Table 1. Sub-sampling windows for the FX trade volume data. The overall sampling period between Mon 6 Jan and Sun 21 Dec 2014 is split into sixteen sub-sample windows for the backtesting analysis.

usefulness of our model in these two disperate applications.

Figure 1 gives a snapshot of our equity data. In the top panel, we observe several recurrent spikes in volume near the moment of market opening or closure. These extreme observations make the upper tail of the empirical distribution of data very long. (See Table 2.) The volume fluctuates a lot throughout the day. The smoothed IBM30s series in the right column of Figure 1 reflects that there is a diurnal U-shaped pattern in trading activity on every trading day. High volume in the morning trading hours can be caused by news transmitted over night. The level of activity declines towards lunch time (bottoming out at around 1pm) as overnight information is processed, but picks up again in the afternoon as traders re-balance their positions before the market closes. In the equity context, other measures of trading activity based on durations (such as trade durations, midquote change durations, and volume durations) exhibit similar diurnal patterns, but the shape is inverted so that the duration is long during lunch. (See Hautsch (2012, p.41).)

Figure 2 gives a snapshot of our FX data. Trade volume tends to be high around midday and low in the evening (in GMT). It also fluctuates a lot throughout the day. On Friday 5 September 2014, there is an extreme spike in volume at around 1.30pm in GMT. This coincides with the release of non-farm payroll data in the US, which is known to be one of the most important events in

Series	Obs.	Mean	S.D.	Skew	Max.	Max-99% Q	Zero freq.
IBM30s	19,500	10,539	26,071	29	1,652,100	1,591,073	0.47%
IBM1m	9,750	21,297	39,114	18	1,652,100	1,532,175	0.06%

Table 2. Sample statistics of IBM trade volume. Sampling period is Mon 28 Feb - Fri 31 Mar 2000. The skewness statistics must be interpreted with care as the theoretical skewness may not exist.

	Obs.	Mean	S.D.	Skew	Max	Max-99%Q	Zero freq.
EURUSD (10 mins)	50,112	136	316	9.5	10,017	8,668	2.1%
USDJPY (10 mins)	50,112	121	250	8.8	7,379	6,322	2.0%

Table 3. Sample statistics of EURUSD and USDJPY trade volume. Sampling period: Mon 6 Jan - Fri 19 Dec 2014. The skewness statistics must be interpreted with care as the theoretical skewness may not exist.

Window	1	2	3	4	5	6	7	8
EURUSD (10 mins)	0.1%	0.1%	0.6%	0.8%	2.6%	2.2%	2.2%	2.1%
USDJPY (10 mins)	0.2%	0.0%	0.4%	0.9%	1.9%	2.4%	2.1%	1.9%
Window	9	10	11	12	13	14	15	16
EURUSD (10 mins)	2.7%	1.9%	1.8%	1.3%	1.0%	0.8%	0.3%	0.7%
USDJPY (10 mins)	2.7%	1.9%	1.5%	1.3%	0.9%	0.7%	0.4%	0.0%

Table 4. The percentage of samples for each window that are zero-valued. The sixteen sub-sample windows are listed in Table 1.

FX. This spike highlights the importance of the announcement effect (see, for instance, Andersen and Bollerslev (1998) and Lo and Wang (2010)). There are very few trades between Friday 10pm and Sunday 6pm in GMT. Although trades in FX occur continuously throughout the day during any weekday, the market closes when New York closes on Friday at 9pm in GMT and re-opens when the Auckland market resumes at around 9pm in GMT on Sunday.

Figure 3 shows that the FX volume series also exhibit intra-day periodic patterns. The intra-day percentage distribution of volume appears to have a bimodal pattern for EURUSD and a trimodal pattern for USDJPY. The activity level peaks at around 8am and again at around 2pm in GMT. The market is relatively quiet at around 10pm, but picks up around midnight, particularly for USDJPY. These modes come roughly when trading activity in major markets around the world is high for the day. In GMT outside the daylight saving period, trading is very active in London between 8am and 4pm, in New York between

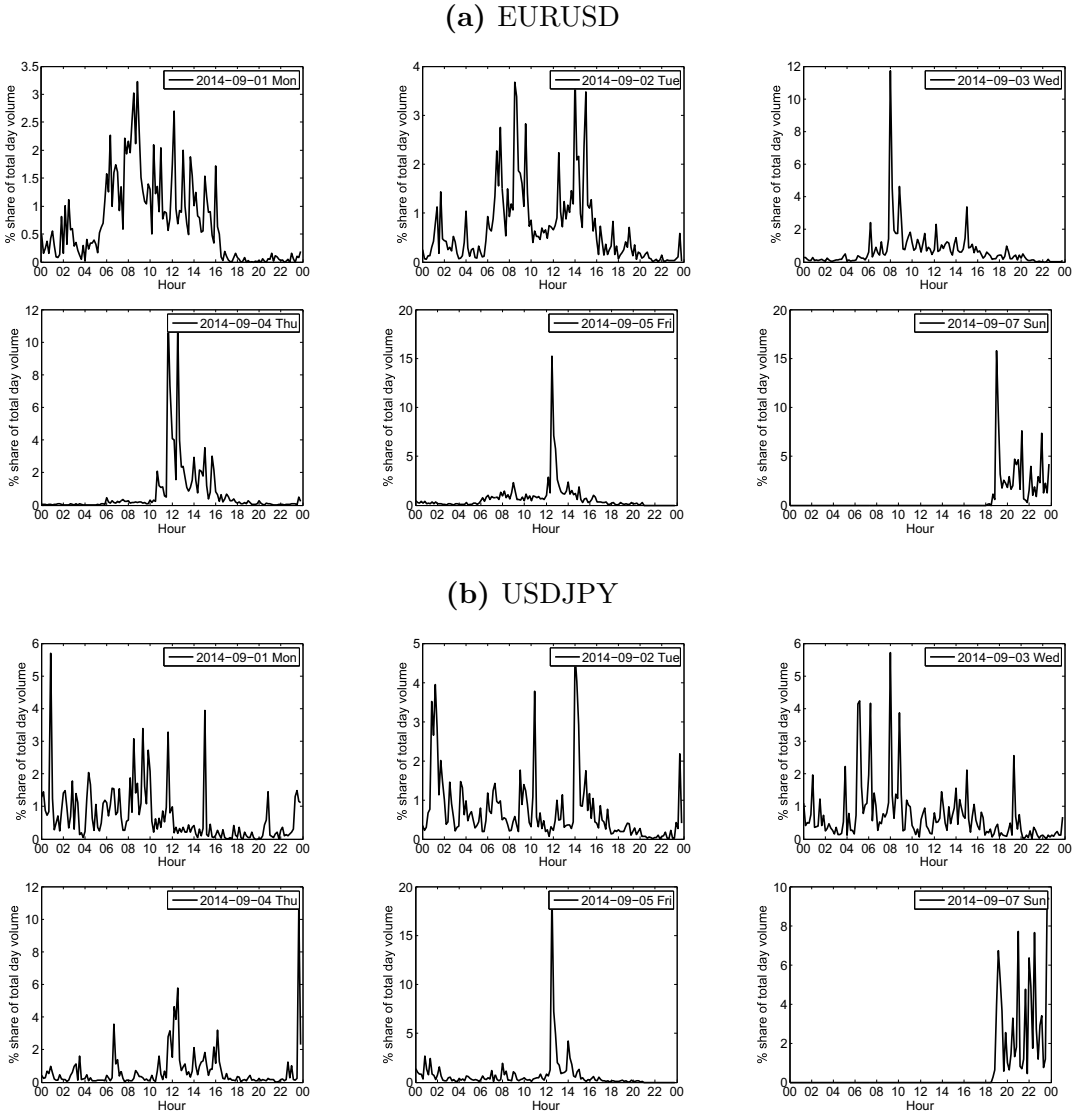


Figure 2. The percentage of total day volume attributed to each intra-day 10-minute bin. The series sum to 100% in each picture. The days are between Mon 1 Sep 2014 and Sun 7 Sep 2014 excluding Sat 6 Sep 2014. Each day covers the 24-hour period (in GMT). Intra-day time on the x-axis.

1pm and 9pm, and in Tokyo between 11pm and 7am. However, trading is not restricted to these hours; for instance, many traders in London trade between 7am and 5pm in the London local time. For EURUSD, volume is particularly high between 1pm and 4pm when the active period in London and New York overlaps. The London and New York markets attract high volume since the bid-ask spread tends to be tighter there for popular currency pairs than in the Asian markets. From this figure, it is not clear whether the pattern of intra-day trading activity differs substantially by the day of the week.

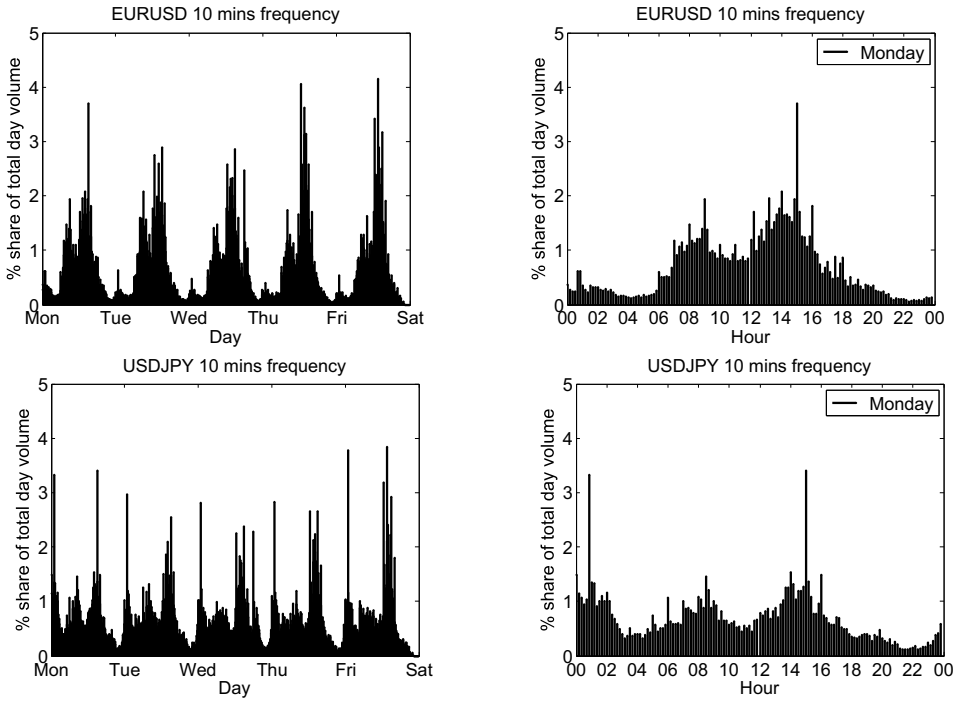


Figure 3. The percentage of total day volume attributed to each intra-day bin. The x-axis is between Monday and Friday (left) and intra-day hours in GMT on Monday (right). The series are obtained by computing the average trade volume at each intra-day bin on each weekday, and dividing it by the total trade volume of that weekday. The series sum to 100% each day. This uses the data between Monday 6 January and Friday 19 December 2014.

The left and middle columns of Figure 4 show that our series are heavily right-skewed and have a heavy (or long) upper-tail. The length of the upper-tail can be also seen in the difference between the maximum and the 99% sample quantile in Tables 2 and 3. The right column of Figure 4 shows the highly persistent nature of our series. IBM30s exhibits statistically significant autocorrelation that is very slow to decay. Sample autocorrelation decays faster for series with wider aggregation interval. Moreover, our series contain a non-negligible number of zero-valued observations. The number of zero-valued observations increases with the sampling frequency. (See Table 2.) Table 4 shows the percentage of samples for each sub-sampling window that are zero-valued for the FX data. These numbers can be compared with the estimated parameter value of p (defined in Section 3), which is the probability mass of zero-valued observations in the Spline-DCS model.

The discussion so far suggests that our model needs a periodic component to capture the intra-day periodic patterns. Moreover, non-periodic factors may coexist with the periodic component. One such factor is a highly persistent

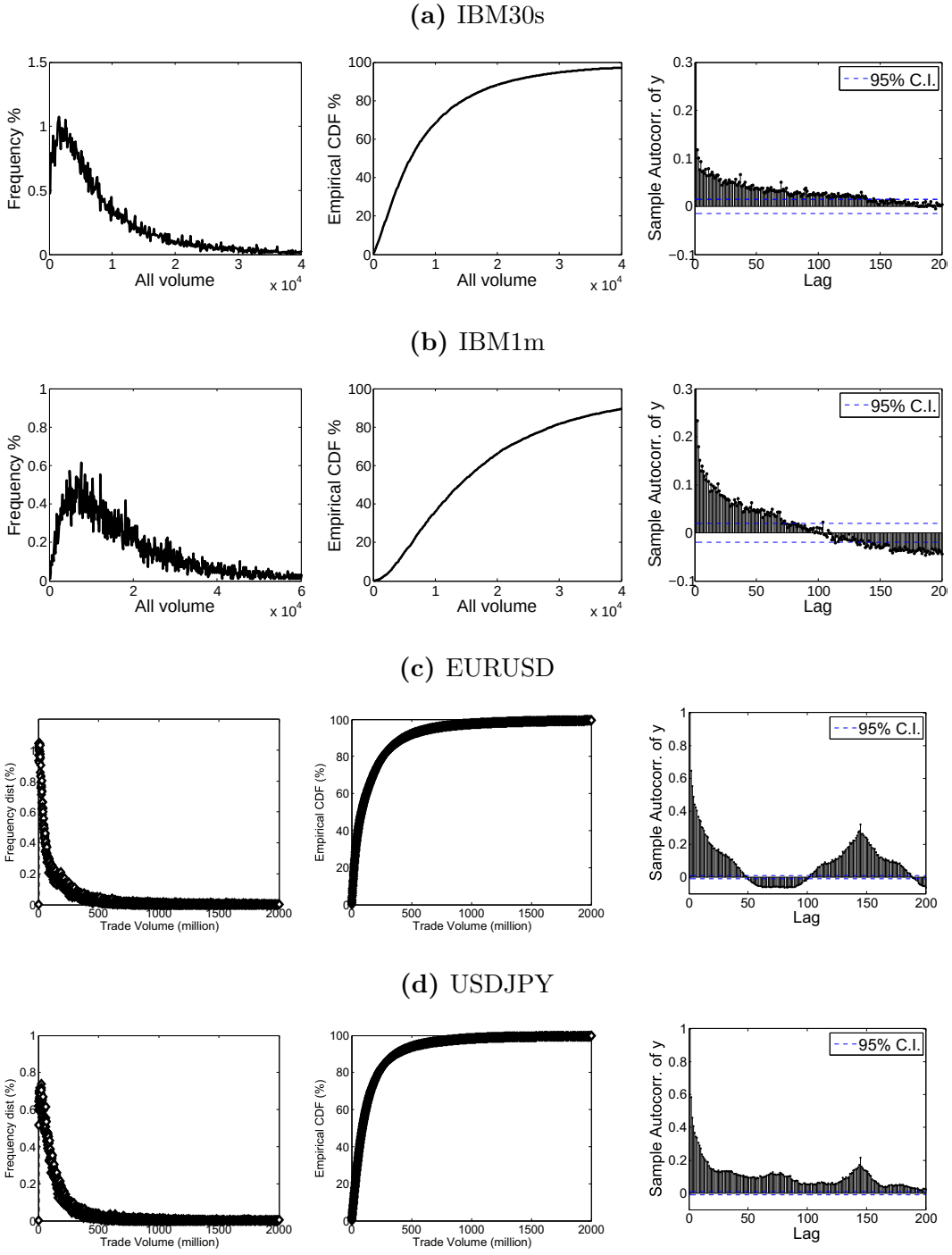


Figure 4. The empirical frequency distribution (left), the empirical cumulative distribution function (middle), and the sample autocorrelation (right). The sampling period is Mon 28 Feb - Fri 31 Mar 2000 for the equity data and Mon 6 Jan - Fri 19 Dec 2014 for the FX data. The 200th lag corresponds approximately to 1.5 hours prior for IBM30s, 3 hours prior for IBM1m, and 1.4 days prior for the FX data.

low-frequency component. The empirical autocorrelation structure suggests that we should allow for highly persistent behavior similar to long memory. This can be captured by a combination of autoregressive components. The presence of a non-negligible number of zero-valued observations can be explained by a binary component governing whether the next observation is zero or non-zero. The next section gives the formal definition of each of these components.

3 The Spline-DCS model

3.1 Notations

We divide each trading day into $I \in \mathbb{N}_{>0}$ intra-day bins. $y_{t,\tau}$ denotes the observation of trade volume at the τ-th intra-day bin on the t-th trading day for $t = 1, \ldots, T$ and $\tau = 1, \ldots, I$. $T \in \mathbb{N}_{>0}$ and $H \in \mathbb{N}_{>0}$ denote the number of in-sample and out-of-sample days, respectively. We set $y_{t,T} = y_{t+1,0}$ for all $t \in \mathbb{N}_{>0}$, so that $\tau = 1$ is the location of the first aggregated observation for each trading day. Then, over a given T consecutive trading days, we have $I \times T$ observations. Our data is generated on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $(\mathcal{F}_{t,\tau})_{t,\tau=1}^{T,I}$. Any $\mathcal{F}_{1,1}$-measurable random variables are almost surely constant. We denote in-sample estimates by $\widehat{\cdot}$ and forecast quantities by $\widetilde{\cdot}$. We use the following set notations;

$$\begin{aligned} \Psi_{T,\,I} &= \{(t,\tau) \in \{1,2,\ldots,T\} \times \{1,2,\ldots,I\}\}, \\ \Psi_{T,I>0} &= \{(t,\tau) \in \{1,2,\ldots,T\} \times \{1,2,\ldots,I\} : y_{t,\tau} > 0\}. \end{aligned}$$

3.2 Spline-DCS

Formally, our spline-DCS model is defined by:

$$
\begin{aligned}
y_{t,\tau} &= \varepsilon_{t,\tau} \exp(\lambda_{t,\tau}), \qquad \varepsilon_{t,\tau} \sim \text{i.i.d. } F(\varepsilon;\theta) \\
\lambda_{t,\tau} &= \omega + \mu_{t,\tau} + \eta_{t,\tau} + s_{t,\tau} + e_{t,\tau} \\
\mu_{t,\tau} &= \mu_{t,\tau-1} + \kappa_\mu u_{t,\tau-1}, \\
\eta_{t,\tau} &= \eta^{(1)}_{t,\tau} + \eta^{(2)}_{t,\tau}, \\
\eta^{(1)}_{t,\tau} &= \phi^{(1)}_1 \eta^{(1)}_{t,\tau-1} + \phi^{(1)}_2 \eta^{(1)}_{t,\tau-2} + \kappa^{(1)}_\eta u_{t,\tau-1} + \kappa^{(1)}_{\eta,\mathrm{a}} \mathrm{sign}(-r_{t,\tau-1})(u_{t,\tau-1} + \nu\xi), \\
\eta^{(2)}_{t,\tau} &= \phi^{(2)}_1 \eta^{(2)}_{t,\tau-1} + \kappa^{(2)}_\eta u_{t,\tau-1} + \kappa^{(2)}_{\eta,\mathrm{a}} \mathrm{sign}(-r_{t,\tau-1})(u_{t,\tau-1} + \nu\xi), \\
e_{t,\tau} &= \phi_e e_{t,\tau-1} + \boldsymbol{\kappa}_e^\top \boldsymbol{d}_{t,\tau}, \\
\boldsymbol{d}_{t,\tau} &= (d_{t,\tau,1}, \ldots, d_{t,\tau,m}), \quad d_{t,\tau,i} = \mathbb{1}_{\{\text{type } i \text{ event at time } (t,\tau)\}}, \quad i = 1, \ldots, m.
\end{aligned}
\tag{1}
$$

for $(t,\tau) \in \Psi_{T,I}$ and $\omega \in \mathbb{R}$. The non-periodic components are $\mu_{t,\tau}$ and $\eta_{t,\tau}$. $e_{t,\tau}$ is the event component. The periodic component, $s_{t,\tau}$, is defined first in Section 3.3. Then we define the distribution, F, in order to define the score variable $u_{t,\tau}$, which is another defining feature of DCS and drives the dynamics of the non-periodic components. Then we define the rest of the components.

3.3 The cubic spline

The periodic component, $s_{t,\tau}$, captures the pattern of intra-day periodicity. This is an application of the seminal work by Harvey and Koopman (1993). We refer to the version of the spline studied here as the *static* (cubic) spline, which assumes that the pattern of periodicity does not change over time. This is a standard assumption in the existing literature. Ito (2013) challenges this assumption by introducing a generalized Spline-DCS specification with the dynamic version of the cubic spline. The empirical merit of such a generalization in high-frequency finance is illustrated by Ito (2013).

Some of the technical details are omitted in the following sections, but we give the complete mathematical construction in Appendix B.

3.3.1 Static daily spline

The cubic spline is termed a *daily* spline if the periodicity is complete over one trading day. The static daily spline assumes that the shape of intra-day periodic patterns is the same every day. The daily spline is a continuous piecewise function of time and connected at $k+1$ knots for some $k \in \mathbb{N}_{>0}$ such that $k < I$. The coordinates of the knots along the time axis are denoted by $\tau_0 < \cdots < \tau_k$, where $\tau_0 = 1$, $\tau_k = I$, and $\tau_j \in \{2, \ldots, I-1\}$ for $j = 1, \ldots, k-1$. The set of the

knots is also called *mesh*. The y-coordinates (height) of the knots are denoted by $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_k)^\top$. The static daily spline ($s_{t,\tau} = s_\tau$) is defined as

$$s_\tau = \sum_{j=1}^{k} \mathbb{1}_{\{\tau \in [\tau_{j-1}, \tau_j]\}} \, \boldsymbol{z}_j(\tau) \cdot \boldsymbol{\gamma}, \qquad \tau = 1, \ldots, I, \tag{2}$$

where $\boldsymbol{z}_j : [\tau_{j-1}, \tau_j]^{k+1} \to \mathbb{R}^{k+1}$ for $j = 1, \ldots k$ is a $(k+1)$-dimensional vector of deterministic functions that conveys all information about the *polynomial order*, *continuity*, and *zero-sum conditions* of the spline. See Appendix B for the definition of these conditions and the derivation of $\boldsymbol{z}_j(\tau)$. The zero-sum condition ensures that the parameters in $\boldsymbol{\gamma}$ are identified. To impose the zero-sum condition, we also need to set $\gamma_k = -\sum_{i=0}^{k-1} w_{*i}\gamma_i / w_{*k}$, where $\boldsymbol{w}_* = (w_{*0}, \ldots, w_{*k})^\top$ is defined in the appendix.

For the equity data, we capture the overnight effect that arises from the regular overnight market closure by relaxing the *periodicity condition* of the spline, and allowing for a discrepancy in the spline between the end and the beginning of any two consecutive trading days. That is, we assume that $(\tau_k, \gamma_k) \neq (\tau_0, \gamma_0)$. The definition of $\boldsymbol{z}_j(\tau)$ given above is this case, and it is different from the one defined by Harvey and Koopman (1993). See Appendices B.1 and B.2. Harvey and Koopman (1993) impose the periodicity condition, since their hourly electricity demand data is collected 24 hours a day. This feature is the same as the FX data. Thus, the spline we use for the FX data maintains the periodicity condition, and it is fully defined in Appendices B.3 and B.4. (2), (3), (4), and Figure 5 are the version for the equity data and capture the overnight effect.

3.3.2 Location of daily knots and overnight effect

The location of knots, $\tau_1, \ldots, \tau_{k-1}$, and the size of k depend on the empirical shape of periodicity and the number of intra-day observations. Increasing k does not necessarily improve the fit of the model, and using too many knots deteriorates the speed of computation. We give tips on how to select the location and the number of knots in Section 7.

For the FX data, based on the empirical observations we made in Section 2 and from Figure 3, we find that the following locations of knots (in hours) along the intra-day time axis works well:

`1, 2, 3.30, 5, 6, 7, 8, 9.30, 11, 12, 13, 14, 15, 16, 17.30, 19, 20, 21, 22, 23, 24.`

For the FX data, the period between Friday 10pm and Sunday 10pm is treated as the regular period of missing data over the weekend. Although there is no

regular and decisive moments on Friday and Sunday at which transactions end and begin for the week, we omit data points over the specified weekend period for simplicity. Then, in the static daily spline for FX, we impose the periodicity condition on the knots at Friday 10pm and Sunday 10pm.

For the equity data, we find the following location of knots (in hours) along the intra-day time axis works well:

`9.30, 11, 12.30, 14.30, 16`

The shape of the spline up to 12.30pm captures the busy trading hours in the morning, between 12.30pm and 2.30pm captures the quiet lunch hours, and after 2.30pm captures any acceleration in trading activities before the market closes. There is little to no improvement in the goodness of fit of the model to the data when the number of knots per day increases from this specification.

3.3.3 Static weekly spline

The static spline becomes a static *weekly spline* if we set the periodicity to be complete over one trading week instead of one day. The static weekly spline allows the shape of periodic patterns to be different for each weekday, but it assumes that the overall shape for the whole week cannot change from week to week.

For this version, we redefine $\tau_0, \tau_1, \ldots, \tau_k$ as follows. We let $\widetilde{\tau}_0 < \widetilde{\tau}_1 < \cdots < \widetilde{\tau}_{k'}$ denote the coordinates along the time-axis of the *intra-day mesh*, where $k' < I$, $\widetilde{\tau}_0 = 1$, $\widetilde{\tau}_{k'} = I$, and $\widetilde{\tau}_j \in \{2, \ldots, I-1\}$ for $j = 1, \ldots, k'-1$. Then the coordinates, $\tau_0, \tau_1, \ldots, \tau_k$, along the time-axis of the total mesh for the whole week is defined as $\tau_{i(k'+1)+j} = \widetilde{\tau}_j + iI$ for $i = 0, \ldots, 4$ and $j = 0, \ldots, k'$. Then $(\tau_j)_{j=0}^k$ is still an increasing sequence. The total number of knots for one whole week is $k+1 = 5(k'+1)$. The height of the knots are $\gamma_0, \gamma_1, \ldots, \gamma_{k'}$ for Monday, $\gamma_{k'+1}, \gamma_{k'+2}, \ldots, \gamma_{2(k'+1)}$ for Tuesday, and so on.

Here, we do not impose periodicity condition between τ_k and τ_0 in order to allow for the effect of weekend news on trading patterns. (Also see Section 3.3.4.) Moreover, for the equity data, we capture the overnight effect of weeknights by relaxing the continuity and polynomial order restrictions between $\widetilde{\tau}_{k'}$ and $\widetilde{\tau}_0$ of any two successive weekdays. Thus, the procedure for computing $\boldsymbol{z}_j(\tau)$ is different from the daily spline. See Appendix B.2 for the complete specification of the weekly spline with these features for the equity data. For the FX data, we maintain the continuity and polynomial order conditions on $\widetilde{\tau}_{k'}$ and $\widetilde{\tau}_0$. Then the number of knots for the whole week is $k+1 = 5k'+1$. See Appendix B.4.

The weekly spline can be used to capture the day-of-the-week effect. It allows

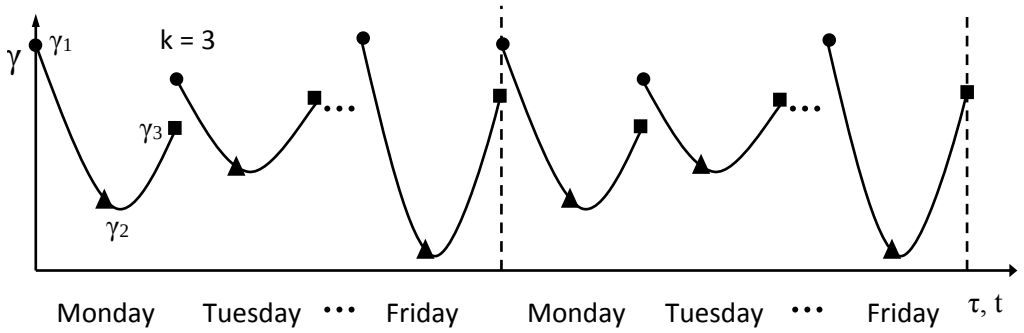


Figure 5. Picture illustration of the static weekly spline with overnight effect and weekend effect when the number of knots per day is $k = 3$. The shape depends on the day of the week.

for the overall shape of the pattern of periodicity to depend on the day of the week by varying the height of the knots for different weekdays. That is, the weekly spline allows for

$$(\gamma_0, \gamma_1, \ldots, \gamma_{k'}) \neq (\gamma_{k'+1}, \gamma_{k'+2}, \ldots, \gamma_{2(k'+1)}) \neq \cdots \neq (\gamma_{4k'+4}, \gamma_{4k'+5}, \ldots, \gamma_{5(k'+1)}).$$

See Figure 5.

3.3.4 Restricted weekly spline

One disadvantage of the weekly spline is that the total number of knots for one week increases quickly (fivefold) with the number of daily knots. If we use 5 knots per trading day ($k' = 4$) as we specified for the daily spline, the total number of knots for the week is 25 ($k = 24$) for the weekly spline with the overnight effect. Estimating such a high number of knots can be computationally costly.

Instead of letting the coordinates of the mesh be free for each weekday, we can restrict them to be the same on selected days. We term this special case the *restricted weekly spline*. We restrict the pattern to be the same on mid-weekdays (Tuesday-Thursday) and let the pattern be different on Monday and Friday. This restricted weekly spline captures a special type of the day-of-the-week effect called the *weekend effect*, which, in the context of this paper, refers to the tendency of trade volume before and after the weekend to display distinct patterns compared to mid-weekdays. This effect arises from news disseminated over the weekend period.

Formally, we define (the static version of) the restricted weekly spline as:

$$s_{t,\tau} = \sum_{j=1}^{k} \mathbb{1}_{\{I\times(t-1)\bmod(5)+\tau\in[\tau_{j-1},\tau_j]\}}\, \boldsymbol{z}_j(\tau)\cdot \mathbf{S}\boldsymbol{\gamma} \tag{3}$$

where $\boldsymbol{\gamma} = (\widetilde{\boldsymbol{\gamma}}_1^\top, \widetilde{\boldsymbol{\gamma}}_2^\top, \widetilde{\boldsymbol{\gamma}}_3^\top)^\top$ and $\widetilde{\boldsymbol{\gamma}}_i$ for $i = 1, 2, 3$ are $(k'+1)$-dimensional mesh vectors for Monday, mid-weekdays (Tuesday-Thursday), and Friday, respectively, and $\boldsymbol{z}_j(\tau)$ is of the weekly spline. $\mathbf{S}$ is the following $5(k'+1) \times 3(k'+1)$ matrix of zeros and ones:

$$\mathbf{S} = \begin{bmatrix} \mathbf{I}_{(k'+1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{(k'+1)} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{I}_{(k'+1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{(k'+1)} \end{bmatrix}.$$

where $\mathbf{I}_{(k'+1)}$ is the identity matrix of size $(k'+1)$. We can rewrite (3) as

$$s_{t,\tau} = \sum_{j=1}^{k} \mathbb{1}_{\{I \times (t-1) \bmod (5) + \tau \in [\tau_{j-1}, \tau_j]\}} \, \widetilde{\boldsymbol{z}}_j(\tau) \cdot \boldsymbol{\gamma} \tag{4}$$

where $\widetilde{\boldsymbol{z}}_j(\tau) = \mathbf{S}^\top \boldsymbol{z}_j(\tau)$. If we place 5 knots per trading day (i.e. $k' = 4$) as we specified for the daily spline for the equity data, we have 15 unrestricted knots for one week.

The evidence for the day-of-the-week effect in financial data is generally mixed and largely depends on the estimation method and the variable being estimated. For instance, Andersen and Bollerslev (1998) found that the day-of-the-week effect is insignificant in the deutsche mark-dollar FX returns once the calendar effect (e.g. daylight saving and public holidays) and the effects of major macroeconomic announcements are taken into account. However, they found indications of a weak (but clear) seasonality on Monday mornings and Friday afternoons. Lo and Wang (2010) found that some financial assets exhibit a strong day-of-the-week effect. Their volume data as measured by turnover is roughly constant on all days except on Mondays and Fridays when turnover is slightly lower than mid-weekdays. The weekend effect is studied in many other studies (mainly in the context of asset returns) and appears to be more pronounced than the more general day-of-the-week effect. Thus, given the computational cost of the fully flexible weekly spline, estimating the restricted weekly spline may suffice for us.

3.4 Dealing with zero-valued observations

The cumulative distribution function (c.d.f.), $F : \mathbb{R}_{\geq 0} \to [0, 1]$, with the constant parameter vector θ of a standard random variable $X \sim F$ is defined as

$$\mathbb{P}_F(X = 0) = p \in (0, 1), \quad \mathbb{P}_F(X > 0) = 1 - p, \quad \mathbb{P}_F(X \leq x | X > 0) = F^*(x; \theta^*).$$

for any $x > 0$. $F^* : \mathbb{R}_{\geq 0} \to [0, 1]$ is the c.d.f. of some conventional standard continuous random variable with the time-invariant parameter vector θ^*. We write $\theta = (\theta^{*\top}, p)^\top$ and use the notations, f and f^*, to denote the probability density function (p.d.f.) of F and F^*, respectively. The distribution, F, explicitly deals with zero-valued observations, and it is useful given that the presence of a non-negligible number of zero-valued observations cannot be explained by conventional continuous distributions as the probability of observing a particular value is typically zero by definition. The unconditional n-th moment of X is well-defined as long as it is well-defined for F^* because $\mathbb{E}[X^n] = (1-p)\int_0^\infty x^n f^*(x)dx$. F^* is chosen parametrically, and the quality of its fit to the empirical distribution of data is tested using the standardized observations $\widehat{\varepsilon}_{t,\tau} \equiv y_{t,\tau}/\exp(\widehat{\lambda}_{t,\tau})$. The properties of this type of distributions are studied formally in Hautsch et al. (2014). This decomposition technique is a standard one in econometrics and similar to the ones studied in McCulloch and Tsay (2001) and Rydberg and Shephard (2003).

$u_{t,\tau}$ is the score of F^*. We set $u_{t,\tau} = \inf_{s\in\Omega} u_{t,\tau}(s)$ whenever $y_{t,\tau} = 0$. This means that, if F^* is the generalized beta distribution of the second kind (GB2) formally defined in Appendix A.1, we set $u_{t,\tau} = -\nu\xi$ whenever $y_{t,\tau} = 0$. It is natural to suspect that p may change throughout the day because the probability of observing a trade must change with the level of trading activity. Rydberg and Shephard (2003) and Hautsch et al. (2014) independently study decomposition models for estimating the conditional dynamics of p via the logit link. An interesting extension of our model is a hybrid Spline-DCS model for the intra-day dynamics of p as well as the level of volume. However, in this paper, we assume p to be constant for simplicity and leave this extension for future research. This simplification is inconsequential in our application as the fraction of samples that are zero-valued is small.

3.5 Non-periodic components

The stationary component is $\eta_{t,\tau}$. $\eta_{t,\tau}$ consists of two stationary components, $\eta^{(1)}_{t,\tau}$ and $\eta^{(2)}_{t,\tau}$. This structure allows us to capture highly persistent dynamics similar to long memory.[5] $\eta^{(1)}_{t,\tau}$ and $\eta^{(2)}_{t,\tau}$ are stationary if $-\phi^{(1)}_1 + \phi^{(1)}_2 < 1$, $\phi^{(1)}_2 > -1$,

[5] $\eta_{t,\tau}$ can be generalized as:

$$\begin{aligned}
\eta_{t,\tau} &= \sum_{j=1}^{J} \eta^{(j)}_{t,\tau}, \qquad \forall\,(t,\tau) \in \Psi_{T,\,I} \\
\eta^{(j)}_{t,\tau} &= \phi^{(j)}_1 \eta^{(j)}_{t,\tau-1} + \phi^{(j)}_2 \eta^{(j)}_{t,\tau-2} + \cdots + \phi^{(j)}_{m^{(j)}} \eta^{(j)}_{t,\tau-m^{(j)}} + \kappa^{(j)}_\eta u_{t,\tau-1}, \qquad j = 1, \ldots, J
\end{aligned}$$

$0 < \phi_1^{(1)} + \phi_2^{(1)} < 1$, and $0 < \phi_1^{(2)} < 1$ (see, for instance, Harvey (1993, p.19)). The non-stationary component, $\mu_{t,\tau}$, captures the slowly changing movements that is non-periodic. The estimation results in Section 4 suggest that this component can do a good job in capturing the low-frequency dynamics of our data.

The role of each component is such that $\mu_{t,\tau}$ should be less sensitive to changes in $u_{t,\tau-1}$ than $\eta_{t,\tau}^{(1)}$, which should be, in turn, less sensitive than $\eta_{t,\tau}^{(2)}$. That is, we should typically expect $|\kappa_\mu| < |\kappa_\eta^{(1)}| < |\kappa_\eta^{(2)}|$ (although this condition is not imposed during the estimation). Moreover, the scale of trade volume should increase in the wake of positive news. Thus we would expect $\kappa_\mu > 0$. We set $\eta_{1,1}^{(1)} = \eta_{1,1}^{(2)} = 0$ as we have $\mathbb{E}[\eta_{t,\tau}^{(1)}] = \mathbb{E}[\eta_{t,\tau}^{(2)}] = 0$.[6] Since $\mathbb{E}[\mu_{t,\tau}] = \mu_{1,1}$, we assume $\mu_{1,1} = 0$ so that ω is identified. The identification conditions of the parameters in $s_{t,\tau}$ are as laid out in Section 3.3.

3.5.1 Asymmetric effect

For the equity data, analogously to the well-documented leverage effects in equity return volatility, we can test for asymmetric effects in volume related to the direction of price change by testing the significance of the coefficients, $\kappa_{\eta,\mathrm{a}}^{(1)}$ and $\kappa_{\eta,\mathrm{a}}^{(2)}$. $\kappa_{\eta,\mathrm{a}}^{(1)} > 0$ (or $\kappa_{\eta,\mathrm{a}}^{(1)} < 0$) gives an increase (decrease) in the scale of volume when price falls (i.e. when the return, $r_{t,\tau}$, which is the log-difference in price, is negative). We use the sign function (instead of the indicator function) in order to capture the asymmetric effect of price change in both the positive and negative directions. That is, the sign function with $\kappa_{\eta,\mathrm{a}}^{(i)} > 0$ (or $\kappa_{\eta,\mathrm{a}}^{(i)} < 0$) gives a decrease (increase) in the scale of volume when price increases for $i = 1, 2$. How to model leverage effects in the DCS models is discussed in Harvey (2013).

For the FX data, the ways in which we should assess the existence and the directional impact of such an effect is less straight forward than the equity data. For instance, a fall in the exchange rate of the US dollar per one euro means that euro depreciated against the dollar, but also that the dollar appreciated against euro. Then a sudden sizable strengthening of euro against the dollar can cause panic and trigger asymmetric effects if the market generally expected the dollar to stay strong against euro. In this view, the statistical significance of this effect, and the sign of the asymmetry term if the effect exists, must depend on market expectation. Since we do not have a measure of market expectation or any other coherent measure to assess whether estimation results for the asymmetry term

for some $J \in \mathbb{N}_{>0}$. We assume that $m^{(j)} \in \mathbb{N}_{>0}$ and $\eta_{t,\tau}^{(j)}$ is stationary for all $j = 1, \ldots, J$. $J = 2$ works well for our application.

[6] With the asymmetry terms, this assumes that $r_{t,\tau}$ and $y_{t,\tau}$ are independent for any $(t, \tau) \in \Psi_{T,I}$, and $\mathbb{E}[r_{t,\tau}] = 0$.

Currency	EURUSD		USDJPY	
Event category	Count	Frequency (%)	Count	Frequency (%)
Small	1024	93%	543	90%
Intermediate	65	6%	51	8%
US non-farm payroll	12	1%	12	2%

Table 5. Event schemes between Monday 6 January - Sunday 21 December 2014.

makes sense, we set $\kappa_{\eta,\mathrm{a}}^{(i)} = 0$ for $i = 1, 2$ for the FX data for simplicity.

3.5.2 Announcement effect

The event component, $e_{t,\tau}$, captures the effect of anticipated macroeconomic events. Its dynamics are assumed to be deterministic. Any deviation of market response from the deterministic pattern at each event is assumed to be captured by other non-deterministic components. We set $e_{1,1} = 0$. $e_{t,\tau}$ reverts to zero if $|\phi_e| < 1$.

In FX, major macroeconomic events that affect our selected currency pairs include monetary policy announcements and macroeconomic data releases (e.g. US non-farm payrolls, gross domestic product (GDP), and consumer prices). Since there are usually many events per day that can impact our currency pairs, we categorize events by the anticipated size of the impact as tabulated in Table 5. We used the information provided in the Forex Economic Calendar by DailyFX (www.dailyfx.com) as a benchmark to collate events. Then we assign a dummy variable $(d_{t,\tau,i}, i = 1, \ldots, m)$ to each category. The small event category include scheduled releases of retail sales data, manufacturing data, home sales data, and various indicators of house prices. The intermediate events category include some of the central bank announcements (e.g. by the European Central Bank and the Federal Reserve Board for EURUSD) and the release of GDP data, employment figures, and consumer price data by governments in the relevant currency areas. The release of US non-farm payroll data on the first Friday of each month is assigned its own category. The elements of $\boldsymbol{d}_{t,\tau}$ are ordered by the anticipated size of the impact. The size of m varies across the sub-sample windows. The first element of $\boldsymbol{d}_{t,\tau}$ correspond to the event category with the largest anticipated impact for that window. Multiple events usually occur simultaneously. If there are more than one small event occurring simultaneously, they are treated as one small event. We do the same for events in other categories.

For the equity data, the events that affect the trade volume of the IBM stock

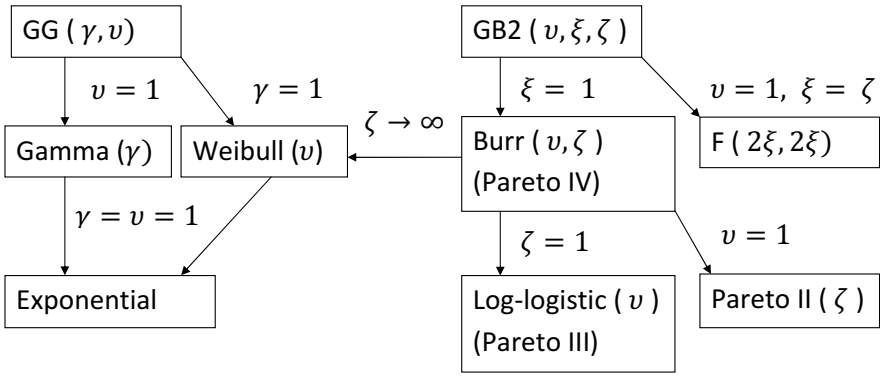


Figure 6. Nested diagram of some of the useful non-negative distributions. The scale and location parameters are assumed to be one and zero, respectively, in all of these cases.

include the company's quarterly or annual earnings and dividend announcements, the earnings and dividend announcements by the competitors (e.g. Accenture, Hewlett-Packard, and Microsoft), and important news in the technology industry. To our knowledge, IBM did not make any earnings or dividend related announcements during the sampling period. Since the sampling period of our data is relatively old, we could not find the exact timing of news releases in intra-day hours for all of the companies mentioned here. Due to the limitation with acquiring information, we remove the event component, $e_{t,\tau}$, for the equity data for simplicity.

3.6 The estimation method

All of the parameters of the model are estimated by ML. For non-negative series, F^* can be a number of distributions including Weibull, Gamma, Burr, and log-normal, many of which are special cases of the generalized gamma (GG) and GB2 distributions. See Figure 6 and, more formally, Appendix A and Kleiber and Kotz (2003) for their definitions and the relationship between these distributions.

The joint log-likelihood function given by F for the set of observations $(y_{t,\tau})_{(t,\tau)\in\Psi_{T,I}}$ is

$$\log L = A\log(1-p) + (T\times I - A)\log(p) + \sum_{(t,\tau)\in\Psi_{T,I>0}} \log\left(\exp(-\lambda_{t,\tau}) f^*(\varepsilon_{t,\tau};\boldsymbol{\theta}^*)\right), \tag{5}$$

where $A = |\Psi_{T,I>0}|$. It is easy to check that the ML estimator (MLE) of p is $\widehat{p} = (T\times I - A)/(T\times I)$. We do not formally verify that the consistency and the asymptotic normality of MLE in DCS (along the lines of the discussions by Ito

(2016) and by Blasques et al. (2014) and Harvey (2013)) also extend to our particular specification in which the exponential link function is used and $\lambda_{t,\tau}$ is nonstationary. We simply assume that they hold, and leave the verification for future research.

3.6.1 Derivatives for standard errors

Let $\boldsymbol{\vartheta}$ denote the vector of all of the constant parameters of Spline-DCS including the distribution parameters in θ. Denoting the single log-likelihood of an observation, $y_{t,\tau}$, as $\log f_Y(y_{t,\tau};\boldsymbol{\vartheta})$,[7] we compute the standard error of MLE for the i-th element of $\boldsymbol{\vartheta}$ (denoted by $\widehat{\boldsymbol{\vartheta}}_i$) using the outer-product of the first-derivative of the joint log-likelihood as:

$$\text{S.E.}(\widehat{\boldsymbol{\vartheta}}_{\text{i}}) = \sqrt{\left(\sum_{(t,\tau)\in\Psi_{T,I}} \frac{\partial \log f_Y(y_{t,\tau};\widehat{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\vartheta}} \frac{\partial \log f_Y(y_{t,\tau};\widehat{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\vartheta}^\top}\right)^{-1}_{ii}},$$

where $\cdot_{ii}$ for $i = 1,\ldots,\dim(\boldsymbol{\vartheta})$ denotes the i-th diagonal element.

If F^* is GB2, $\boldsymbol{\vartheta}$ includes $\theta = (\theta^{*\top}, p)^\top$ with $\theta^* = (\nu,\xi,\zeta)^\top$, as well as the parameters of $\lambda_{t,\tau}$. Then we have the following derivatives for the parameters in θ:

$$\begin{aligned}
\frac{\partial \log f_Y(y_{t,\tau};\boldsymbol{\vartheta})}{\partial \nu} &= \frac{1}{\nu} + \xi \log(y_{t,\tau}e^{-\lambda_{t,\tau}}) + u_{t,\tau}\frac{\partial \lambda_{t,\tau}}{\partial \nu},\\
\frac{\partial \log f_Y(y_{t,\tau};\boldsymbol{\vartheta})}{\partial \xi} &= \log(b_{t,\tau}) - \frac{1}{B(\xi,\zeta)}\frac{\partial B(\xi,\zeta)}{\partial \xi} + u_{t,\tau}\frac{\partial \lambda_{t,\tau}}{\partial \xi},\\
\frac{\partial \log f_Y(y_{t,\tau};\boldsymbol{\vartheta})}{\partial \zeta} &= -\frac{1}{B(\xi,\zeta)}\frac{\partial B(\xi,\zeta)}{\partial \zeta} + \log(1-b_{t,\tau}) + u_{t,\tau}\frac{\partial \lambda_{t,\tau}}{\partial \zeta},\\
\frac{\partial \log f_Y(y_{t,\tau};\boldsymbol{\vartheta})}{\partial p} &= \mathbb{1}_{\{y_{t,\tau}=0\}}/p - \mathbb{1}_{\{y_{t,\tau}>0\}}/(1-p),
\end{aligned}$$

where $u_{t,\tau} \equiv \nu(\xi+\zeta)b_{t,\tau} - \nu\xi$ is the score of GB2 and $b_{t,\tau} \equiv 1/(1+(y_{t,\tau}\exp(-\lambda_{t,\tau}))^{-\nu})$. (See (A.1) and (A.2) in Appendix A.1). We also have $\partial B(\xi,\zeta)/\partial\xi = B(\xi,\zeta)(\psi(\xi) - \psi(\xi+\zeta))$ with $\partial B(\xi,\zeta)/\partial\zeta = \partial B(\zeta,\xi)/\partial\zeta$ by the symmetry of the Beta function. $\psi(\cdot)$ is the digamma function. For other parameters of $\boldsymbol{\vartheta}$ (denoted by $\boldsymbol{\vartheta}_{-\theta}$), we have

$$\frac{\partial \log f_Y(y_{t,\tau};\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}_{-\theta,i}} = u_{t,\tau}\frac{\partial \lambda_{t,\tau}}{\partial \boldsymbol{\vartheta}_{-\theta,i}},$$

where $\boldsymbol{\vartheta}_{-\theta,i}$ is the i-th element of $\boldsymbol{\vartheta}_{-\theta}$. Denoting the i-th element of $\boldsymbol{\vartheta}$ by $\boldsymbol{\vartheta}_i$, the

[7] The relationship between f_Y and f are given in Appendix A.

derivatives of $\lambda_{t,\tau}$ are given by the following recursions:

$$\begin{aligned}
\frac{\partial \lambda_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} &= \mathbb{1}_{\{\boldsymbol{\vartheta}_i=\omega\}} + \frac{\partial \mu_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} + \frac{\partial \eta_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} + \frac{\partial s_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} + \frac{\partial e_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} \\
\frac{\partial \mu_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} &= \frac{\partial \mu_{t,\tau-1}}{\partial \boldsymbol{\vartheta}_i} + u_{t,\tau}\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\kappa_\mu\}} + \kappa_\mu \frac{\partial u_{t,\tau-1}}{\partial \boldsymbol{\vartheta}_i} \\
\frac{\partial \eta_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} &= \frac{\partial \eta^{(1)}_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} + \frac{\partial \eta^{(2)}_{t,\tau}}{\partial \boldsymbol{\vartheta}_i}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \eta^{(j)}_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} &= \eta^{(j)}_{t,\tau-1}\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\phi^{(j)}_1\}} + \eta^{(j)}_{t,\tau-2}\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\phi^{(1)}_2\}} + u_{t,\tau-1}\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\kappa^{(j)}_\eta\}} \\
&\quad + \phi^{(j)}_1 \frac{\partial \eta^{(j)}_{t,\tau-1}}{\partial \boldsymbol{\vartheta}_i} + \phi^{(j)}_2 \frac{\partial \eta^{(j)}_{t,\tau-2}}{\partial \boldsymbol{\vartheta}_i} + \kappa^{(j)}_\eta \frac{\partial u_{t,\tau-1}}{\partial \boldsymbol{\vartheta}_i} \\
&\quad + \operatorname{sign}(-r_{t,\tau})(u_{t,\tau-1} + \nu\xi)\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\kappa^{(j)}_{\eta,a}\}} \\
&\quad + \kappa^{(j)}_{\eta,a}\operatorname{sign}(-r_{t,\tau})\left(\frac{\partial u_{t,\tau-1}}{\partial \boldsymbol{\vartheta}_i} + \xi\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\nu\}} + \nu\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\xi\}}\right), \quad j = 1, 2, \\
\frac{\partial e_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} &= e_{t,\tau-1}\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\phi_e\}} + \phi_e \frac{\partial e_{t,\tau-1}}{\partial \boldsymbol{\vartheta}_i} + d_{t,\tau,m}\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\kappa_{e,m}\}}, \quad m = 1, \ldots, \dim(\boldsymbol{d}_{t,\tau}),
\end{aligned}$$

where $\kappa_{e,m}$ is m-th element of $\boldsymbol{\kappa}_e$. As for the spline component, we have

$$\frac{\partial s_\tau}{\partial \boldsymbol{\vartheta}_i} = \sum_{j=1}^{k} \mathbb{1}_{\{\tau\in[\tau_{j-1},\tau_j]\}}\boldsymbol{z}_{j,l}(\tau)\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\gamma_l\}},$$

if it is a static daily spline, and

$$\frac{\partial s_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} = \sum_{j=1}^{k} \mathbb{1}_{\{I\times(t-1)\bmod(5)+\tau\in[\tau_{j-1},\tau_j]\}}\widetilde{\boldsymbol{z}}_{j,l}(\tau)\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\gamma_l\}},$$

if it is a restricted weekly spline, where $l = 0, \ldots, k-1$ and k is the number of knots. Given our assumption about $\mathcal{F}_{1,1}$, these recursions are assumed to begin from zero. Finally, for the score variable, we have

$$\begin{aligned}
\frac{\partial u_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} &= ((\xi+\zeta)b_{t,\tau} - \xi)\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\nu\}} + \nu(b_{t,\tau}-1)\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\xi\}} + \nu b_{t,\tau}\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\zeta\}} + \nu(\xi+\zeta)\frac{\partial b_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} \\
\frac{\partial b_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} &= \begin{cases} -b_{t,\tau}(1-b_{t,\tau})\log(y_{t,\tau}e^{-\lambda_{t,\tau}})(y_{t,\tau}e^{-\lambda_{t,\tau}})\frac{\partial \lambda_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} & \text{if } \boldsymbol{\vartheta}_i = \nu, \\ -\nu b_{t,\tau}(1-b_{t,\tau})\frac{\partial \lambda_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} & \text{otherwise.} \end{cases}
\end{aligned}$$

3.6.2 Market closure

For the equity data, public holidays can be treated in the same way as the overnight period or weekends where data is missing due to market closure. Whenever $(t,\tau) \in \mathcal{H} \equiv \{(t,\tau) \in \{1,\ldots,I\} \times \{1,\ldots,T\} : \text{the market is closed}\}$, we set $\mu_{t,\tau} = \mu_{t,\tau-1}$, $\eta^{(1)}_{t,\tau} = \eta^{(1)}_{t,\tau-1}$, $\eta^{(2)}_{t,\tau} = \eta^{(2)}_{t,\tau-1}$, and $s_{t,\tau} = s_{t,\tau-1}$ so that $\lambda_{t,\tau}$ is unchanged. The joint log-likelihood is defined only for days $t \in \mathcal{H}^c$.

4 In-sample estimation results

4.1 Choice of distribution

Given the shape of the empirical distribution of data, the candidate distributions we considered for F^* include the log-normal distribution, the Pareto distributions, GB2 and its special cases, and GG and its special cases. GB2 nests log-logistic and Burr, and is closely related to the Pareto class of distributions. See Appendix A and Figure 6 for the formal definitions of these distributions and the relationship between them. Taking a general-to-specific approach, we sequentially estimate GB2, Burr, and then log-logistic for the GB2 class of distributions. As for the GG class of distributions, we sequentially estimate GG, and then Gamma and Weibull.

For the FX data, we found that choosing GB2 as the error distribution, F^*, gives a good fit to the empirical distribution of the data.[8] Burr, which is a special case of GB2 with $\xi = 1$, was found to fit the empirical distribution of the equity data well.[9]

Figures 7 and 8 illustrate the impressive fit of the chosen F^*. The empirical c.d.f. of non-zero $\widehat{\varepsilon}_{t,\tau}$ appears to overlap the c.d.f. of GB2$(\widehat{\nu}, \widehat{\xi}, \widehat{\zeta})$ or Burr$(\widehat{\nu},\widehat{\zeta})$.[10] The closeness of the fit can be also checked by inspecting the the probability integral transform (PIT) of non-zero $\widehat{\varepsilon}_{t,\tau}$ computed under the assumption that it is from $F^*(\cdot;\widehat{\theta}^*)$. The empirical c.d.f. of the PIT values lies along the diagonal, indicating that the PIT values are close to being standard uniformly distributed (denoted by $U[0,1]$). The results are similar for IBM1m and other sampling windows that are tested for the FX data. The model appeared to be robust to the choice of sampling frequency, sampling period, or the choice of initial parameter values. The computing time for the ML estimation to converge when F^* is GB2 or Burr is generally short (up to 5 minutes). This suggests that the theoretical moments and quantiles of GB2$(\widehat{\nu}, \widehat{\xi}, \widehat{\zeta})$ or Burr$(\widehat{\nu}, \widehat{\zeta})$ scaled by $\exp(\widehat{\lambda}_{t,\tau})$ should approximate the time-varying (conditional) moments and

[8] The score variable $(u_{t,\tau})$ for when F^* is GB2 is defined in Appendix A.1. $\theta^* = (\nu, \xi, \zeta)^\top$ are the shape parameters of GB2.

[9] For the equity case, we found it difficult to estimate the GB2 distribution. This could be due to the relatively high sampling frequency. Some of the features in the empirical distribution of data that arise from market microstructure (e.g. the discreteness of volume) may have become more dominant when the sampling frequency is as high as 1 minute or 30 seconds. In such cases, imposing restrictions on some of the distribution parameters and testing relatively more specific distribution specifications can work well.

[10] The large sample size would make formal tests such as the Kolgomorov-Smirnov test sensitive to small departures from the theoretical distribution being tested.

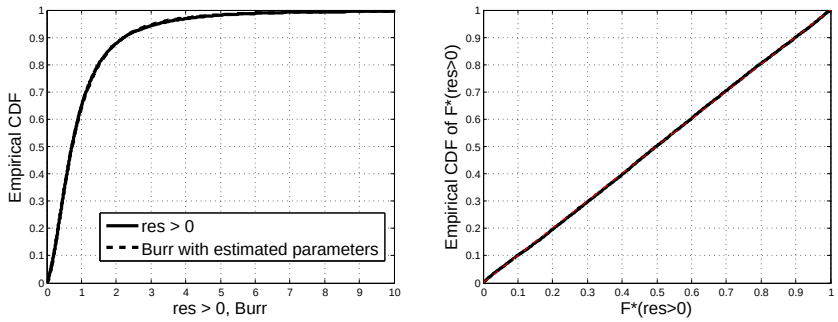


Figure 7. Fitting Burr to IBM30s over 28 February - 31 March 2000. The empirical c.d.f. of $\widehat{\varepsilon}_{t,\tau} > 0$ plotted against the theoretical c.d.f. of Burr$(\widehat{\nu}, \widehat{\zeta})$ (left). The empirical c.d.f. of the PIT values of $\widehat{\varepsilon}_{t,\tau} > 0$ when $F^*(\cdot; \widehat{\boldsymbol{\theta}}^*)$ is Burr$(\widehat{\nu}, \widehat{\zeta})$ (right). Spline-DCS with the static daily spline was used.

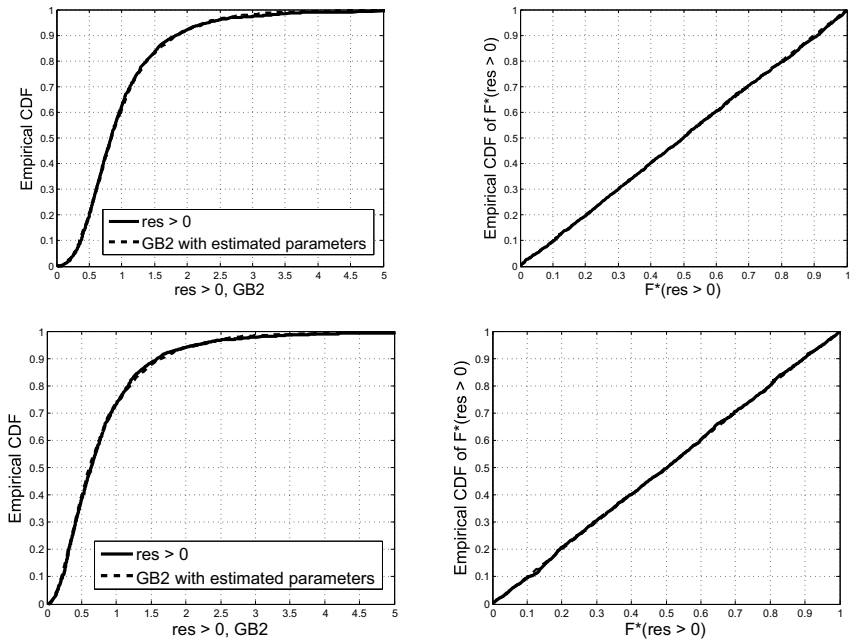


Figure 8. Fitting GB2 to the trade volume of EURUSD (top) and USDJPY (bottom) in Window 1. The empirical c.d.f. of $\widehat{\varepsilon}_{t,\tau} > 0$ plotted against the theoretical c.d.f. of GB2$(\widehat{\nu}, \widehat{\xi}, \widehat{\zeta})$ (left). The empirical c.d.f. of the PIT values of $\widehat{\varepsilon}_{t,\tau} > 0$ when $F^*(\cdot; \widehat{\boldsymbol{\theta}}^*)$ is GB2$(\widehat{\nu}, \widehat{\xi}, \widehat{\zeta})$ (right). Spline-DCS with the static daily spline was used.

quantiles of volume well.

4.2 Autocorrelation

Spline-DCS does a good job in capturing the dynamics of the data. Figure 9 shows that the estimation residuals, $\widehat{\varepsilon}_{t,\tau}$, and the score, $\widehat{u}_{t,\tau}$, exhibit little to no signs of serial correlation. For the FX data, the results are generally similar for other sampling windows. The large sample size makes the Ljung-Box statistics sensitive to small departures from zero autocorrelation. This can be seen in the

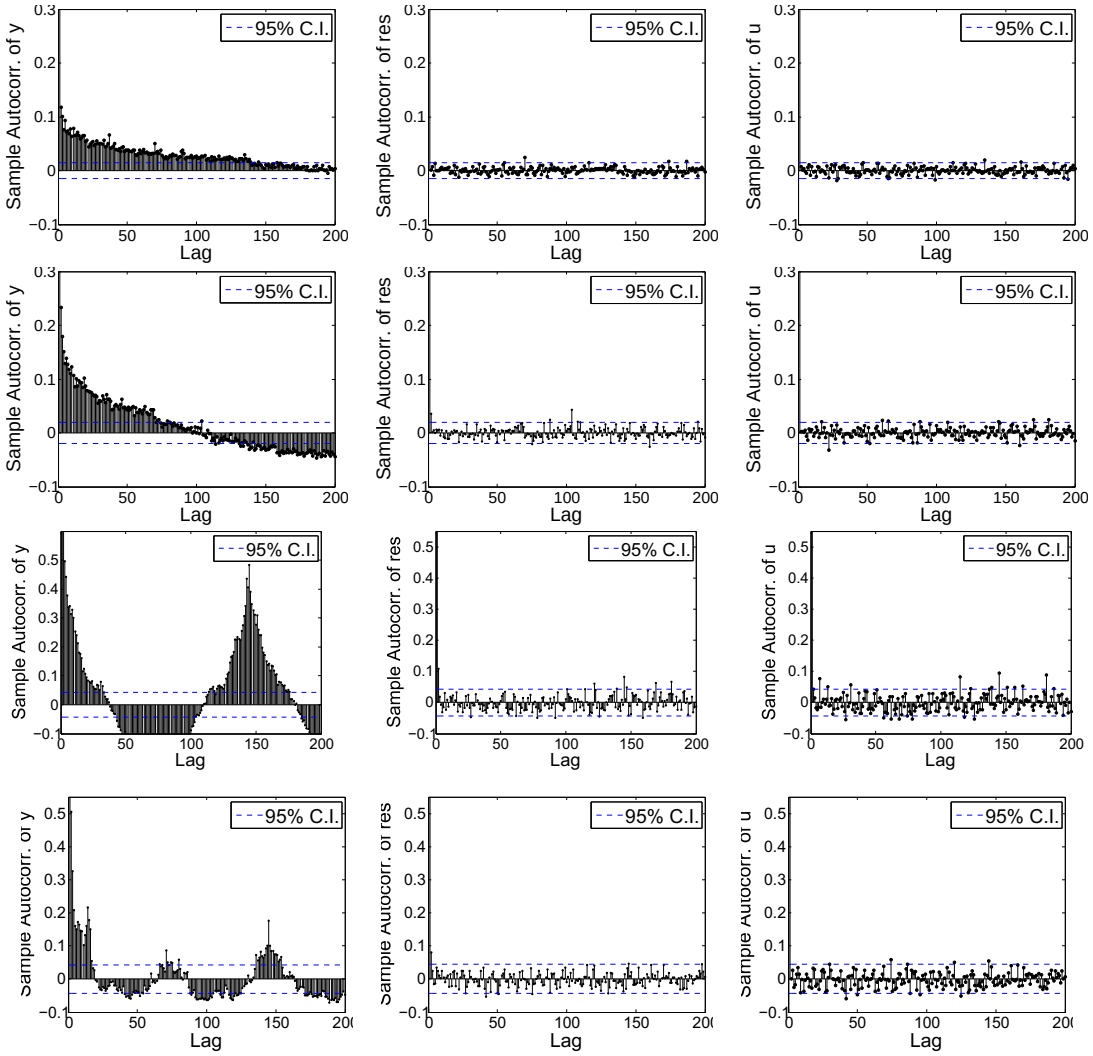


Figure 9. The sample autocorrelation of trade volume (left), $\widehat{\varepsilon}_{t,\tau}$ (middle), and $\widehat{u}_{t,\tau}$ (right). IBM30s (top row), IBM1m (second row), EURUSD (Window 1, third row), and USDJPY (Window 6, bottom row). Spline-DCS with the static daily spline. The 95% confidence interval is computed at ± 2 standard errors.

statistics reported in Table 6, which pick up statistically significant autocorrelation. [11]

4.3 Estimated coefficients

Table 7 shows the estimated coefficients. The estimates are obtained without imposing any restrictions on the parameters. The standard errors are computed analytically using the methodology outlined in Section 3.6.1. (For the FX data,

[11] Note that the sample autocorrelation of $\widehat{u}_{t,\tau}$ may exhibit stronger serial correlation than that of $\widehat{\varepsilon}_{t,\tau}$, because the score weighs down (and thus it is robust to) the effect of extreme observations.

	$\widehat{\varepsilon}_{t,\tau}$						$\widehat{\varepsilon}^2_{t,\tau}$					
Window	$\widehat{\rho}_1$	$\widehat{\rho}_{day}$	Q_1	Q_{day}	p-val. 1	p-val. day	$\widehat{\rho}_1$	$\widehat{\rho}_{day}$	Q_1	Q_{day}	p-val. 1	p-val. day
1	0.120	-0.033	31.048	169.116	0.000	0.000	0.080	-0.010	13.961	53.550	0.000	1.000
2	0.147	-0.018	46.998	188.952	0.000	0.000	0.125	-0.006	33.870	43.940	0.000	1.000
3	0.187	-0.002	74.982	277.065	0.000	0.000	0.094	-0.006	18.854	68.910	0.000	0.999
4	0.079	-0.009	13.309	142.910	0.000	0.016	0.069	0.000	10.271	92.462	0.001	0.872
5	0.069	-0.023	10.246	150.444	0.001	0.005	0.037	-0.009	2.862	84.197	0.091	0.963
6	0.080	0.009	13.494	157.383	0.000	0.002	0.177	0.008	65.872	191.092	0.000	0.000
7	0.050	-0.001	5.387	130.444	0.020	0.089	0.007	0.007	0.095	16.065	0.758	1.000
8	0.099	-0.030	20.931	163.575	0.000	0.000	0.156	-0.014	51.776	192.510	0.000	0.000
9	0.223	-0.019	104.555	268.999	0.000	0.000	0.388	0.003	317.345	361.753	0.000	0.000
10	0.092	0.030	18.075	174.686	0.000	0.000	0.050	0.016	5.386	201.496	0.020	0.000
11	0.131	-0.017	36.784	182.354	0.000	0.000	0.047	-0.002	4.613	16.008	0.032	1.000
12	0.084	-0.006	14.891	161.028	0.000	0.001	0.038	-0.002	3.075	98.504	0.079	0.755
13	0.088	-0.024	16.557	171.534	0.000	0.000	0.017	-0.006	0.613	206.254	0.434	0.000
14	0.128	-0.033	34.979	201.509	0.000	0.000	0.045	-0.004	4.436	29.458	0.035	1.000
15	0.188	-0.022	75.888	195.670	0.000	0.000	0.090	-0.002	17.539	23.249	0.000	1.000
16	0.082	-0.036	12.476	129.013	0.000	0.082	0.060	-0.011	6.660	25.878	0.010	1.000

	$\widehat{u}_{t,\tau}$						$\widehat{u}^2_{t,\tau}$					
Window	$\widehat{\rho}_1$	$\widehat{\rho}_{day}$	Q_1	Q_{day}	p-val. 1	p-val. day	$\widehat{\rho}_1$	$\widehat{\rho}_{day}$	Q_1	Q_{day}	p-val. 1	p-val. day
1	0.029	-0.031	1.855	155.616	0.173	0.002	0.081	-0.028	14.022	205.040	0.000	0.000
2	0.007	-0.003	0.101	298.334	0.750	0.000	0.066	-0.026	9.491	197.152	0.002	0.000
3	0.044	-0.002	4.139	233.940	0.042	0.000	0.127	0.012	34.561	175.616	0.000	0.000
4	0.004	-0.016	0.037	196.560	0.847	0.000	0.095	-0.018	19.508	173.661	0.000	0.000
5	0.024	-0.009	1.252	178.331	0.263	0.000	0.084	0.000	15.044	191.530	0.000	0.000
6	0.002	0.001	0.006	183.804	0.936	0.000	0.072	0.033	10.861	162.538	0.001	0.001
7	0.008	-0.035	0.144	137.420	0.705	0.039	0.077	0.002	12.462	149.393	0.000	0.007
8	0.001	-0.022	0.003	133.777	0.957	0.047	0.042	-0.009	3.734	264.000	0.053	0.000
9	0.017	-0.035	0.625	151.152	0.429	0.004	0.101	-0.014	21.569	180.611	0.000	0.000
10	0.044	0.024	4.172	150.216	0.041	0.005	0.063	0.016	8.341	150.025	0.004	0.006
11	0.012	-0.036	0.326	166.686	0.568	0.000	0.077	-0.015	12.615	203.411	0.000	0.000
12	0.036	0.000	2.780	188.838	0.095	0.000	0.075	-0.004	11.865	188.038	0.001	0.000
13	0.018	-0.036	0.732	141.613	0.392	0.017	0.090	0.006	17.540	280.275	0.000	0.000
14	0.008	-0.008	0.139	169.272	0.709	0.000	0.121	0.037	31.273	262.037	0.000	0.000
15	0.008	-0.022	0.127	170.162	0.721	0.000	0.121	0.003	31.749	214.484	0.000	0.000
16	0.025	-0.033	1.160	265.047	0.281	0.000	0.071	-0.010	9.441	215.045	0.002	0.000

Table 6. Residual analysis for Spline-DCS with the static daily spline fitted to the trade volume of EURUSD. Q_l is the Ljung-Box statistic to test the null of no autocorrelation up to the l-th lag.

the additional results for other sampling windows are given in Appendix C.) In Table 7, we have $\widehat{\kappa}_\eta^{(2)} > \widehat{\kappa}_\eta^{(1)} > \widehat{\kappa}_\mu > 0$, which means that $\eta_{t,\tau}^{(2)}$ is more sensitive to changes in $u_{t,\tau-1}$ than $\eta_{t,\tau}^{(1)}$, which is, in turn, more sensitive than $\mu_{t,\tau}$. We also have $0 < \widehat{\phi}_1^{(2)} < 1$ so $\eta_{t,\tau}^{(2)}$ is stationary. The stationarity of $\eta_{t,\tau}^{(1)}$ requires $\phi_1^{(1)} + \phi_2^{(1)} < 1$, $-\phi_1^{(1)} + \phi_2^{(1)} < 1$, and $\phi_2^{(1)} > -1$. (See Harvey (1993, p.19).) It is easy to check that these conditions are satisfied by $\widehat{\phi}_1^{(1)}$ and $\widehat{\phi}_2^{(1)}$ so that $\eta_{t,\tau}^{(1)}$ is also stationary. We have $|\widehat{\phi}_e| < 1$ so that $\widehat{e}_{t,\tau}$ reverts to zero after an event. The estimates of the probability mass at zero are consistent with the sample statistics in Tables 2 and 4 (e.g. $\widehat{p} = 0.0047$ for IBM30s). For the equity data, the estimated asymmetry term, $\widehat{\kappa}_{\eta,a}^{(2)}$, in $\eta_{t,\tau}^{(2)}$ is negative and statistically significant, reflecting the tendency of volume to decrease when price falls. Brownlees et al. (2011) found that the sign of their asymmetry term was positive so that volume on average increases when price falls. The direction of asymmetry appears to be in the opposite direction compared to us. $\widehat{\kappa}_{\eta,a}^{(1)}$ was found to be statistically insignificant for both IBM1m and IBM30s, suggesting that there is no asymmetry effect in the lower frequency component $\eta_{t,\tau}^{(1)}$.

Variable Window	IBM30s —	IBM1m —	USDJPY 2	EURUSD 3
ν	1.632 (0.016)	2.229 (0.033)	1.738 (0.018)	2.002 (0.019)
ξ	—	—	1.738 (0.069)	1.369 (0.048)
ζ	1.484 (0.045)	1.143 (0.044)	2.062 (0.085)	1.501 (0.059)
ω	9.172 (0.199)	9.774 (0.178)	5.521 (0.095)	4.232 (0.068)
κ_μ	0.006 (0.001)	0.007 (0.002)	0.010 (0.003)	0.006 (0.003)
$\phi_1^{(1)}$	0.554 (0.134)	0.391 (0.091)	0.570 (0.017)	0.582 (0.016)
$\phi_2^{(1)}$	0.414 (0.133)	0.555 (0.093)	0.373 (0.017)	0.363 (0.016)
$\kappa_\eta^{(1)}$	0.049 (0.007)	0.047 (0.008)	0.059 (0.007)	0.070 (0.007)
$\phi_1^{(2)}$	0.690 (0.041)	0.610 (0.057)	0.438 (0.083)	0.369 (0.114)
$\kappa_\eta^{(2)}$	0.092 (0.008)	0.067 (0.008)	0.094 (0.009)	0.080 (0.010)
$\kappa_{\eta,a}^{(2)}$	-0.004 (0.004)	-0.008 (0.003)	—	—
p	0.0047 (0.0005)	0.0006 (0.0003)	0.000 (0.000)	0.006 (0.002)
γ_0	1.197 (0.066)	1.119 (0.064)	(omitted)	(omitted)
γ_1	0.061 (0.041)	0.066 (0.041)	(omitted)	(omitted)
γ_2	-0.419 (0.036)	-0.392 (0.036)	(omitted)	(omitted)
γ_3	-0.216 (0.037)	-0.244 (0.037)	(omitted)	(omitted)
ϕ_e	—	—	0.807 (0.047)	0.504 (0.108)
$\kappa_{e,1}$	—	—	2.739 (1.024)	0.998 (0.021)
$\kappa_{e,2}$	—	—	2.457 (1.094)	0.645 (0.097)
$\kappa_{e,3}$	—	—	1.424 (0.021)	1.707 (1.208)

Table 7. The estimated parameter values for Spline-DCS with the static daily spline. The standard errors are computed analytically using the outer-product of the first derivative of the joint log-likelihood (see Section 3.6.1). The estimated values for the parameters of $s_{t,\tau}$ are omitted in this table for the FX data due to the relatively large number of nodes. Burr (GB2 with $\xi = 1$) is estimated for the IBM data due to the reasons described in Section 4.1 and Footnote 9.

4.4 Comparing with log-normal and generalized gamma

The log-normal distribution is widely used to model non-negative time series whenever the logarithm of observations roughly resembles normality. In such cases, the degree of efficiency and bias in estimated parameters depend on how far the distribution of the log-transformed variable is from normality. See, for example, Alizadeh et al. (2002) for a discussion and an empirical example. See Appendix A.3 for the formal definition of log-normal. In our case, the quality of the fit of log-normal was inferior to that of Burr or GB2.[12] This was presumably due to the departure of the log-transformed non-zero data (denoted by log(IBM1m) and so on) from normality, particularly around the tail regions. The quantile-to-quantile (QQ) plot in Figure 10 shows that the normal distribution appears to put too much weight on the lower-tail and too little weight on the

[12]For the equity data, the Bowman-Shenton (Jarque-Bera) test comfortably rejected normality of the logarithm of the estimation residuals when F^* is log-normal.

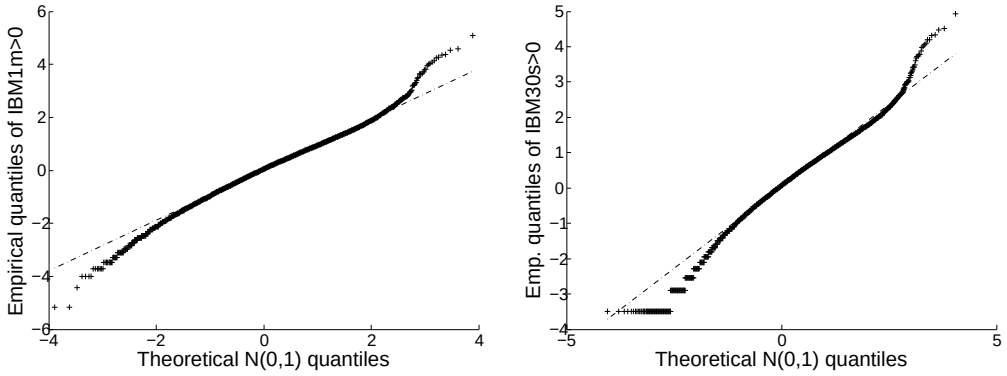


Figure 10. The QQ-plot of log(IBM1m) (left) and of log(IBM30s) (right). The logarithm series are re-centered around mean and standardized by one standard deviation.

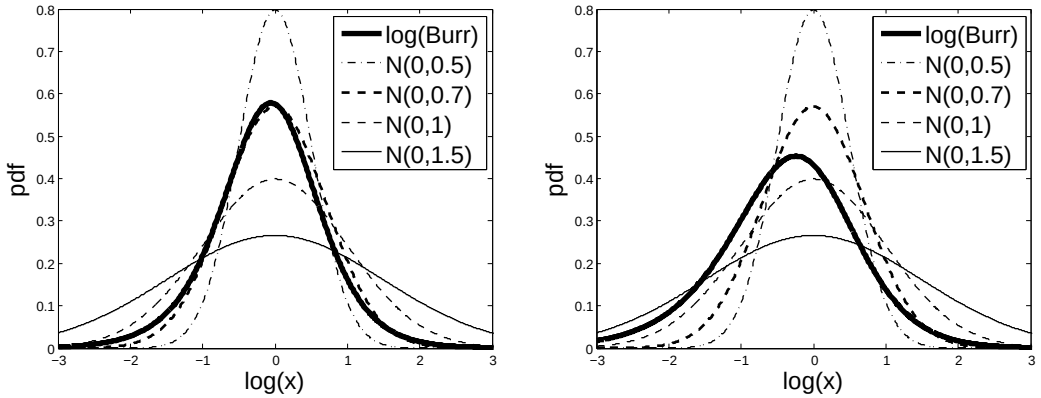


Figure 11. The p.d.f. of $\log(X)$: $X \sim \text{Burr}(\widehat{\nu}\ ,\widehat{\zeta})$ against $X \sim$ log-normal(σ) at different values of σ. Spline-DCS with the static daily spline was fitted to IBM1m (left) and IBM30s (right).

upper-tail.

Burr and GB2 may fit better than log-normal because the shape of log-normal is determined by only one parameter (σ), whereas Burr and GB2 have two to three shape parameters ((ν, ξ, ζ) for GB2 with $\xi = 1$ for Burr). Using the estimation results for the equity data, we compare the shape of Burr and log-normal in Figure 11. It plots the p.d.f. of $\log(X)$ when $X \sim \text{Burr}(\widehat{\nu}, \widehat{\zeta})$[13] against the case when $X \sim$ log-normal(σ) at different values of σ. The asymmetric shape of the p.d.f. of $\log(X)$ when $X \sim \text{Burr}(\widehat{\nu}, \widehat{\zeta})$ illustrates the

[13]If a random variable X follows the standard GB2 distribution with parameters, ν, ξ, and ζ, we say that $\log(X)$ follows the exponential-GB2 distribution with the p.d.f.

$$f(x;\nu,\xi,\zeta) = \frac{\nu \exp(x\nu\xi)(\exp(x\nu)+1)^{-\xi-\zeta}}{B(\xi,\zeta)}, \qquad x > 0, \text{ and } \nu,\xi,\zeta > 0$$

where $B(\cdot,\cdot)$ denotes the Beta function. Setting $\xi = 1$ gives the exponential-Burr distribution. Also see Harvey and Lange (2015) and McDonald and Xu (1995).

Series	Loglike		Likelihood	χ^2_1,
	$\zeta = 1$	$\zeta \neq 1$	ratio stat	p-value
IBM30s	-195,587	-195,490	194.6	0.000
IBM1m	-104,118	-104,112	12.0	0.001

Table 8. The likelihood ratio statistics to test the null $H_0 : \zeta = 1$ (log-logistic) against the alternative $H_1 : \zeta \neq 1$ (Burr).

relatively more flexible shape of Burr compared with the normal distribution. Although GB2 is related to log-normal, the estimated parameters $(\widehat{\nu}, \widehat{\xi}, \widehat{\zeta})$ suggest that the estimated GB2 is comfortably far from being log-normal since $\widehat{\nu}$ is comfortably away from zero, and $(\widehat{\nu}, \widehat{\xi}, \widehat{\zeta})$ are far from being large.

In the existing literature, the gamma distribution is also frequently used when dealing with non-negative variables (see Brownlees et al. (2011)). The gamma distribution is a special case of GG, and GG is a limiting case of GB2 for when ζ is large. Gamma and GG did not fit the empirical distribution of data well compared to GB2. This is consistent with the estimated ζ, which is far from being large. GB2 is not well-known unlike its special or limiting cases (such as Pareto, Burr, log-logistic, F, Weibull, and exponential). We can use the likelihood ratio test to check whether the estimated distributions are statistically significantly different from its special cases. As an illustration, Table 8 shows that the likelihood ratio test rejects the null hypothesis that $\zeta = 1$, which rules out the estimated Burr as being log-logistic. See Harvey and Lange (2015) for discussions and more examples of such tests.

We have $2 < \widehat{\nu}\widehat{\zeta} < 3$ for IBM1m and IBM30s, implying that only the first and second moments exist, and that the theoretical skewness does not exist, under the assumption that F^* is Burr.[14] (See Appendix A.1 for the existence of moments.) For the FX data, $3 < \widehat{\nu}\widehat{\zeta} < 4$, so that the moments up to the third exist. Since $\nu\zeta$ is also the (upper) tail-index of GB2, the estimates suggest that our series are heavy-tailed. Since a gamma distribution can never be heavy-tailed, this could be why the gamma distribution did not work well. GB2 may be preferred to the gamma distribution when dealing with heavy-tailed data. Also see Harvey (2013, p.12, p.189). We ruled out other candidate distributions after inspecting the empirical distribution of the estimation residuals.

[14]We did not check whether this result is because we estimated Burr instead of GB2 since we found it difficult to estimate GB2 for IBM1m and IBM30s as noted in Footnote 9.

Pair	EURUSD				Pair	USDJPY			
Freq	10 mins				Freq	10 mins			
	AIC		BIC			AIC		BIC	
Window	SDS	SWS	SDS	SWS	Window	SDS	SWS	SDS	SWS
1	**11.03**	11.07	**11.13**	11.28	1	**11.54**	11.56	**11.64**	11.77
2	**13.09**	13.13	**13.19**	13.34	2	**11.71**	11.74	**11.81**	11.95
3	**10.78**	10.82	**10.88**	11.03	3	**11.20**	11.22	**11.29**	11.43
4	**10.95**	11.00	**11.04**	11.21	4	**11.12**	11.15	**11.21**	11.35
5	**12.57**	12.61	**12.67**	12.82	5	**12.33**	12.37	**12.43**	12.58
6	**10.14**	10.21	**10.23**	10.42	6	**11.75**	11.78	**11.84**	11.99
7	12.79	**10.48**	12.88	**10.69**	7	**11.99**	12.03	**12.08**	12.23
8	**13.10**	13.14	**13.20**	13.35	8	**11.58**	11.62	**11.68**	11.83
9	**12.28**	12.32	**12.38**	12.53	9	**11.56**	11.60	**11.66**	11.81
10	**12.73**	12.77	**12.83**	12.98	10	**10.33**	10.40	**10.42**	10.60
11	**10.52**	10.56	**10.61**	10.77	11	**10.84**	10.86	**10.94**	11.07
12	**11.29**	11.33	**11.39**	11.53	12	**11.26**	11.29	**11.36**	11.50
13	**11.52**	11.60	**11.62**	11.81	13	**11.70**	11.74	**11.80**	11.94
14	13.77	**11.66**	13.86	**11.87**	14	**12.96**	13.00	**13.06**	13.21
15	**13.84**	13.88	**13.94**	14.09	15	**12.17**	12.19	**12.28**	12.40
16	**10.77**	10.80	**10.87**	11.01	16	**11.08**	11.10	**11.19**	11.31
Total min #	**14**	2	**14**	2	Total min #	**16**	0	**16**	0

	AIC		BIC	
	SDS	SWS	SDS	SWS
IBM30s	**20.0261**	20.0262	**20.032**	20.036
IBM1m	21.3048	**21.3046**	**21.316**	21.323

Table 9. Spline-DCS: in-sample model selection using AIC and BIC to determine the spline specification. SDS stands for the static daily spline and SWS for the static (restricted) weekly spline.

4.5 The estimated spline component

One way of choosing the specification of the periodic component ($s_{t,\tau}$) is to compare the values of the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). If the restricted weekly spline is preferred, it signals that the weekend effect may be present in our data. Table 9 shows that both AIC and BIC are typically in favor of the static daily spline, although they occasionally favor the static weekly spline in some cases. BIC penalizes free parameters more severely than AIC.[15] Thus the evidence for the weekend effect is generally weak and mixed in our data.

Figure 12 shows the estimated static daily spline, $\widehat{s}_{t,\tau}$, for IBM30s. We find that $\widehat{s}_{t,\tau}$ successfully captures the tendency of trade volume to be high in the

[15]Note that likelihood based tests computed under the null $H_0 : \widetilde{\gamma}_1 = \widetilde{\gamma}_2 = \widetilde{\gamma}_3$, against the alternative, which replaces $=$ by $\neq$, are invalid since the zero-sum constraint (or the weights $\boldsymbol{w}_*$ imposed) on $\boldsymbol{\gamma}$ is not the same for the daily spline and the weekly spline.

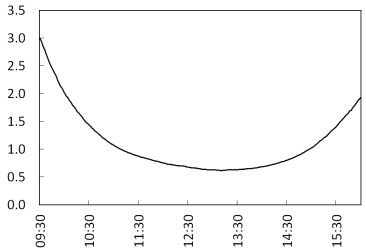
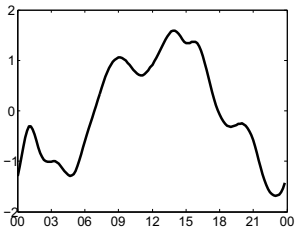
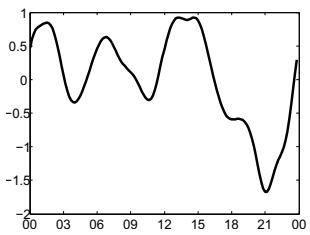


Figure 12. $\exp(\widehat{s}_{t,\tau})$ for IBM30s from market open to close. Spline-DCS with the static daily spline. Intra-day hours in the NY local time along the x-axes.

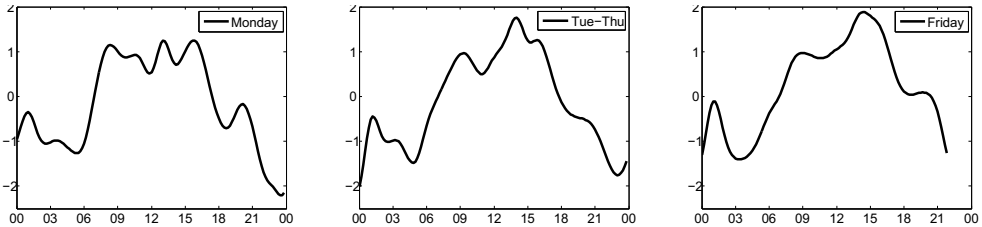
(a) Daily spline for EURUSD



(b) Daily spline for USDJPY



(c) Weekly spline for EURUSD: Mon (left), Tue-Thu (middle), Fri (right)



(d) Weekly spline for USDJPY: Mon (left), Tue-Thu (middle), Fri (right)

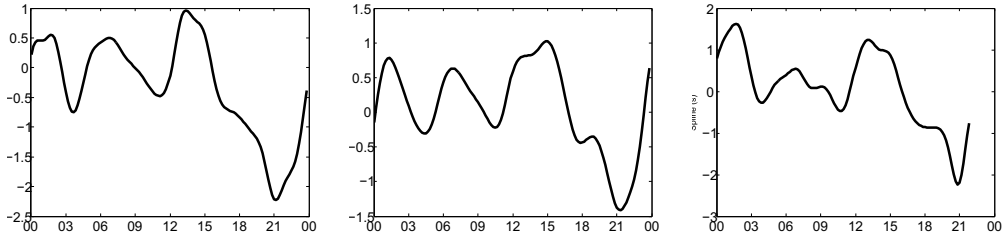


Figure 13. $\widehat{s}_{t,\tau}$ for the EURUSD volume (the in-sample window 2) and the USDJPY volume (the in-sample window 13). Intra-day hours in GMT along the x-axis.

morning, fall during the quiet lunch hours of around 1pm, and pick-up again in the afternoon. The spline takes a step-increase between the end and the beginning of any two consecutive trading days, reflecting the overnight effect. Figure 13 shows the estimated spline component for the FX data. The spline successfully captures the tendency of volume to increase when trading activities

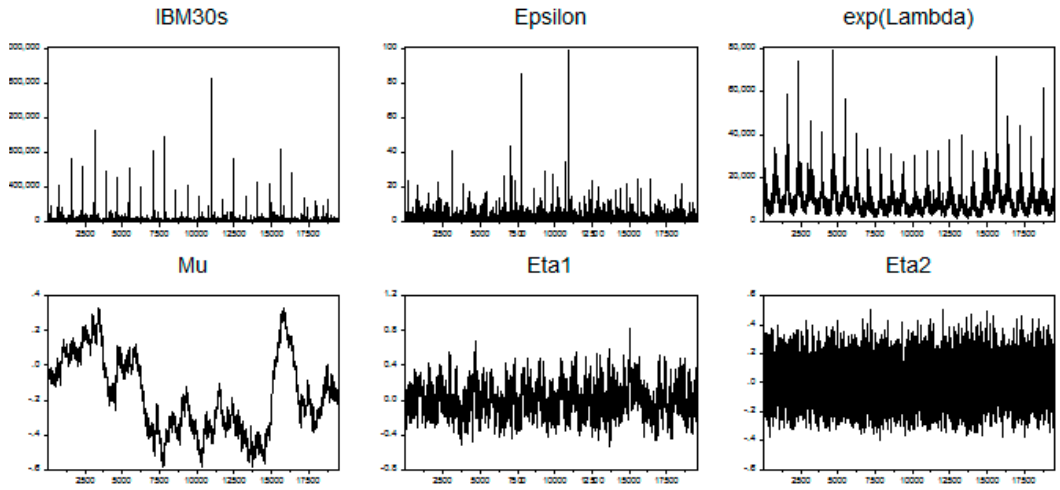


Figure 14. IBM30s (top left), $\widehat{\varepsilon}_{t,\tau}$ (top middle), $\exp(\widehat{\lambda}_{t,\tau})$ (top right), $\widehat{\mu}_{t,\tau}$ (bottom left), $\widehat{s}_{t,\tau}$ (bottom middle), and $\widehat{\eta}^{(1)}_{t,\tau}$ (bottom right). Time along the x-axes between 28 February - 31 March 2000. Spline-DCS with the static daily spline.

in major markets are high as we discussed in Section 2. For both the equity data and the FX data, the shapes we obtained are smooth. The difference in the overall pattern of periodicity between the equity and FX data is presumably because the dynamics of equity trade volume is dominated by the overnight effect, whereas the dynamics of FX trade volume is dominated by the timing of peaks in trade intensity around the world.

4.6 Estimated components

Figure 14 shows $\widehat{\varepsilon}_{t,\tau}$ and the estimated components of $\widehat{\lambda}_{t,\tau}$ for IBM30s. All series are displayed over the entire sampling period between 28 February and 31 March 2000. While the IBM30s volume series clearly exhibits periodic patterns, $\widehat{\varepsilon}_{t,\tau}$ appears free of periodicity. The diurnal U-shaped patterns are captured by $\exp(\widehat{\lambda}_{t,\tau})$ via the spline, $\widehat{s}_{t,\tau}$. $\widehat{\mu}_{t,\tau}$ and $\widehat{\eta}_{t,\tau}$ appear to satisfy their dynamic assumptions as $\widehat{\mu}_{t,\tau}$ resembles a random-walk, and $\widehat{\eta}^{(1)}_{t,\tau}$ and $\widehat{\eta}^{(2)}_{t,\tau}$ resemble stationary AR(2) and AR(1) processes, respectively.

5 Out-of-sample performance

5.1 Model stability: one-step ahead forecasts

We use the predictive c.d.f. to assess the stability of the estimated distribution and parameters, as well as the ability of our model to produce good one-step ahead forecasts over a given out-of-sample prediction period without

re-estimating the in-sample parameter values. Henceforth, we use the following notations

$$\begin{aligned}\Psi_H &= \{(t,\tau) \in \{T+1,\ldots,T+H\} \times \{1,\ldots,I\}\} \\ \Psi_{H,>0} &= \{(t,\tau) \in \{T+1,\ldots,T+H\} \times \{1,\ldots,I\} : y_{t,\tau} > 0\}.\end{aligned}$$

The procedure is as follows. Taking the in-sample parameter estimates obtained using samples up to date T, we recursively update the time-varying scale parameter at each new observation point $y_{T+h,\tau}$ for $\tau = 1,\ldots,I$ and $h = 1,\ldots,H$ to obtain one-step ahead forecasts, $\widetilde{\lambda}_{t,\tau}$, for the out-of-sample period, $(t,\tau) \in \Psi_{H,>0}$. We then compute the predictive c.d.f at each positive observation as

$$F^*(\widetilde{\varepsilon}_{t,\tau}; \widehat{\boldsymbol{\theta}}^*), \quad \widetilde{\varepsilon}_{t,\tau} = y_{t,\tau}/\exp(\widetilde{\lambda}_{t,\tau}) \tag{6}$$

for all $(t,\tau) \in \Psi_{H,>0}$. The predictive c.d.f. in (6) simply gives the PIT values of forecast standardized observations.

Figure 15 shows the empirical c.d.f. of the PIT values for IBM30s over different forecast horizons up to $H = 20$ days ahead. (The results for IBM1m are very similar to Figure 15.) As we have 390 observations per day for IBM1m and 780 observations per day for IBM30s, $H = 20$ corresponds to 7,800 steps ahead for IBM1m and 156,000 steps ahead for IBM30s. Likewise, Figure 16 shows the results for the EURUSD volume in Window 1. The length of the out-of-sample windows is two weeks as in Table 1. As before, the distribution of the PIT values is roughly $U[0,1]$ for this extended out-of-sample period, although there is non-negligible deterioration in the quality of fit. The results are similar for USDJPY and other sampling windows. Thus, Burr and GB2 appears to capture the empirical distribution of out-of-sample observations remarkably well for an extended out-of-sample period. Figure 17 shows the sample autocorrelation of the one-step ahead forecast $\widetilde{\varepsilon}_{t,\tau}$ and $\widetilde{u}_{t,\tau}$ for the EURUSD volume in the out-of-sample window 1. The results are similar for the equity data, as well as other out-of-sample windows and USDJPY. The one-step ahead forecasts by Spline-DCS appear to capture the volume dynamics over this extended out-of-sample period. In summary, our estimated distributions and parameter values appear to be fairly stable and able to provide good one-step-ahead forecasts of the conditional distribution of our data.

5.2 Multi-step ahead forecasts

We now examine multi-step forecasts, which are of greater interest than one-step forecasts as they give us an ultimate assessment of our model's predictive ability.

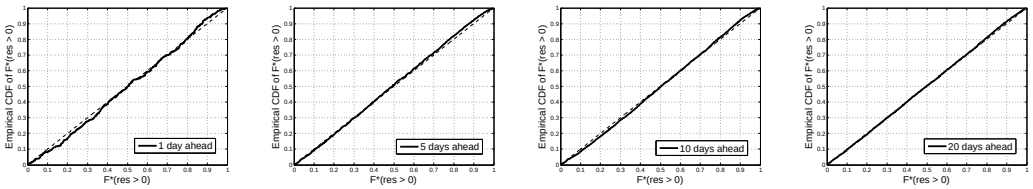


Figure 15. The empirical c.d.f. of the PIT values (predictive c.d.f.) of one-step ahead $\widetilde{\varepsilon}_{t,\tau}$. Computed using the theoretical c.d.f. of Burr$(\widehat{\nu}, \widehat{\zeta})$. Spline-DCS with the static weekly spline is fitted to IBM30s. Forecast horizons: 1 day ahead (left), 5 days ahead (left centre) 10 days ahead (right centre), and 20 days ahead (right). IBM30s for forecast horizons between 3 - 23 April 2000.

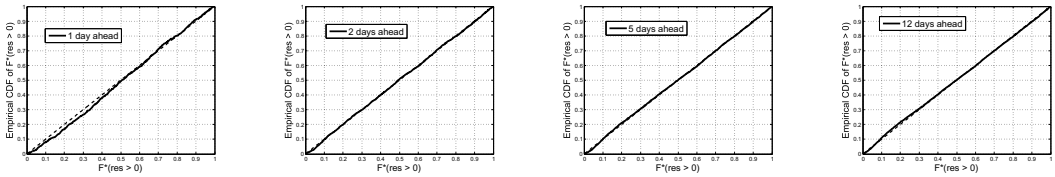


Figure 16. The empirical c.d.f. of the PIT values (predictive c.d.f.) of one-step ahead $\widetilde{\varepsilon}_{t,\tau}$. Computed using the theoretical c.d.f. of GB2$(\widehat{\nu}, \widehat{\xi}, \widehat{\zeta})$. Spline-DCS with the static weekly spline is fitted to EURUSD trade volume in Window 1. Forecast horizons: 1-day ahead (left), 2 days ahead (left centre), 5 days ahead (right centre), and 12 days ahead (right).

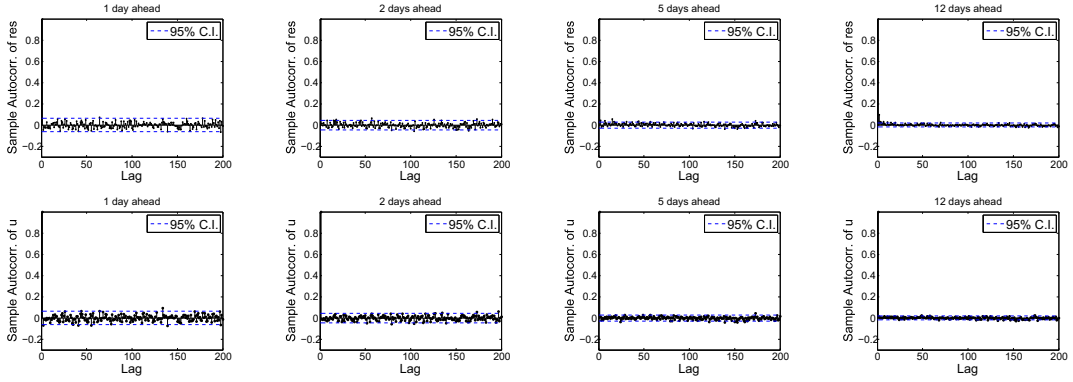


Figure 17. The sample autocorrelation of the out-of-sample $\widetilde{\varepsilon}_{t,\tau}$ (top) and $\widetilde{u}_{t,\tau}$ (bottom) when Spline-DCS with the static weekly spline is fitted to the trade volume of EURUSD (window 1). From the left column, 1-day ahead, 2 days ahead, 5 days ahead, and 12 days ahead.

We produce multi-step ahead density forecasts over a long forecast horizon using the estimation results obtained in Section 4.

We make optimal forecasts of $\exp(\lambda_{t,\tau})$ for $(t,\tau) \in \Psi_H$ conditional on $\mathcal{F}_{T,I}$. This optimality is in the sense of minimizing mean square error (MSE) so that the prediction is $\mathbb{E}\left[\exp(\lambda_{t,\tau})|\mathcal{F}_{T,I}\right] \equiv \overline{\exp(\lambda_{t,\tau})}$ for $(t,\tau) \in \Psi_H$. We use the property of DCS that the analytic expression of the multi-step optimal predictors

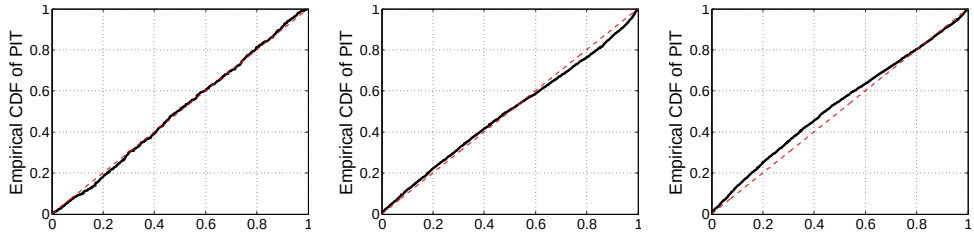


Figure 18. The empirical c.d.f of the PIT of multi-step forecasts. IBM30s for forecast horizons between 3 - 23 April 2000. Forecast horizons: 1 day ahead (left), 5 days ahead (middle), 8 days ahead (right). Spline-DCS with the static daily spline.

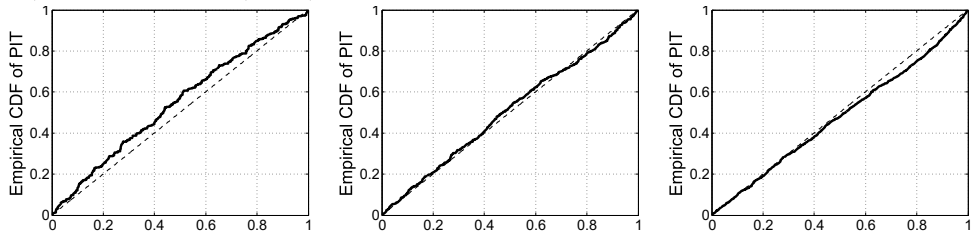


Figure 19. The empirical c.d.f of the PIT of multi-step forecasts. EURUSD (the out-of-sample window 1). Forecast horizons: 2 days ahead (left), 4 days ahead (middle), 8 days ahead (right). Spline-DCS with the static daily spline.

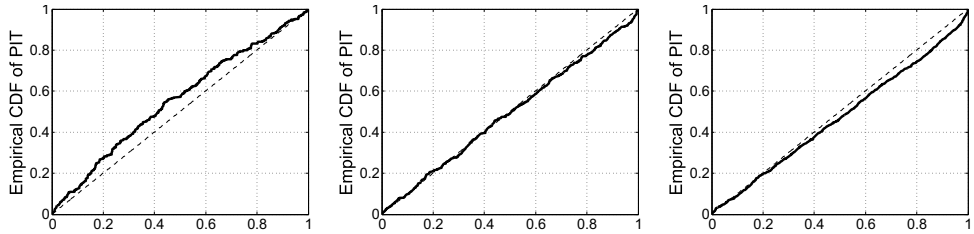


Figure 20. The empirical c.d.f of the PIT of multi-step forecasts. USDJPY (the out-of-sample window 3). Forecast horizons: 2 days ahead (left), 4 days ahead (middle), 8 days ahead (right). Spline-DCS with the static daily spline.

in the sense of minimizing MSE is available whenever the corresponding moments of F and the moment generating function (m.g.f) of $u_{t,\tau}$ exist. The methodology is as described by Harvey (2013). The difference between the notations, $\widetilde{\cdot}_{t,\tau}$ and $\bar{\cdot}_{t,\tau}$, is that the computation of the former does not involve evaluating the conditional moment condition. We then standardize the actual future observations $y_{t,\tau}$ by $\overline{\exp(\lambda_{t,\tau})}$ for all $(t,\tau) \in \Psi_H$. The standardized future observations $\bar{\varepsilon}_{t,\tau} \equiv y_{t,\tau}/\overline{\exp(\lambda_{t,\tau})}$ should be conditionally distributed as Burr$(\widehat{\nu}, \widehat{\zeta})$ for the equity data and GB2$(\widehat{\nu}, \widehat{\xi}, \widehat{\zeta})$ for the FX data at least approximately. Then $F^*(\bar{\varepsilon}_{t,\tau}; \widehat{\boldsymbol{\theta}}^*)$ for $(t,\tau) \in \Psi_{h,>0}$ gives the PIT values of the forecast standardized observations. They should be distributed approximately as $U[0,1]$ and be free of autocorrelation if the predictions are good. We make predictions for up to $H = 8$ future trading days, which corresponds to 3,120 steps ahead for

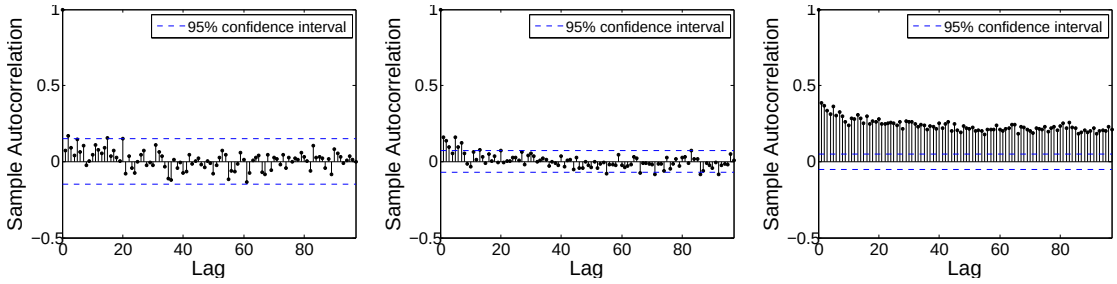


Figure 21. The sample autocorrelation of the PIT values of multi-step forecasts. IBM30s for forecast horizons between 3 - 23 April 2000. Forecast horizons: a quarter of a day ahead (left), one day ahead (middle), two days ahead (right). Spline-DCS with the static daily spline.

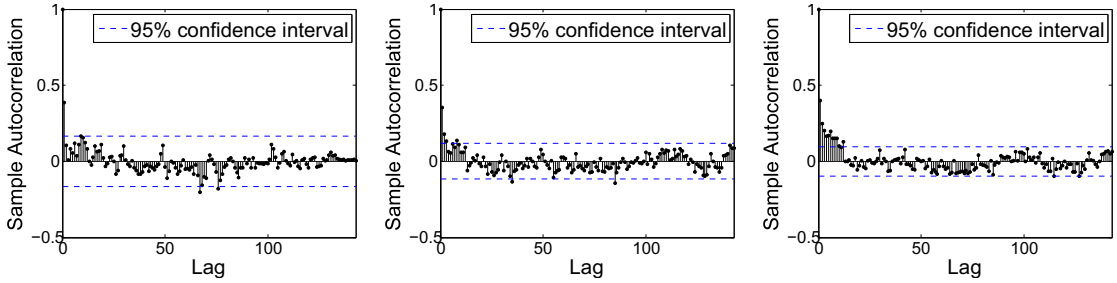


Figure 22. The sample autocorrelation of the PIT values of multi-step forecasts. EURUSD (the out-of-sample window 1). Forecast horizons: 1 day ahead (left), 2 days ahead (middle), 3 days ahead (right). Spline-DCS with the static daily spline.

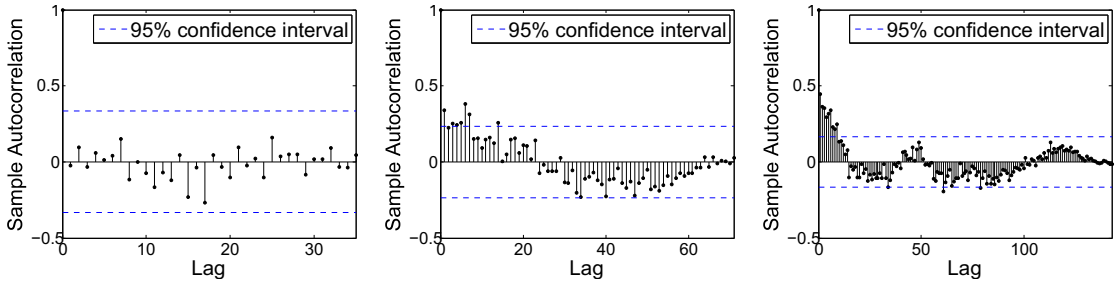


Figure 23. The sample autocorrelation of the PIT values of multi-step forecasts. USDJPY (the out-of-sample window 3). Forecast horizons: a quarter of a day ahead (left), one day ahead (middle), two days ahead (right). Spline-DCS with the static daily spline.

IBM1m, 6,240 steps ahead for IBM30s, and 1,152 steps ahead for the FX data.

For the equity data, the distribution of the PIT values is close to $U[0,1]$ for the first day of forecast horizon for both IBM1m and IBM30s. See Figure 18. We omit presenting the pictures for IBM1m again as they are very similar to IBM30s. This means that the density prediction produced by our model is good for the first 390 steps for IBM1m and 780 steps for IBM30s. A Box-Ljung test

indicates that the PIT values for the first quarter of the first forecast day, equivalent to around 200 steps ahead for IBM30s (100 steps for IBM1m), are not serially correlated. See Figure 21. By the end of the first forecast day, the PIT becomes autocorrelated, although the degree of autocorrelation remains very small. These results are again similar for IBM1m. If we consider this predictive performance in terms of the number of steps, these results appear to be good as they imply that the PIT values are close to being i.i.d. $U[0,1]$ for several hundred steps of the forecast horizon. Beyond the first forecast day, the quality of density forecasts deteriorate and the degree of autocorrelation increases with the length of forecast horizon.

For the FX data, we find that the distribution of the PIT values is fairly close to $U[0,1]$ (particularly around the 4th or 5th day of the forecast horizon for the cases shown in Figures 19 and 20). However, the closeness of the PIT values to being $U[0,1]$ fluctuates, and it is typically determined by the size of errors in forecasting the announcement effect. As before, by the end of the first forecast day (equivalent to 144 steps ahead), there is non-negligible autocorrelation in the PIT values. See Figures 22 and 23.

6 Out-of-sample model comparison

In this section, the FX data is used to compare the out-of-sample predictive performance of Spline-DCS and Fourier-MEM. First, we outline Fourier-MEM.

6.1 Fourier-MEM

Throughout this report, we adhere to the notations of Brownlees et al. (2011). The version of their model we test in this paper is

$$
\begin{aligned}
y_{t,\tau} &= \eta_t \phi_\tau \mu_{t,\tau} e^*_{t,\tau} \varepsilon_{t,\tau}, \quad \varepsilon_{t,\tau} \text{ i.i.d.} \sim (1, \sigma^2), \\
\eta_t &= \alpha_0^{(\eta)} + \beta_1^{(\eta)} \eta_{t-1} + \alpha_1^{(\eta)} y_{t-1}^{(\eta)} \\
\mu_{t,\tau} &= \alpha_0^{(\mu)} + \beta_1^{(\mu)} \mu_{t,\tau-1} + \alpha_1^{(\mu)} y_{t,\tau-1}^{(\mu)} + \alpha_2^{(\mu)} y_{t,\tau-2}^{(\mu)} \\
\phi_{\tau+1} &= \exp\left\{ \sum_{k=1}^{\lfloor I/2 \rfloor} \left[\delta_{1k} \cos\left(\frac{2\pi k \tau}{I}\right) + \delta_{2k} \sin\left(\frac{2\pi k \tau}{I}\right) \right] \right\}, \\
e^*_{t,\tau} &= \exp(e_{t,\tau}), \quad e_{t,\tau} = \beta_1^{(e)} e_{t,\tau-1} + \boldsymbol{\alpha}_1^{(e)\top} \boldsymbol{d}_{t,\tau}.
\end{aligned}
$$

η_t is the daily component, $\mu_{t,\tau}$ is the intra-day non-periodic component, and ϕ_τ is the intra-day periodic component, which approximate periodic patterns using the Fourier series. A new component we add to their model is $e^*_{t,\tau}$, which captures the effect of anticipated macroeconomic announcements. $\boldsymbol{d}_{t,\tau}$ is as

defined in (1) and $\dim(\boldsymbol{\alpha}_1^{(e)}) = m$. $y_t^{(\eta)}$ and $y_{t,\tau}^{(\mu)}$ are the *standardized* daily volume and intra-daily volume, respectively, which are defined as

$$y_t^{(\eta)} = \eta_t I^{-1} \sum_{\tau=1}^{I} \varepsilon_{t,\tau}, \quad \text{and} \quad y_{t,\tau}^{(\mu)} = \mu_{t,\tau}\varepsilon_{t,\tau}.$$

These make the model analogous to the GARCH model in the return volatility literature.[16] There are no asymmetry terms due to the reasons we discussed in Section 3.5.1.

The interpretations and identification conditions of the parameters are as laid out by the authors. To give a refresher, we mention some of them here. $\mu_{t,\tau}$ is stationary if $|\beta_1^{(\mu)} + \alpha_1^{(\mu)} + \alpha_2^{(\mu)}| < 1$. η_t is stationary if $|\beta_1^{(\eta)} + \alpha_1^{(\eta)}| < 1$. For the identification of the parameters, we assume $\mathbb{E}[\mu_{t,\tau}] = 1$, giving the constraint $\alpha_0^{(\mu)} = 1 - (\beta_1^{(\mu)} + \alpha_1^{(\mu)} + \alpha_2^{(\mu)})$. The initial conditions are $\eta_1 = x_1^{(\eta)} = \sum_{t=1}^{5}\sum_{\tau=1}^{I} y_{t,\tau}/5$ (i.e. the daily average volume over the first observation week excluding the weekend period) and $\mu_{1,1} = y_{1,1}^{(\mu)} = \mathbb{E}[\mu_{t,\tau}] = 1$. The continuity condition of data is $\mu_{t,0} = \mu_{t-1,I}$ and $y_{t,0}^{(\mu)} = y_{t-1,I}^{(\mu)}$. A sufficient condition to ensure that volume is positive is that the parameters of η_t and $\mu_{t,\tau}$ are positive. Taking the exponential of the Fourier series ensures that ϕ_τ is positive. $\delta_{2\lfloor I/2 \rfloor} = 0$ if I is even. ϕ_τ assumes that the pattern of periodicity is the same every day. In our case, there is no overnight dummies of the type used by Brownlees et al. (2011) in $\mu_{t,\tau}$ for the reasons described in Section 3.3 as we are modeling the FX data.

Brownlees et al. (2011) apply the model to forecasting the volume turnover of three Exchange Traded Funds that replicate the movements of U.S. stock indices, SPDR S&P 500, Diamonds, and PowerShares QQQ, between 2002 and 2006. The highest sampling frequency they consider is 15 minutes. The authors find that their intra-day volume data tend to cluster and that there is a strong serial-correlation in daily average volume. They also find that their data displays diurnal U-shaped patterns, with low volume around noon and high volume at the beginning and the end of trading day. This is similar to the periodic patterns we found in our equity data, but different to the patterns we found in our FX volume data. Our in-sample estimation results indicated that the model with the second lag-term in $\mu_{t,\tau}$ generally performed well in capturing the volume

[16] η_t and $\mu_{t,\tau}$ are GARCH-like filters because

$$\mathbb{E}[y_{t,\tau}^{(\mu)}|\mathcal{F}_{t,\tau-1}] = \mu_{t,\tau}, \qquad \mathrm{Var}[y_{t,\tau}^{(\mu)}|\mathcal{F}_{t,\tau-1}] = \mu_{t,\tau}^2\sigma^2,$$

and likewise for $y_t^{(\eta)}$ (with the normalization factor $1/I$ for the variance). The conditional moment of volume in the τ-th bin on the t-th observation day is $\mathbb{E}[y_{t,\tau}|\mathcal{F}_{t,\tau-1}] = \eta_t\phi_\tau\mu_{t,\tau}$.

dynamics, which is consistent with Brownlees et al. (2011).

6.1.1 GMM and MLE estimation

Brownlees et al. (2011) estimate Fourier-MEM by the generalized method of moment (GMM) in order to allow for a greater flexibility in the distribution of the error term. We adopt their GMM estimation strategy. The in-sample estimation method and results are given in Appendix D.

In addition to the GMM approach of Brownlees et al. (2011), we report the results of our attempt to estimate Fourier-MEM by ML. We try this alternative estimation method because we found the proposed GMM in Fourier-MEM computationally considerably slower to converge than ML. Also note that the GMM condition does not capture zero-valued observations explicitly. These features of Fourier-MEM are inconsequential if the sampling frequency is sufficiently low (so that the total sample size can be typically small and the number of zero-valued observations is small). Although we report the estimation results for ML in Fourier-MEM to highlight our preliminary investigation, the results must be interpreted with caution since the asymptotic properties of MLE in this model have not been studied and the results were typically sensitive to the choice of initial parameter values. The computational aspects of competing models are discussed in Section 7.

For ML in Fourier-MEM, we redefine the error distribution to be $\varepsilon_{t,\tau} \sim$ i.i.d $F(\varepsilon;\theta)$, where F is as defined in Section 3.4 so that it captures zero-valued observations explicitly. We also use GB2 for F^* in this case. Since Fourier-MEM assumes that the mean of the error distribution is one, we redefine $y_t^{(\eta)}$ and $y_t^{(\mu)}$ as

$$y_t^{(\eta)} = \eta_t (I \times M)^{-1} \sum_{\tau=1}^{I} \varepsilon_{t,\tau}, \quad \text{and} \quad y_{t,\tau}^{(\mu)} = \mu_{t,\tau}\varepsilon_{t,\tau}/M,$$

where $M \equiv \mathbb{E}[\varepsilon_{t,\tau}]$. $M = (1-p)B(\xi + 1/\nu, \zeta - 1/\nu)/B(\xi,\zeta)$ for GB2 defined in Appendix A.1, where $B(\cdot,\cdot)$ is the Beta function and (ν,ξ,ζ) are the distribution parameters as before. The stationarity and identification conditions outlined in Section 6.1 remain the same.

6.2 Comparison method

Since the proposed GMM in Fourier-MEM can be used to forecast the level of volume (hereafter, level forecasts), but not the density, the performance of the competing models are compared using their level forecasts.

For Spline-DCS, the one-step ahead level forecasts of volume (denoted by

$\widetilde{y}_{T+h,\tau}$) we evaluate are

$$\widetilde{y}_{T+h,\tau} \equiv \mathbb{E}[y_{T+h,\tau}|\mathcal{F}_{T+h,\tau-1}] = \exp(\widetilde{\lambda}_{T+h,\tau}) \int_0^\infty x f(x;\widehat{\theta}) dx \tag{7}$$

(i.e. one-step ahead (conditional first) moment forecast), and

$$\widetilde{y}_{T+h,\tau} \equiv \exp(\widetilde{\lambda}_{T+h,\tau}) Q_{0.5}(\widehat{\theta}), \quad \text{where} \quad \mathbb{P}\left(\varepsilon_{T+h,\tau} \leq Q_{0.5}(\widehat{\theta})|\mathcal{F}_{T+h,\tau-1}\right) = 0.5 \tag{8}$$

(i.e. (conditional) median forecast) for $h = 1, \ldots, H$ and $\tau = 1, \ldots, I$. Since the distribution of $\varepsilon_{t,\tau}$ is fully specified by GB2, the median and the first moment of $F(\varepsilon;\widehat{\theta})$ can be computed analytically.

For Fourier-MEM, we produce one-step ahead forecasts of $\eta_t \phi_\tau \mu_{t,\tau} e^*_{t,\tau}$ without updating the in-sample parameter estimates by following the same recursive updating methodology as Spline-DCS. For GMM in Fourier-MEM, the moment condition, $\mathbb{E}[\varepsilon_{t,\tau}] = 1$, means that the forecast path of $\eta_t \phi_\tau \mu_{t,\tau} e^*_{t,\tau}$ gives the model's (conditional first moment) level forecasts of volume. There is no other type of level forecasts we derived from it given the minimal GMM assumption for the error distribution. As for ML in Fourier-MEM, since the error distribution is completely defined in this case, we can compute two types of level forecasts using the mean and the median of $F(\varepsilon;\widehat{\theta})$ by following the same methodology as Spline-DCS.

For each in-sample window listed in Table 1, we estimate the models and produce forecasts for the corresponding out-of-sample window. We do this for all of the sixteen windows listed in Table 1. Then, we compute the forecast errors $(y_{t,\tau} - \widetilde{y}_{t,\tau})$ for all $(t,\tau) \in \Psi_H$ of each out-of-sample window. Our loss functions are the daily mean absolute errors (MAE):

$$L\left((y_{T+h,\tau}, \widetilde{y}_{T+h,\tau})_{\tau=1}^I\right) = (I)^{-1} \sum_{\tau=1}^I |y_{T+h,\tau} - \widetilde{y}_{T+h,\tau}|,$$

and the daily root mean squared errors (RMSE)

$$L\left((y_{T+h,\tau}, \widetilde{y}_{T+h,\tau})_{\tau=1}^I\right) = \sqrt{(I)^{-1} \sum_{\tau=1}^I (y_{T+h,\tau} - \widetilde{y}_{T+h,\tau})^2}.$$

They are computed for each $h = 1, \ldots, H$ of each out-of-sample window. The conditional first moment is theoretically optimal in the sense of minimizing RMSE. We consider not only RMSE, but also MAE, since few extreme observations can drive the conclusions of forecast evaluation studies that rely only on RMSE. The conditional median is theoretically the optimal predictor if the loss function is MAE.

Since the volume data, $y_{t,\tau}$ is highly volatile, the value of the daily loss

functions are typically dominated by large errors and, in some sense, can obscure the picture of comparative performance of competing models in minimizing loss function values. To get around this, we produce benchmark forecasts using some relatively ad-hoc ARMA-style model, and assess, by how much in percentage terms, each model is able to improve on the daily loss function values of the ARMA-style model. We call this model the *baseline* model. It is defined in Appendix E. We compute

$$100 \times \frac{L\left((y_{T+h,\tau}, \widetilde{y}_{T+h,\tau})_{\tau=1}^{I}\right)_{Model\ i} - L\left((y_{T+h,\tau}, \widetilde{y}_{T+h,\tau})_{\tau=1}^{I}\right)_{Baseline}}{L\left((y_{T+h,\tau}, \widetilde{y}_{T+h,\tau})_{\tau=1}^{I}\right)_{Baseline}} \tag{9}$$

for each out-of-sample day $h = 1, \ldots, H$. This is the percentage difference between the loss function value of a given model and that of the baseline on a given out-of-sample day h. The idea is to look for a model with forecast errors substantially below the errors of the baseline in percentage terms (i.e. small (9)). (9) is a good indicator of comparative performance for us since it can normalize errors that are due to extreme observation points in $(y_{t,\tau})_{(t,\tau)\in\Psi_{H,>0}}$.

6.3 Forecast comparison

Table 10 summarizes the forecast comparison results. In terms of minimizing the daily MAE, the median forecasts (i.e. the (8) quantities) by Spline-DCS yield the largest (16%) improvement on the daily MAE of the baseline on average each day with relatively small standard deviation (17% or 18%). The confidence intervals tabulated in Table 10 show that these sample averages are statistically significantly and comfortably different from zero at the 5% significance level. In evaluating the confidence intervals, we note that there are 158 out-of-sample days in total.[17] In terms of the number of days, the median forecasts (i.e. the (8) quantities) by Spline-DCS achieve the largest improvement on the daily MAE of the baseline on at least 113 days (or 72%) of the 158 out-of-sample days.

In terms of minimizing the daily RMSE, the confidence intervals suggest that Spline-DCS could not yield statistically significant improvement on the daily RMSE of the baseline. In contrast, the relatively wider confidence intervals in the positive region for Fourier-MEM suggest that it was typically outperformed by the baseline. The confidence intervals signal that the predictive performance of Spline-DCS does not fluctuate as much as that of Fourier-MEM.

The daily MAE appears to favor Spline-DCS more strongly than does the

[17] Two days (26 Nov. and 27 Nov. 2014) of the 15th out-of-sample window were excluded from the forecast evaluation due to some sampling errors in the data. More specifically, the intra-day times that records the moment observations $(y_{t,\tau})$ are collected were not consecutive for some of the samples on these dates.

			Daily MAE					
			(9), % diff. from baseline		95% confidence interval		# days it is the best	
			Average	S.D.	Lower bd	Upper bd	Count	Win rate
EURUSD	Fourier-MEM GMM	Moment f (7)	-3%	28%	-7%	2%	21	13%
10 mins	Fourier-MEM MLE	Moment f (7)	54%	164%	28%	79%	7	4%
		Median f (8)	31%	119%	12%	49%	7	4%
	Spline-DCS (SDS)	Moment f (7)	-8%	18%	-10%	-5%	7	4%
		Median f (8)	-16%	17%	-19%	-13%	116	73%
USDJPY	Fourier-MEM GMM	Moment f (7)	-3%	27%	-7%	1%	27	17%
10 mins	Fourier-MEM MLE	Moment f (7)	36%	239%	-1%	73%	7	4%
		Median f (8)	17%	174%	-10%	44%	10	6%
	Spline-DCS (SDS)	Moment f (7)	-7%	21%	-10%	-3%	1	1%
		Median f (8)	-16%	18%	-19%	-13%	113	72%
			Daily RMSE					
			(9), % diff. from baseline		95% confidence interval		# days it is the best	
			Average	S.D.	Lower bd	Upper bd	Count	Win rate
EURUSD	Fourier-MEM GMM	Moment f (7)	8%	51%	0%	16%	54	34%
10 mins	Fourier-MEM MLE	Moment f (7)	76%	437%	8%	144%	10	6%
		Median f (8)	51%	328%	0%	102%	8	5%
	Spline-DCS (SDS)	Moment f (7)	-2%	53%	-10%	6%	46	29%
		Median f (8)	-4%	41%	-10%	2%	40	25%
USDJPY	Fourier-MEM GMM	Moment f (7)	16%	79%	4%	29%	57	36%
10 mins	Fourier-MEM MLE	Moment f (7)	29%	142%	7%	51%	17	11%
		Median f (8)	18%	101%	2%	33%	15	9%
	Spline-DCS (SDS)	Moment f (7)	3%	56%	-5%	12%	43	27%
		Median f (8)	1%	40%	-5%	7%	26	16%

Table 10. Fourier-MEM and Spline-DCS: out-of-sample forecast comparison. SDS stands for the static daily spline. Note that the figures in the last column may add up to slightly below 100% due to the rounding of decimal places.

daily RMSE. This is because loss functions such as RMSE penalizes models for occasionally throwing up large errors more severely than MAE. When the data being estimated is heavy-tailed, a robust approach would prefer the use of MAE because it is not clear whether we should penalize forecast errors squarely when the timing and the size of tail observations are very difficult to predict.

Figure 24 shows the empirical distribution[18] of the (9) quantities for the selected models. A given model can improve on the baseline forecasts if the distribution rests in the negative territory. The relatively less disperse shape of the distribution for Spline-DCS compared to Fourier-MEM illustrates that the ability of Spline-DCS in improving on the daily MAE of the baseline can be more reliable and less variable for the forecasting horizons listed in Table 1.

[18]This is obtained using a smoothing kernel.

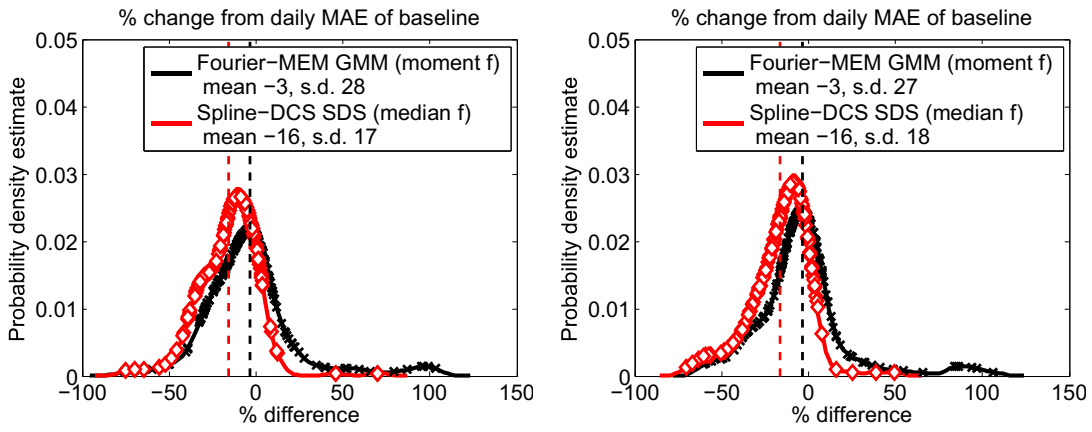


Figure 24. The empirical distribution of the (9) quantities for the selected model. For the EURUSD data (left) and the USDJPY data (right). The (9) quantity is computed for each out-of-sample day listed in Table 1. The vertical dashed lines correspond to the sample average of the distribution of the same color.

7 Discussions

7.1 Computing time: Fourier versus spline

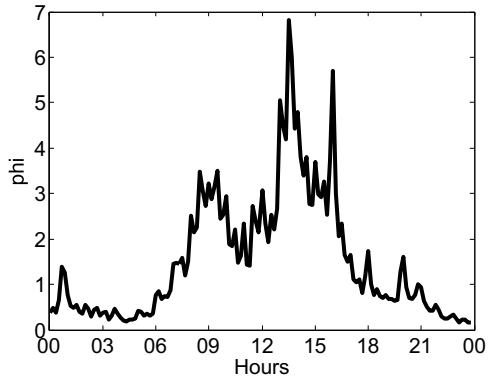
We found MLE in Spline-DCS faster and easier to compute than GMM in Fourier-MEM. With the FX data and the same convergence tolerance, the optimization procedure for Spline-DCS converged in about 5 minutes, whereas the Fourier-MEM usually took hours to converge (about 4 hours or longer). The main source of computational cost for the latter is the MM criterion (given in Appendix D), which includes the gradient vector of each component with respect to the parameters. The number of recursive dynamic equations to evaluate at each parameter value being tested is (the number of components)$\times$(1+the number of parameters), which can be large when the number of intra-day bins is high due to the Fourier component. Since the FX volume data is collected 24 hours a day, the total number of intra-day data is 144 at 10-minute frequency. Then, the number of Fourier coefficients become $\lfloor 144/2 \rfloor \times 2 - 1 = 143$ for the sine and cosine terms combined. This is a large number of parameters to estimate.

Figure 25 shows a snapshot of the estimated Fourier component, $\widehat{\phi}_\tau$. This compares with the estimated spline component ($\widehat{s}_{t,\tau}$) of Spline-DCS we showed earlier in Figure 13. The Fourier component successfully captures the bimodal or trimodal intra-day patterns in trade volume. However, unlike the estimated

spline of Spline-DCS, which is smooth and has no spikes, the estimated Fourier component has a lot of spikes. Some of the larger spikes seem as if they are picking up some of the announcement effects, especially at around 1.30pm in GMT for the USDJPY case, since regular data releases in the US are scheduled at this time of the day. These spikes are somewhat unintuitive as we expect the periodic component to capture smooth and regular transitions in volume and the non-periodic components to capture any (erratic) deviations from the regular patterns. The spikes in $\widehat{\phi}_\tau$ are presumably due to the well-known fact that the Fourier approximation does not work well at the discontinuous points of the function being approximated. The number of Fourier coefficients needed to achieve a good approximation is large if the function being approximated has many points of discontinuity. In our high-frequency context, the more the volume series fluctuate, the higher the number of Fourier terms we would need. Using substantially fewer Fourier terms than $\lfloor I/2 \rfloor$ not only makes the Fourier component smoother, but also reduces the computing time. However, the quality of approximation can also deteriorate substantially, especially around the points of extreme movements in volume. There seems to be a non-trivial trade-off between the quality of approximation and the computing time for GMM in Fourier-MEM.

As for Spline-DCS, the number of parameters used in the spline component is determined by the number of intra-day knots. As we mentioned in Section 3.3.2, the static daily spline has 21 parameters (or knots) for the FX data. Increasing the number of knots does not necessarily improve the quality of the fit of the model to data. The location and the number of knots play a key role in achieving good estimation outcomes. For both the equity and FX data, there were two rules of thumb that worked well. The first is to place one knot approximately every 1 hour to 1.5 hours along the intra-day time axis. The second is to place relatively more knots around the hours in which the intensity of trading activity changes fast. These hours correspond to the hours after NYSE opens or before it closes for the equity data, and the hours trading in major markets around the world peaks for the FX data. We placed fewer knots around the hours in which trading intensity does not change much. These hours typically correspond to lunch hours (in the NY local time) for the equity data and the evening hours (in GMT) when both New York and Asia are closed for FX. It is useful to sketch how a piecewise function of cubic polynomials would fit the empirical shape of intra-day patterns when determining the location and the number of knots.

(a) EURUSD volume, Window 2



(b) USDJPY volume, Window 10

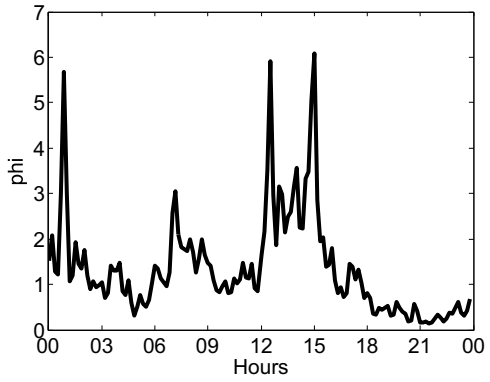


Figure 25. The estimated Fourier component ϕ_τ of Fourier-MEM. The x-axis is intra-day hours in GMT.

7.2 Computing time: GMM versus ML

The computing time of Fourier-MEM can be reduced substantially if it is estimated by ML, since this eliminates the need to compute a large number of recursive dynamic equations for the gradient vector in the MM criterion. However, as we noted in Section 6.1.1, the asymptotic properties of MLE in Fourier-MEM have not been studied, and the estimation results appeared to be sensitive to the choice of initial parameter values.

A main objection to an ML approach is that it requires the error distribution to be fully defined. GMM in Fourier-MEM may be preferred to ML in Spline-DCS if, for instance, there are reasons to believe that no parametrically chosen distribution can reasonably describe the empirical distribution of data. This does not seem to be the case at least in our application, since GB2 and Burr appeared to work well. If a parametrically chosen distribution works well, ML is useful since it allows us to produce density forecasts, from which we can compute other quantities such as the level forecasts we considered and other quantile or moment forecasts relating to risk measures (e.g. value at risk and expected shortfall).

8 Concluding remarks

This paper developed the Spline-DCS model for forecasting the dynamics of high-frequency trade volume with intra-day periodic patterns. We showed that it captures salient features of the high-frequency data such as the pattern of

periodicity, autocorrelation, and highly non-Gaussian empirical distribution of the data. The out-of-sample forecast results show that the in-sample estimation results are stable, and that our model can produce good density forecasts for a relatively long forecast horizon. Our estimation results are robust to the choice of sampling frequency and sampling period. We also found that the model can outperform the component MEM model (which we called Fourier-MEM) introduced by Brownlees et al. (2011) in forecasting the level of trade volume. The ease of computation and the intuitive shape of the estimated spline are two of the key advantages of Spline-DCS.

Burr and GB2 achieved a very good fit to the empirical distribution of the data. The estimated parameter values indicated that our data is heavy tailed. The flexible shape of these distributions presumably meant that they worked better than the gamma or log-normal distribution in our application due to the heavy-tailed feature and the departure of the log-transformed data from normality.

The pattern of periodicity was assumed to be the same every day in this paper. Ito (2013) challenges this standard assumption by introducing Spline-DCS with a dynamic spline, and show the empirical merit of such a generalization.

The object of our empirical analysis is trade volume, and, as such, this study also contributes to the literature dedicated to the analysis of market activity and intensity. It would be interesting to test our model to other variables such as asset price using heavy-tailed two-sided distributions. We studied the movements of volume in complete isolation from price, which is ultimately not satisfactory if one is interested in studying the interaction of price and quantity dynamics. Thus, the next natural step is to construct multivariate intra-day DCS that models price and volume simultaneously. One can also extend our framework to study correlation of scale with other variables, or construct a model for panel-data using a composite likelihood.

For the FX data, the forecast accuracy of Spline-DCS can be improved by adjusting the location of the knots of the spline to reflect the onset or the end of daylight saving. We estimated the announcement effect using relatively short in-sample period. Since the announcement effect can substantially dominate volume dynamics in FX, the accuracy of forecasts reported in this paper can be improved substantially by making the in-sample window longer to include a more comprehensive set of data points at anticipated events. These extensions are left for future research.

REFERENCES

Alizadeh, S., Brandt, M. W., and Diebold, F. X. (2002), "Range-Based Estimation of Stochastic Volatility Models," *Journal of Finance*, 57, 1047–1091.

Andersen, T. G. and Bollerslev, T. (1998), "Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Annoucements, and Longer Run Dependencies," *Journal of Finance*, 53, 219–265.

Blasques, F., Koopman, S. J., and Lucas, A. (2014), "Maximum Likelihood Estimation for Generalized Autoregressive Score Models," Tinbergen Institute Discussion Paper TI2014-029/III, Tinbergen Institute.

Bowsher, C. G. and Meeks, R. (2008), "The Dynamics of Economic Functions: Modelling and Forecasting the Yield Curve," *Journal of the American Statistical Association*, 103, 1419–1437.

Brownlees, C. T., Cipollini, F., and Gallo, G. M. (2011), "Intra-Daily Volume Modelling and Prediction for Algorithmic Trading," *Journal of Financial Economics*, 9, 489–518.

Creal, D. D., Koopman, S. J., and Lucas, A. (2011), "A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations," *Journal of Business and Economic Statistics*, 29, 552–563.

— (2013), "Generalized Autoregressive Score Models With Applications," *Journal of Applied Econometrics*, 28, 777–795.

Harvey, A. C. (1993), *Time Series Models*, Harvester: Wheatsheaf, 2nd ed.

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

Harvey, A. C. and Chakravarty, T. (2008), "Beta-t-(E)GARCH," Cambridge Working Papers in Economics CWPE0840, University of Cambridge.

Harvey, A. C. and Koopman, S. J. (1993), "Forecasting Hourly Electricity Demand Using Time-Varying Splines," *Journal of the American Statistical Association*, 88, 1228–1236.

Harvey, A. C., Koopman, S. J., and Riani, M. (1997), "The Modeling and Seasonal Adjustment of Weekly Observations," *Journal of Business & Economic Statistics*, 15, 354–68.

Harvey, A. C. and Lange, R.-J. (2015), "Volatility Modeling with a Generalized t-Distribution," Cambridge Working Papers in Economics CWPE1517, University of Cambridge.

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

Hautsch, N., Malec, P., and Schienle, M. (2014), "Capturing the Zero: A New Class of Zero-Augmented Distributions and Multiplicative Error Processes," *Journal of Financial Econometrics*, 12, 89–121.

Ito, R. (2013), "Modeling Dynamic Diurnal Patterns in High Frequency Financial Data," Cambridge Working Papers in Economics CWPE1315, University of Cambridge.

— (2016), "Asymptotic Theory for Beta-t-GARCH," Cambridge Working Papers in Economics, University of Cambridge, forthcoming.

Kleiber, C. and Kotz, S. (2003), *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley: New York.

Lo, A. W. and Wang, J. (2010), "Stock Market Trading Volume," in *Handbook of Financial Econometrics*, eds. Aït-Sahalia, Y. and Hansen, L., North-Holland: New York, vol. 2.

McCulloch, R. E. and Tsay, R. S. (2001), "Nonlinearity in High-Frequency Financial Data and Hierarchical Models," *Studies in Nonlinear Dynamics and Econometrics*, 5, 1–17.

McDonald, J. B. and Xu, Y. J. (1995), "A Generalization of the Beta Distribution with Applications," *Journal of Econometrics*, 66, 133–152.

Poirier, D. (1976), *The Econometrics of Structural Change: With Special Emphasis on Spline Functions*, North-Holland.

Rydberg, T. H. and Shephard, N. (2003), "Dynamics of Trade-By-Trade Price Movements: Decomposition and Models," *Journal of Financial Econometrics*, 1, 2–25.

Appendix A: List of distributions

A.1 The generalized beta distribution of the second kind

The (standard) generalized beta distribution of the second kind (GB2) has the p.d.f.:

$$f(x;\nu,\xi,\zeta) = \frac{\nu x^{\nu\xi-1}(x^\nu+1)^{-\xi-\zeta}}{B(\xi,\zeta)}, \qquad x>0, \text{ and } \nu,\xi,\zeta>0$$

where $B(\cdot,\cdot)$ denotes the Beta function. GB2 becomes the Burr distribution when $\xi=1$ and the log-logistic distribution when $\xi=\zeta=1$. The Burr distribution is also called the Pareto Type IV distribution (Pareto IV). Log-logistic is also called Pareto III. Burr becomes Pareto II when $\nu=1$. Burr

becomes Weibull (defined in Appendix A.2) when $\zeta \to \infty$. GB2 with $\nu = 1$ and $\xi = \zeta$ is a special case of the F distribution with the degrees of freedom $\nu_1 = \nu_2 = 2\xi$. GB2 is related to the generalized gamma (GG) distribution as its limiting case.

If a non-standardized random variable Y follows the GB2 distribution, its p.d.f. $f_Y : \mathbb{R}_{>0} \to \mathbb{R}$ with the scale parameter $\alpha > 0$ is $f_Y(y; \alpha, \nu, \xi, \zeta) = f(y/\alpha; \nu, \xi, \zeta)/\alpha$ for $y > 0$. To obtain GG defined in Appendix A.2 from GB2, we replace α by $\alpha\zeta^{1/\nu}$ and replace ζ by the tail index, $\nu\zeta$. Then GB2 becomes GG by setting $\xi = \gamma$ and letting $\nu\zeta \to \infty$. GG becomes log-normal when $\gamma \to \infty$, provided that other parameters satisfy additional conditions. See Kleiber and Kotz (2003) or Harvey and Lange (2015). For a set of i.i.d. observations $y_1, \ldots, y_T$ where each follows the non-standardized GB2 distribution, the log-likelihood function of a single observation y_t can be written using the exponential link function $\alpha = \exp(\lambda)$ with the link parameter $\lambda \in \mathbb{R}$ as:

$$\log f_Y(y_t) = \log(\nu) - \nu\xi\lambda + (\nu\xi - 1)\log(y_t) - \log B(\xi, \zeta) - (\xi + \zeta)\log[(y_t e^{-\lambda})^\nu + 1]. \tag{A.1}$$

The *score* u_t of the non-standardized GB2 computed at y_t is

$$u_t \equiv \frac{\partial \log f_Y(y_t)}{\partial \lambda} = \frac{\nu(\xi + \zeta)(y_t e^{-\lambda})^\nu}{(y_t e^{-\lambda})^\nu + 1} - \nu\xi = \nu(\xi + \zeta)b_t - \nu\xi \tag{A.2}$$

where we used the notation $b_t \equiv (y_t e^{-\lambda})^\nu / ((y_t e^{-\lambda})^\nu + 1)$. By the property of the GB2 distribution, we know that b_t follows the beta distribution with parameters ξ and ζ. The beta distribution characterized by the m.g.f. is

$$M_b(z; \xi, \zeta) \equiv \mathbb{E}[e^{bz}] = 1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \left(\frac{\xi + r}{\xi + \zeta + r} \right) \frac{z^k}{k!} \right).$$

It is easy to check that $\mathbb{E}[u_t] = 0$. $b_t(\xi, \zeta)$ is bounded between 0 and 1, which means that we have $-\nu\xi \leq u_t \leq \nu\zeta$.

A.2 The generalized gamma distribution

The (standard) generalized gamma (GG) distribution has the p.d.f.:

$$f(x;\gamma,\nu) = \frac{\nu}{\Gamma(\gamma)} x^{\nu\gamma-1} \exp\left(-x^{\nu}\right), \qquad 0 < x, \text{ and } \gamma,\nu > 0,$$

where $\Gamma(\cdot)$ is the gamma function. The GG distribution becomes the gamma distribution when $\nu = 1$, the Weibull distribution when $\gamma = 1$, and the exponential distribution when $\nu = \gamma = 1$.

If a non-standardized random variable Y follows the GG distribution, its p.d.f. $f_Y : \mathbb{R}_{>0} \to \mathbb{R}$ with the scale parameter $\alpha > 0$ is $f_Y(y;\alpha,\gamma,\nu) = f(y/\alpha;\gamma,\nu)/\alpha$ for $y > 0$. For a set of i.i.d. observations $y_1,\ldots,y_T$ where each follows the non-standardized GG distribution, the log-likelihood function of a single observation y_t can be written using the exponential link function $\alpha = \exp(\lambda)$ as:

$$\log f_Y(y_t) = \log(\nu) - \lambda + (\nu\gamma - 1)\log(y_t e^{-\lambda}) - (y_t e^{-\lambda})^{\nu} - \log\Gamma(\gamma).$$

The score u_t of the non-standardized GG computed at y_t is

$$u_t \equiv \frac{\partial \log f_Y(y_t)}{\partial\lambda} = \nu(y_t e^{-\lambda})^{\nu} - \nu\gamma = \nu g_t - \nu\gamma,$$

where we used the notation $g_t \equiv (y_t e^{-\lambda})^{\nu}$. By the property of the GG distribution, we know that g_t follows the (standard) gamma distribution with parameter γ, which is characterized by the m.g.f., $\mathbb{E}[e^{gz}] = (1-z)^{-\gamma}$, for $z < 1$. We also have $\mathbb{E}[u_t] = 0$ and $u_t > -\nu\gamma$ by the property of the gamma distribution.

A.3 The log-normal distribution

The (non-standardized) log-normal distribution has the p.d.f.:

$$f_Y(y;\alpha,\sigma) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{\log(y)-\log(\alpha)}{\sigma}\right)^2\right), \qquad y > 0, \text{ and } \alpha,\sigma > 0.$$

For a log-normally distributed random variable Y, the moments of all orders can be obtained easily using the m.g.f. of the normal distribution as $\mathbb{E}[Y^m] = \mathbb{E}[e^{m\log(Y)}]$ for all $m \in \mathbb{N}_{>0}$.

For a set of i.i.d. observations $y_1,\ldots,y_T$, where each follows the log-normal distribution, the log-likelihood function of a single observation y_t can be written using the exponential link function $\alpha = \exp(\lambda)$ as:

$$\log f_Y(y_t) = -\log(y_t) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\log(y_t)-\lambda}{\sigma}\right)^2.$$

The score u_t of the log-normal computed at y_t is

$$u_t \equiv \frac{\partial \log f_Y(y_t)}{\partial \lambda} = \frac{\log(y_t) - \lambda}{\sigma^2}.$$

Thus, u_t is a Gaussian random variable. We also have $\mathbb{E}[u_t] = 0$ as λ is the first moment of $\log(y_t)$.

Appendix B: The spline component

In this section, we formally explain the mathematical construction of the spline component $s_{t,\tau}$.

The spline is termed a *daily* spline if the periodicity is complete over one trading day. The static daily spline assumes that the shape of intra-day periodic patterns is the same for every trading day.

The daily spline is a continuous piecewise function of time and connected at $k+1$ knots for some $k \in \mathbb{N}_{>0}$ such that $k < I$. The coordinates of the knots along the time axis are denoted by $\tau_0 < \cdots < \tau_k$, where $\tau_0 = 1$, $\tau_k = I$, and $\tau_j \in \{2, \ldots, I-1\}$ for $j = 1, \ldots, k-1$. The set of the knots is also called *mesh*. The y-coordinates (height) of the knots are denoted by $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_k)^\top$. We denote the distance between the knots along the time-axis by $h_j = \tau_j - \tau_{j-1}$ for $j = 1, \ldots k$. We begin by defining the cubic spline function $g : [\tau_0, \tau_k] \to \mathbb{R}$, which is a piecewise function of the form

$$g(\tau) = \sum_{j=1}^{k} g_j(\tau) \mathbb{1}_{\{\tau \in [\tau_{j-1}, \tau_j]\}}, \qquad \forall\, \tau \in [\tau_0, \tau_k],$$

where each function $g_j : [\tau_{j-1}, \tau_j] \to \mathbb{R}$ is a polynomial of order up to three for all $j = 1, \ldots, k$. We can set g to be continuous at each knot (τ_j, γ_j); that is, $g_j(\tau_j) = \gamma_j$ and $g_j(\tau_{j-1}) = \gamma_{j-1}$ for all $j = 1, \ldots, k$. This means we have

$$g_j(\tau_{j-1}) = g_{j-1}(\tau_{j-1}) \qquad \text{and} \qquad g_j'(\tau_{j-1}) = g_{j-1}'(\tau_{j-1}) \tag{B.1}$$

for $j = 2, \ldots, k$. (B.1) is the *continuity condition* of g. The polynomial order of each g_j means that $g_j''(\cdot)$ is a linear function on $[\tau_{j-1}, \tau_j]$ for $j = 1, \ldots, k$. This implies that

$$g_j''(\tau) = a_{j-1} + \frac{\tau - \tau_{j-1}}{h_j}(a_j - a_{j-1}) = \frac{(\tau_j - \tau)}{h_j} a_{j-1} + \frac{(\tau - \tau_{j-1})}{h_j} a_j, \tag{B.2}$$

for $\tau \in [\tau_{j-1}, \tau_j]$ and $j = 1, \ldots, k$, where $a_0 = g_1''(\tau_0)$ and $a_j = g_j''(\tau_j)$ for $j = 1, \ldots, k$. We call (B.2) the *polynomial order condition* of g.

We integrate (B.2) with respect to τ to find the expressions for g_j' and g_j. That is, we evaluate $g_j'(\tau) = \int g_j''(\tau) d\tau$ and $g_j(\tau) = \int \int g_j''(\tau) d\tau$ for each $j = 1, \ldots, k$, where we recover the integration constant using (B.1). Then we

obtain

$$g_j'(\tau) = -\left[\frac{1}{2}\frac{(\tau_j-\tau)^2}{h_j} - \frac{h_j}{6}\right] a_{j-1} + \left[\frac{1}{2}\frac{(\tau-\tau_{j-1})^2}{h_j} - \frac{h_j}{6}\right] a_j, \quad \text{(B.3)}$$

$$g_j(\tau) = \boldsymbol{r}_j(\tau)\cdot\boldsymbol{\gamma} + \boldsymbol{s}_j(\tau)\cdot\boldsymbol{a} \quad \text{(B.4)}$$

for $\tau \in [\tau_{j-1}, \tau_j]$ and $j = 1, \ldots, k$, where $\boldsymbol{a} = (a_0, a_1, \ldots, a_k)^\top$, and $\boldsymbol{r}_j(\tau)$ and $\boldsymbol{s}_j(\tau)$ are the following k-dimensional vectors

$$\boldsymbol{r}_j(\tau) = \left(0, \ldots, 0, \frac{(\tau_j-\tau)}{h_j}, \frac{(\tau-\tau_{j-1})}{h_j}, 0, \ldots, 0\right)^\top,$$

$$\boldsymbol{s}_j(\tau) = \left(0, \ldots, 0, (\tau_j-\tau)\frac{(\tau_j-\tau)^2 - h_j^2}{6h_j}, (\tau-\tau_{j-1})\frac{(\tau-\tau_{j-1})^2 - h_j^2}{6h_j}, 0, \ldots, 0\right)^\top. \quad \text{(B.5)}$$

The non-zero elements of $\boldsymbol{r}_j(\tau)$ and $\boldsymbol{s}_j(\tau)$ are at the jth and $(j+1)$th entries.

B.1 Static daily spline with overnight effect

The conditions for g'_j in (B.1) and (B.3) give

$$\frac{h_j}{h_j+h_{j+1}}a_{j-1}+2a_j+\frac{h_{j+1}}{h_j+h_{j+1}}a_{j+1}=\frac{6\gamma_{j-1}}{h_j(h_j+h_{j+1})}-\frac{6\gamma_j}{h_jh_{j+1}}+\frac{6\gamma_{j+1}}{h_{j+1}(h_j+h_{j+1})}$$

for $j=1,\dots,k-1$. From these, we obtained a system of $k-1$ equations with $k+1$ unknowns $a_0,\dots,a_k$. Following Poirier (1976) we set $a_0=a_k=0$ (the *natural condition* for a spline). We can write this system of equations in a matrix form as $\mathbf{P}\boldsymbol{a}=\mathbf{Q}\boldsymbol{\gamma}$, where $\mathbf{P}$ and $\mathbf{Q}$ are the following square matrices of size $(k+1)$:

$$\mathbf{P}=\begin{bmatrix} 2 & 0 & 0 & 0 & \dots & 0 & 0 \\ \frac{h_1}{h_1+h_2} & 2 & \frac{h_2}{h_1+h_2} & 0 & \dots & 0 & 0 \\ 0 & \frac{h_2}{h_2+h_3} & 2 & \frac{h_3}{h_2+h_3} & \cdots & 0 & 0 \\ 0 & 0 & \frac{h_3}{h_3+h_4} & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 2 & \frac{h_k}{h_{k-1}+h_k} \\ 0 & 0 & 0 & 0 & \dots & 0 & 2 \end{bmatrix},$$

$$\mathbf{Q}=\begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ \frac{6}{h_1(h_1+h_2)} & -\frac{6}{h_1h_2} & \frac{6}{h_2(h_1+h_2)} & \dots & 0 & 0 \\ 0 & \frac{6}{h_2(h_2+h_3)} & -\frac{6}{h_2h_3} & \dots & 0 & 0 \\ 0 & 0 & \frac{6}{h_3(h_3+h_4)} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\frac{6}{h_{k-1}h_k} & \frac{6}{h_k(h_{k-1}+h_k)} \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

The first and the last rows of $\mathbf{P}$ and $\mathbf{Q}$ ensure that $a_0=a_k=0$. For a non-singular $\mathbf{P}$, we have $\boldsymbol{a}=\mathbf{P}^{-1}\mathbf{Q}\boldsymbol{\gamma}$. Then (B.4) can be written as $g_j(\tau)=\boldsymbol{w}_j(\tau)\cdot\boldsymbol{\gamma}$ for $\tau\in[\tau_{j-1},\tau_j]$, where $\boldsymbol{w}_j(\tau)^\top=\boldsymbol{r}_j(\tau)^\top+\boldsymbol{s}_j(\tau)^\top\mathbf{P}^{-1}\mathbf{Q}$. Finally, we obtain the following expression for the daily cubic spline

$$s_\tau=g(\tau)=\sum_{j=1}^{k}\mathbb{1}_{\{\tau\in[\tau_{j-1},\tau_j]\}}\boldsymbol{w}_j(\tau)\cdot\boldsymbol{\gamma},\qquad\forall\,\tau\in[\tau_0,\tau_k].\tag{B.6}$$

The elements of $\boldsymbol{\gamma}$ are the parameters of the model to be estimated. For the parameters to be identified, we impose the following zero-sum constraint on the elements of $\boldsymbol{\gamma}$:

$$\sum_{\tau\in[\tau_0,\tau_k]}s_\tau=\sum_{\tau\in[\tau_0,\tau_k]}\sum_{j=1}^{k}\mathbb{1}_{\{\tau\in[\tau_{j-1},\tau_j]\}}\boldsymbol{w}_j(\tau)\cdot\boldsymbol{\gamma}=\boldsymbol{w}_*\cdot\boldsymbol{\gamma}=0,$$

where

$$\boldsymbol{w}_* = (w_{*0}, w_{*1}, \ldots, w_{*k})^\top = \sum_{\tau\in[\tau_0,\tau_k]} \sum_{j=1}^{k} \mathbb{1}_{\{\tau\in[\tau_{j-1},\tau_j]\}} \boldsymbol{w}_j(\tau).$$

We can impose this condition by setting $\gamma_k = -\sum_{i=0}^{k-1} w_{*i}\gamma_i / w_{*k}$. Then (B.8) becomes

$$s_\tau = \sum_{j=1}^{k} \mathbb{1}_{\{\tau\in[\tau_{j-1},\tau_j]\}} \sum_{i=0}^{k-1} \left(w_{ji}(\tau) - \frac{w_{jk}(\tau) w_{*i}}{w_{*k}} \right) \gamma_i = \sum_{j=1}^{k} \mathbb{1}_{\{\tau\in[\tau_{j-1},\tau_j]\}} \boldsymbol{z}_j(\tau) \cdot \boldsymbol{\gamma} \tag{B.7}$$

for $\tau \in [\tau_0, \tau_k]$. $w_{ji}(\tau)$ denotes the ith element of $\boldsymbol{w}_j(\tau)$, and the ith element of $\boldsymbol{z}_j(\tau)$ is

$$z_{ji}(\tau) = \begin{cases} w_{ji}(\tau) - w_{jk}(\tau) w_{*i}/w_{*k} & i \neq k \\ 0 & i = k \end{cases}$$

for $\tau \in [\tau_{j-1}, \tau_j]$ and each $i = 0, \ldots, k$ and $j = 1, \ldots, k$. When estimating the model, it is convenient to compute $\boldsymbol{w}_*$ using the equation $\boldsymbol{w}_*^\top = \boldsymbol{r}_*^\top + \boldsymbol{s}_*^\top \mathbf{P}^{-1}\mathbf{Q}$, where $\boldsymbol{r}_*$ and $\boldsymbol{s}_*$ are k-dimensional vectors computed using the rules of arithmetic and polynomial series as

$$\begin{aligned} \boldsymbol{r}_* &= \left(\frac{\tau_1 - \tau_0 + 1}{2}, \frac{\tau_2 - \tau_0}{2}, \ldots, \frac{\tau_{k-1} - \tau_{k-3}}{2}, \frac{\tau_k - \tau_{k-1} + 1}{2} \right)^\top, \\ \boldsymbol{s}_* &= \left(\frac{h_1 - h_1^3}{24}, \frac{\tau_2 - \tau_0 - h_2^3 - h_1^3}{24}, \ldots, \frac{\tau_{k-1} - \tau_{k-3} - h_{k-1}^3 - h_{k-2}^3}{24}, \frac{h_k - h_k^3}{24} \right)^\top. \end{aligned}$$

Note that these formulae for computing $\boldsymbol{w}_*$, $\boldsymbol{r}_*$, and $\boldsymbol{s}_*$ are different from those of Harvey and Koopman (1993) due to the removal of the periodicity condition.

B.2 Static weekly spline with overnight effect

The static spline becomes a static *weekly spline* if we set the periodicity to be complete over one trading week instead of one day. For this spline, recall that we redefine $\tau_0, \tau_1, \ldots, \tau_k$ as follows. We let $\widetilde{\tau}_0 < \widetilde{\tau}_1 < \cdots < \widetilde{\tau}_{k'}$ denote the coordinates along the time-axis of the *intra-day mesh*, where $k' < I$, $\widetilde{\tau}_0 = 1$, $\widetilde{\tau}_{k'} = I$, and $\widetilde{\tau}_j \in \{2, \ldots, I-1\}$ for $j = 1, \ldots, k'-1$. Then the coordinates $\tau_0, \tau_1, \ldots, \tau_k$ along the time-axis of the total mesh for the whole week is defined as $\tau_{i(k'+1)+j} = \widetilde{\tau}_j + iI$ for $i = 0, \ldots, 4$ and $j = 0, \ldots, k'$. Then $(\tau_j)_{j=0}^k$ is still an increasing sequence. The total number of knots for one whole week is $k + 1 = 5(k' + 1)$. The height of the knots are $\gamma_0, \gamma_1, \ldots, \gamma_{k'}$ for Monday, $\gamma_{k'+1}, \gamma_{k'+2}, \ldots, \gamma_{2(k'+1)}$ for Tuesday, and so on.

As before, there is no periodicity condition between τ_k and τ_0, so that $\gamma_k \neq \gamma_0$ is allowed and we can capture the effect of weekend news on trading patterns.

Moreover, we capture the overnight effect of weeknights by relaxing the continuity and polynomial order restrictions, (B.1)-(B.2), between $\widetilde{\tau}_{k'}$ and $\widetilde{\tau}_0$ of any two successive weekdays. Thus, the procedure for computing $\boldsymbol{z}_j(\tau)$ is different from the daily spline described above. This redefines $\mathbf{P}$ and $\mathbf{Q}$ matrices as follows. For the $\mathbf{P}$ matrix, we replace the off-diagonal entries in the $i(k'+1)$th and $(i(k'+1)+1)$th rows by zeros for each $i = 1, \ldots (k+1)/(k'+1)$. For the $\mathbf{Q}$ matrix, we replace *all* entries in the $i(k'+1)$th and $(i(k'+1)+1)$th rows by zeros for each $i = 1, \ldots (k+1)/(k'+1)$. If we use five knots per day as we specified in the main text, $\mathbf{P}$ and $\mathbf{Q}$ become

$$\mathbf{P} = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \frac{h_1}{h_1+h_2} & 2 & \frac{h_2}{h_1+h_2} & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \frac{h_2}{h_2+h_3} & 2 & \frac{h_3}{h_2+h_3} & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \frac{h_3}{h_3+h_4} & 2 & \frac{h_4}{h_3+h_4} & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{h_6}{h_5+h_6} & 2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 2 & \frac{h_k}{h_{k-1}+h_k} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 2 \end{bmatrix},$$

$$\mathbf{Q} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \frac{6}{h_1(h_1+h_2)} & -\frac{6}{h_1h_2} & \frac{6}{h_2(h_1+h_2)} & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \frac{6}{h_2(h_2+h_3)} & -\frac{6}{h_2h_3} & \frac{6}{h_3(h_2+h_3)} & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \frac{6}{h_3(h_3+h_4)} & -\frac{6}{h_3h_4} & \frac{6}{h_4(h_3+h_4)} & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{6}{h_6(h_5+h_6)} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & -\frac{6}{h_{k-1}h_k} & \frac{6}{h_k(h_{k-1}+h_k)} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

B.3 Static daily spline with no overnight effect

Since the FX data is collected 24 hours a day, we assume the *periodicity condition* in this case; that is, g_1 and g_k satisfy $\gamma_0 = \gamma_k$, $g_1'(\tau_0) = g_k'(\tau_k)$, and $g_1''(\tau_0) = g_k''(\tau_k)$ so that $a_0 = a_k$. This condition is the same as Harvey and Koopman (1993) since their hourly electricity demand data is also collected 24 hours a day.

By the periodicity condition, we have $\gamma_0 = \gamma_k$ and $a_0 = a_k$ so that γ_0 and a_0 become redundant during estimation. Moreover, the conditions for g_j' in (B.1) and (B.3) give

$$\frac{h_j}{h_j + h_{j+1}} a_{j-1} + 2a_j + \frac{h_{j+1}}{h_j + h_{j+1}} a_{j+1} = \frac{6\gamma_{j-1}}{h_j(h_j + h_{j+1})} - \frac{6\gamma_j}{h_j h_{j+1}} + \frac{6\gamma_{j+1}}{h_{j+1}(h_j + h_{j+1})}$$

for $j = 2, \ldots, k-1$ and

$$\frac{h_1}{h_1+h_2}a_k + 2a_1 + \frac{h_2}{h_1+h_2}a_2 = \frac{6\gamma_k}{h_1(h_1+h_2)} - \frac{6\gamma_1}{h_1h_2} + \frac{6\gamma_2}{h_2(h_1+h_2)}, \qquad j = 1,$$
$$\frac{h_k}{h_k+h_1}a_{k-1} + 2a_k + \frac{h_1}{h_k+h_1}a_1 = \frac{6\gamma_{k-1}}{h_k(h_k+h_1)} - \frac{6\gamma_k}{h_kh_1} + \frac{6\gamma_1}{h_1(h_k+h_1)} \qquad j = k.$$

From these, we obtained k equations for k "unknowns" $a_1, \ldots, a_k$. Using notations $\boldsymbol{\gamma}^\dagger = (\gamma_1^\dagger, \ldots, \gamma_k^\dagger)^\top = (\gamma_1, \ldots, \gamma_k)^\top$ and $\boldsymbol{a}^\dagger = (a_1^\dagger, \ldots, a_k^\dagger)^\top = (a_1, \ldots, a_k)^\top$, we can write this system of equations in a matrix form as $\mathbf{P}\boldsymbol{a}^\dagger = \mathbf{Q}\boldsymbol{\gamma}^\dagger$, where $\mathbf{P}$ and $\mathbf{Q}$ are the following square matrices of size k:

$$\mathbf{P} = \begin{bmatrix} 2 & \frac{h_2}{h_1+h_2} & 0 & 0 & \ldots & 0 & \frac{h_1}{h_1+h_2} \\ \frac{h_2}{h_2+h_3} & 2 & \frac{h_3}{h_2+h_3} & 0 & \ldots & 0 & 0 \\ 0 & \frac{h_3}{h_3+h_4} & 2 & \frac{h_4}{h_3+h_4} & \cdots & 0 & 0 \\ 0 & 0 & \frac{h_4}{h_4+h_5} & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 2 & \frac{h_k}{h_{k-1}+h_k} \\ \frac{h_1}{h_1+h_k} & 0 & 0 & 0 & \ldots & \frac{h_k}{h_1+h_k} & 2 \end{bmatrix},$$

$$\mathbf{Q} = \begin{bmatrix} -\frac{6}{h_1h_2} & \frac{6}{h_2(h_1+h_2)} & 0 & \ldots & 0 & \frac{6}{h_1(h_1+h_2)} \\ \frac{6}{h_2(h_2+h_3)} & -\frac{6}{h_2h_3} & \frac{6}{h_3(h_2+h_3)} & \cdots & 0 & 0 \\ 0 & \frac{6}{h_3(h_3+h_4)} & -\frac{6}{h_3h_4} & \ldots & 0 & 0 \\ 0 & 0 & \frac{6}{h_4(h_4+h_5)} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -\frac{6}{h_{k-1}h_k} & \frac{6}{h_k(h_{k-1}+h_k)} \\ \frac{6}{h_1(h_1+h_k)} & 0 & 0 & \ldots & \frac{6}{h_k(h_1+h_k)} & -\frac{6}{h_1h_k} \end{bmatrix}.$$

For a non-singular $\mathbf{P}$, we have $\boldsymbol{a}^\dagger = \mathbf{P}^{-1}\mathbf{Q}\boldsymbol{\gamma}^\dagger$. Then (B.4) can be written as $g_j(\tau) = \boldsymbol{w}_j(\tau) \cdot \boldsymbol{\gamma}^\dagger$ for $\tau \in [\tau_{j-1}, \tau_j]$, where $\boldsymbol{w}_j(\tau)^\top = \boldsymbol{r}_j(\tau)^\top + \boldsymbol{s}_j(\tau)^\top\mathbf{P}^{-1}\mathbf{Q}$. $\boldsymbol{r}_j(\tau)$ and $\boldsymbol{s}_j(\tau)$ are now $k \times 1$ vectors by the periodicity condition with

$$\boldsymbol{r}_1(\tau) = \left(\frac{\tau-\tau_0}{h_1}, 0, \ldots, 0, \frac{\tau_1-\tau}{h_1}\right)^\top,$$
$$\boldsymbol{s}_1(\tau) = \left((\tau-\tau_0)\frac{(\tau-\tau_0)^2-h_1^2}{6h_1}, 0, \ldots, 0, (\tau_1-\tau)\frac{(\tau_1-\tau)^2-h_1^2}{6h_1}\right)^\top,$$

and $\boldsymbol{r}_j(\tau)$ and $\boldsymbol{s}_j(\tau)$ for $j = 2, \ldots, k$ are as defined in (B.5) but the non-zero elements are shifted to $(j-1)$-th and j-th entries. Finally, we obtain the

following expression for the daily cubic spline

$$s_\tau = g(\tau) = \sum_{j=1}^{k} \mathbb{1}_{\{\tau\in[\tau_{j-1},\tau_j]\}} \boldsymbol{w}_j(\tau)\cdot\boldsymbol{\gamma}^\dagger, \qquad \forall\,\tau\in[\tau_0,\tau_k]. \tag{B.8}$$

The elements of $\boldsymbol{\gamma}^\dagger$ are the parameters of the model to be estimated. For the parameters to be identified, we impose the following zero-sum constraint on the elements of $\boldsymbol{\gamma}^\dagger$

$$\sum_{\tau\in[\tau_0,\tau_k]} s_\tau = \sum_{\tau\in[\tau_0,\tau_k]} \sum_{j=1}^{k} \mathbb{1}_{\{\tau\in[\tau_{j-1},\tau_j]\}} \boldsymbol{w}_j(\tau)\cdot\boldsymbol{\gamma}^\dagger = \boldsymbol{w}_*\cdot\boldsymbol{\gamma}^\dagger = 0,$$

where

$$\boldsymbol{w}_* = (w_{*0}, w_{*1}, \ldots, w_{*k})^\top = \sum_{\tau\in[\tau_0,\tau_k]} \sum_{j=1}^{k} \mathbb{1}_{\{\tau\in[\tau_{j-1},\tau_j]\}} \boldsymbol{w}_j(\tau).$$

We can impose this condition by setting $\gamma_k = -\sum_{i=0}^{k-1} w_{*i}\gamma_i/w_{*k}$. Then (B.8) becomes

$$s_\tau = \sum_{j=1}^{k} \mathbb{1}_{\{\tau\in[\tau_{j-1},\tau_j]\}} \sum_{i=1}^{k-1} \left(w_{ji}(\tau) - \frac{w_{jk}(\tau)w_{*i}}{w_{*k}} \right) \gamma_i^\dagger = \sum_{j=1}^{k} \mathbb{1}_{\{\tau\in[\tau_{j-1},\tau_j]\}} \boldsymbol{z}_j(\tau)\cdot\boldsymbol{\gamma}^\dagger \tag{B.9}$$

for $\tau\in[\tau_0,\tau_k]$. $w_{ji}(\tau)$ denotes the ith element of $\boldsymbol{w}_j(\tau)$, and the ith element of $\boldsymbol{z}_j(\tau)$ is

$$z_{ji}(\tau) = \begin{cases} w_{ji}(\tau) - w_{jk}(\tau)w_{*i}/w_{*k} & i\neq k \\ 0 & i = k \end{cases}$$

for $\tau\in[\tau_{j-1},\tau_j]$ and $i,j=1,\ldots,k$. Thus, $\boldsymbol{z}_j : [\tau_{j-1},\tau_j]^k \to \mathbb{R}^k$ for $j=1,\ldots k$ is a k-dimensional vector of deterministic functions that conveys all information about the polynomial order, continuity, periodicity, and zero-sum conditions of the spline. (B.9) is the static daily spline we estimate in this chapter.

When estimating the model, it is convenient to compute $\boldsymbol{w}_*$ using the equation $\boldsymbol{w}_*^\top = \boldsymbol{r}_*^\top + \boldsymbol{s}_*^\top\mathbf{P}^{-1}\mathbf{Q}$, where $\boldsymbol{r}_*$ and $\boldsymbol{s}_*$ are k-dimensional vectors computed using the rules of arithmetic and polynomial series as

$$\boldsymbol{r}_* = \left(\frac{\tau_2-\tau_0}{2}, \ldots, \frac{\tau_k-\tau_{k-2}}{2}, \frac{\tau_1-\tau_0+\tau_k-\tau_{k-1}}{2} \right)^\top,$$

$$\boldsymbol{s}_* = \left(\frac{\tau_2-\tau_0-h_2^3-h_1^3}{24}, \ldots, \frac{\tau_k-\tau_{k-2}-h_k^3-h_{k-1}^3}{24}, \frac{h_1(1-h_1^2)+h_k(1-h_k^2)}{24} \right)^\top.$$

Note that these formulae for computing $\boldsymbol{w}_*$, $\boldsymbol{r}_*$, and $\boldsymbol{s}_*$ are different from the static daily spline with overnight effect.

B.4 Static weekly spline with no overnight effect

The way we redefine $\tau_0, \tau_1, \ldots, \tau_k$ here is slightly different to the static weekly spline with overnight effect.

$\widetilde{\tau}_0 < \widetilde{\tau}_1 < \cdots < \widetilde{\tau}_{k'}$ still denote the coordinates along the time-axis of the intra-day mesh, where $k' < I$, $\widetilde{\tau}_0 = 1$, $\widetilde{\tau}_{k'} = I$, and $\widetilde{\tau}_j \in \{2, \ldots, I-1\}$ for $j = 1, \ldots, k'-1$. Then the coordinates $\tau_0, \tau_1, \ldots, \tau_k$ along the time-axis of the total mesh for the whole week is defined as $\tau_0 = \widetilde{\tau}_0$ and $\tau_{ik'+j} = \widetilde{\tau}_j + iI$ for $i = 0, \ldots, 4$ and $j = 1, \ldots, k'$. (Note the difference here.) Then $(\tau_j)_{j=0}^{k}$ is still an increasing sequence. The total number of knots for one whole week is $k + 1 = 5k' + 1$.

The height of the knots are $\gamma_1^\dagger, \gamma_2^\dagger, \ldots, \gamma_{k'}^\dagger$ for Monday, $\gamma_{k'+1}^\dagger, \gamma_{k'+2}^\dagger, \ldots, \gamma_{2k'}^\dagger$ for Tuesday, and so on. We capture the weekend effect by allowing for $(\tau_k, \gamma_k^\dagger) \neq (\tau_0, \gamma_0^\dagger)$. The rest of the derivations that give $\boldsymbol{z}_j(\cdot)$ are the same as Appendix B.1. This weekly spline can capture the day-of-the-week effect by allowing for

$$(\gamma_1^\dagger, \gamma_2^\dagger, \ldots, \gamma_{k'}^\dagger)^\top \neq (\gamma_{k'+1}^\dagger, \gamma_{k'+2}, \ldots, \gamma_{2k'}^\dagger)^\top \neq \cdots \neq (\gamma_{4k'+1}^\dagger, \gamma_{4k'+2}^\dagger, \ldots, \gamma_{5k'}^\dagger)^\top.$$

The restricted weekly spline is obtained by pre-multiplying $\boldsymbol{\gamma}^\dagger$ of this weekly spline by the matrix $\mathbf{S}$ as before.

Appendix C: Additional in-sample results for Spline-DCS

Pair	USDJPY					
Spline	SDS		SDS		SWS	
Window	11		13		11	
	Estimate	Std.Error	Estimate	Std.Error	Estimate	Std.Error
ν	1.576	0.017	1.702	0.016	1.688	0.019
ξ	1.459	0.052	1.605	0.063	1.334	0.047
ζ	1.982	0.078	1.866	0.072	1.822	0.074
ω	4.414	0.088	4.184	0.084	4.638	0.080
κ_μ	0.031	0.006	0.013	0.004	0.025	0.005
$\phi_1^{(1)}$	0.439	0.091	0.581	0.024	0.427	0.090
$\phi_2^{(1)}$	0.258	0.091	0.328	0.024	0.305	0.091
$\kappa_\eta^{(1)}$	0.087	0.012	0.091	0.009	0.076	0.012
$\phi_1^{(2)}$	0.522	0.107	0.608	0.080	0.506	0.111
$\kappa_\eta^{(2)}$	0.094	0.013	0.092	0.010	0.088	0.012
ϕ_e	0.582	0.089	0.819	0.036	0.604	0.086
p	0.015	0.003	0.009	0.002	0.015	0.003
$\kappa_{e,1}$	2.854	0.021	2.527	0.021	2.239	0.021
$\kappa_{e,2}$	1.130	0.161	0.825	0.139	1.136	0.157
$\kappa_{e,3}$	NaN	NaN	0.677	0.728	NaN	NaN
Pair	EURUSD					
Spline	SDS		SDS		SWS	
Window	6		14		11	
	Estimate	Std.Error	Estimate	Std.Error	Estimate	Std.Error
ν	1.973	0.021	1.844	0.021	1.754	0.019
ξ	1.089	0.037	1.427	0.049	1.379	0.048
ζ	1.538	0.062	2.119	0.088	1.903	0.078
ω	3.366	0.072	5.067	0.073	4.156	0.050
κ_μ	0.010	0.005	0.004	0.003	0.016	0.004
$\phi_1^{(1)}$	0.486	0.011	0.550	0.017	0.473	0.081
$\phi_2^{(1)}$	0.495	0.011	0.384	0.017	0.279	0.081
$\kappa_\eta^{(1)}$	0.062	0.008	0.079	0.007	0.070	0.010
$\phi_1^{(2)}$	0.638	0.073	0.526	0.079	0.455	0.122
$\kappa_\eta^{(2)}$	0.094	0.011	0.086	0.009	0.076	0.010
ϕ_e	0.477	0.135	0.755	0.066	0.596	0.068
p	0.022	0.003	0.008	0.002	0.018	0.003
$\kappa_{e,1}$	0.575	0.089	2.632	0.021	2.830	0.021
$\kappa_{e,2}$	0.214	0.021	0.569	0.085	0.908	0.087
$\kappa_{e,3}$	0.025	2.049	0.570	0.827	NaN	NaN

Table 11. The estimated parameter values for Spline-DCS for selected in-sample windows. The standard errors are computed analytically as described in Section 3.6.1. SDS stands for the static daily spline, and SWS stands for the static (restricted) weekly spline.

Appendix D: Fourier-MEM estimation

Let $\boldsymbol{\varphi}$ denote the vector of all parameters of Fourier-MEM. Brownlees et al. (2011) show that the GMM estimator, $\widehat{\boldsymbol{\varphi}}_{IT}$, of $\boldsymbol{\varphi}$ solves the MM equation

$$\frac{1}{IT}\sum_{t=1}^{T}\sum_{\tau=1}^{I}\boldsymbol{a}_{t,\tau}u_{t,\tau}=\mathbf{0},$$

where

$$\begin{aligned}\boldsymbol{a}_{t,\tau}&=\eta_t^{-1}\nabla_{\boldsymbol{\varphi}}\eta_t+\mu_{t,\tau}^{-1}\nabla_{\boldsymbol{\varphi}}\mu_{t,\tau}+\phi_\tau^{-1}\nabla_{\boldsymbol{\varphi}}\phi_\tau+e_{t,\tau}^{*-1}\nabla_{\boldsymbol{\varphi}}e_{t,\tau}^{*},\\ u_{t,\tau}&=y_{t,\tau}/(\eta_t\phi_\tau\mu_{t,\tau}e_{t,\tau}^{*})-1.\end{aligned}$$

Under certain regularity conditions, $\widehat{\boldsymbol{\varphi}}_{IT}$ is asymptotically normal. The asymptotic covariance matrix is consistently estimated by

$$\widehat{\mathrm{Avar}}(\widehat{\boldsymbol{\varphi}}_{IT})=\frac{1}{IT}\sum_{t=1}^{T}\sum_{\tau=1}^{I}\widehat{u}_{t,\tau}^{2}\left[\sum_{t=1}^{T}\sum_{\tau=1}^{I}\boldsymbol{a}_{t,\tau}\boldsymbol{a}_{t,\tau}^{\top}\right]^{-1},$$

where $\widehat{u}_{t,\tau}^{2}=y_{t,\tau}/(\widehat{\eta}_t\widehat{\phi}_\tau\widehat{\mu}_{t,\tau}\widehat{e}_{t,\tau}^{*})-1$.

D.1 Estimation results

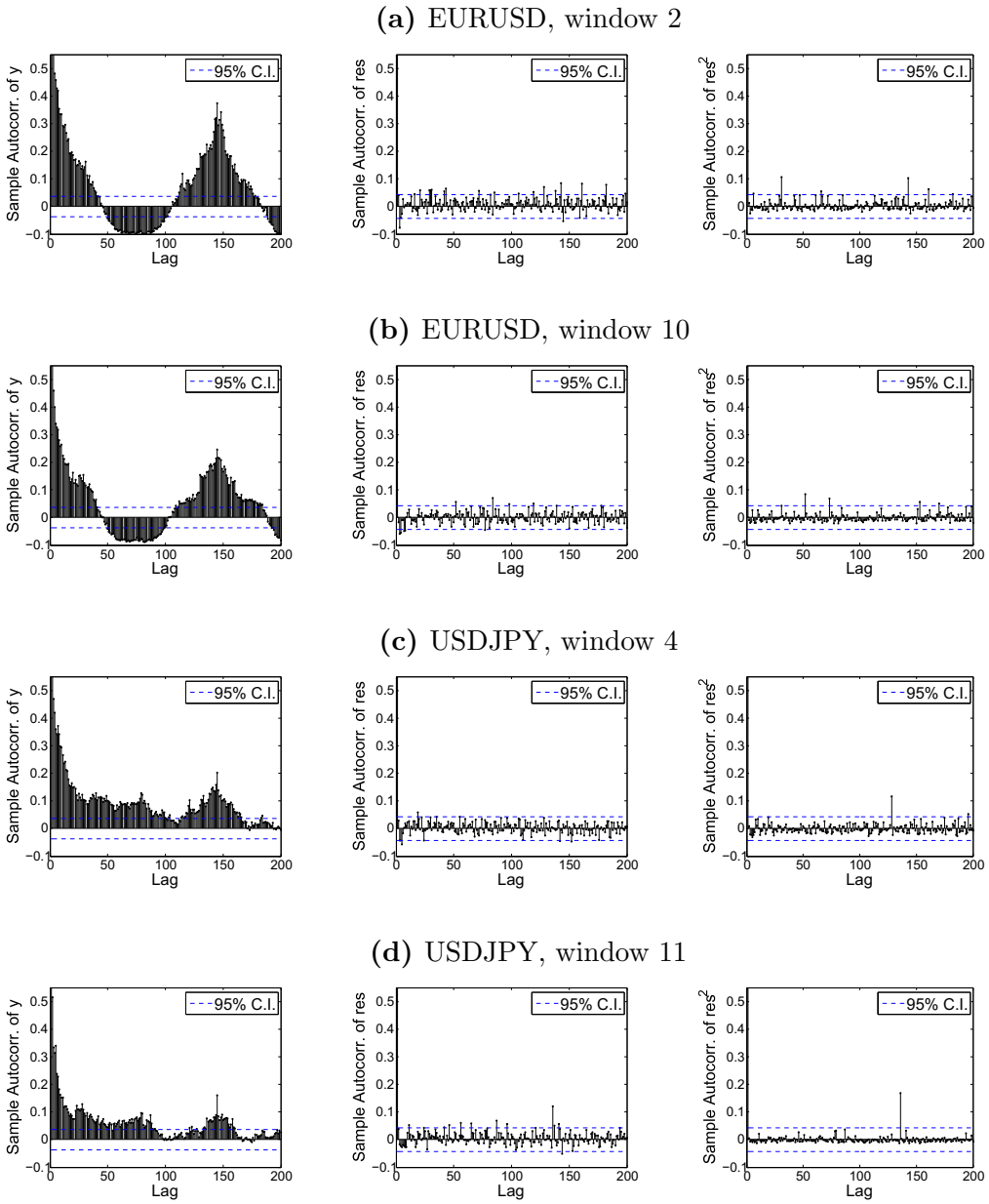


Figure 26. The sample autocorrelation of trade volume ($y_{t,\tau}$, left), $\widehat{\varepsilon}_{t,\tau}$ (middle), and $\widehat{\varepsilon}^2_{t,\tau}$ (right) in Fourier-MEM estimated by GMM. The sampling frequency is 10 minutes. The 95% confidence intervals are based on the numerical standard errors.

(a) EURUSD

	$\widehat{\varepsilon}_{t,\tau}$						$\widehat{\varepsilon}^2_{t,\tau}$					
Window	$\widehat{\rho}_1$	$\widehat{\rho}_{day}$	Q_1	Q_{day}	p-val. 1	p-val. day	$\widehat{\rho}_1$	$\widehat{\rho}_{day}$	Q_1	Q_{day}	p-val. 1	p-val. day
1	0.005	-0.005	0.060	142.769	0.807	0.000	-0.006	-0.004	0.066	11.130	0.797	1.000
2	0.038	0.021	3.186	228.207	0.074	0.000	0.012	-0.002	0.297	140.553	0.586	0.000
3	0.017	-0.035	0.610	215.108	0.435	0.000	0.001	-0.021	0.001	123.142	0.971	0.000
4	-0.004	-0.036	0.031	159.340	0.860	0.000	0.003	-0.027	0.024	147.307	0.878	0.000
5	-0.014	0.004	0.397	115.537	0.529	0.000	-0.006	-0.005	0.078	34.249	0.781	0.644
6	-0.007	0.019	0.116	149.841	0.733	0.000	-0.012	0.020	0.286	120.798	0.593	0.000
7	-0.027	0.005	1.496	144.800	0.221	0.000	-0.010	-0.005	0.196	58.338	0.658	0.024
8	0.003	0.033	0.016	144.045	0.900	0.000	0.002	0.002	0.008	27.122	0.929	0.883
9	-0.024	0.041	1.212	179.660	0.271	0.000	-0.015	0.069	0.482	119.112	0.488	0.000
10	-0.009	0.006	0.173	137.055	0.677	0.000	-0.006	-0.007	0.087	52.116	0.768	0.063
11	0.023	0.004	1.143	213.948	0.285	0.000	0.012	0.003	0.286	109.275	0.593	0.000
12	-0.003	-0.007	0.016	122.563	0.899	0.000	-0.005	-0.006	0.062	50.420	0.803	0.086
13	0.008	-0.016	0.146	160.541	0.703	0.000	0.003	-0.017	0.017	114.118	0.895	0.000
14	0.028	-0.018	1.739	162.601	0.187	0.000	0.006	0.001	0.090	103.395	0.765	0.000
15	0.049	-0.031	5.188	194.885	0.023	0.000	0.013	-0.017	0.379	167.052	0.538	0.000
16	-0.021	-0.018	0.825	132.971	0.364	0.000	-0.002	-0.002	0.007	2.656	0.935	1.000

(b) USDJPY

Window	$\widehat{\rho}_1$	$\widehat{\rho}_{day}$	Q_1	Q_{day}	p-val. 1	p-val. day	$\widehat{\rho}_1$	$\widehat{\rho}_{day}$	Q_1	Q_{day}	p-val. 1	p-val. day
1	0.040	-0.039	3.405	208.051	0.065	0.000	0.012	-0.022	0.330	95.807	0.566	0.000
2	0.054	-0.012	6.219	175.668	0.013	0.000	0.006	-0.009	0.071	91.761	0.790	0.000
3	0.034	0.016	2.548	162.420	0.110	0.000	0.008	0.018	0.140	71.724	0.708	0.001
4	0.000	0.004	0.000	129.530	0.991	0.000	0.000	0.002	0.000	78.487	0.992	0.000
5	0.023	-0.005	1.108	282.144	0.293	0.000	-0.001	-0.001	0.002	50.904	0.966	0.079
6	0.085	0.036	15.300	210.363	0.000	0.000	0.066	0.033	9.243	161.357	0.002	0.000
7	0.003	0.035	0.021	56.842	0.884	0.032	-0.002	0.004	0.007	0.678	0.935	1.000
8	0.019	0.009	0.790	151.175	0.374	0.000	0.000	0.011	0.000	29.212	0.984	0.846
9	0.039	0.028	3.260	189.737	0.071	0.000	0.009	0.048	0.171	247.381	0.680	0.000
10	0.051	0.048	5.558	213.444	0.018	0.000	0.002	0.053	0.012	130.015	0.914	0.000
11	0.030	0.003	1.929	210.740	0.165	0.000	0.002	-0.004	0.006	84.452	0.939	0.000
12	0.038	0.001	3.076	152.323	0.079	0.000	-0.001	-0.004	0.005	164.788	0.945	0.000
13	0.046	0.039	4.467	103.816	0.035	0.000	0.000	0.010	0.000	2.641	0.985	1.000
14	0.026	-0.042	1.423	169.537	0.233	0.000	0.007	-0.032	0.110	146.309	0.740	0.000
15	0.162	0.044	56.645	175.068	0.000	0.000	0.038	0.001	3.050	4.265	0.081	1.000
16	0.005	0.017	0.044	26.517	0.833	0.848	-0.001	0.000	0.003	0.351	0.957	1.000

Table 12. Residual analysis for Fourier-MEM. The sampling frequency is 10 minutes. Q_l is the Ljung-Box statistic to test the null of no autocorrelation up to the l-th lag.

Pair	EURUSD							
Freq	10mins							
Window	3		6		9		12	
	Estimate	Std.Error	Estimate	Std.Error	Estimate	Std.Error	Estimate	Std.Error
$\alpha_0^{(\eta)}$	4.239	1.707	3.064	1.577	2.981	1.343	12.726	2.649
$\beta_1^{(\eta)}$	0.504	0.016	0.410	0.019	0.486	0.017	0.516	0.018
$\alpha_1^{(\eta)}$	0.495	0.017	0.497	0.019	0.478	0.017	0.420	0.018
$\beta_1^{(e)}$	0.699	0.105	0.500	0.159	0.628	0.206	0.681	0.137
$\beta_1^{(\mu)}$	0.498	0.005	0.507	0.002	0.517	0.004	0.534	0.002
$\alpha_1^{(\mu)}$	0.336	0.004	0.353	0.002	0.322	0.003	0.358	0.002
$\alpha_2^{(\mu)}$	0.000	0.007	0.018	0.004	0.000	0.006	0.097	0.004
$\alpha_{1,1}^{(e)}$	1.291	0.457	0.936	0.521	0.221	0.123	2.110	0.530
$\alpha_{1,2}^{(e)}$	0.569	0.132	0.682	0.132	-0.105	0.874	0.302	0.137
$\alpha_{1,3}^{(e)}$	-0.609	0.864	0.582	1.001	0.792	0.920	-0.135	0.970
Pair	USDJPY							
Freq	10mins							
Window	1		6		8		10	
	Estimate	Std.Error	Estimate	Std.Error	Estimate	Std.Error	Estimate	Std.Error
$\alpha_0^{(\eta)}$	6.335	3.006	3.162	1.672	4.709	2.539	13.490	2.016
$\beta_1^{(\eta)}$	0.525	0.016	0.512	0.018	0.520	0.019	0.431	0.021
$\alpha_1^{(\eta)}$	0.498	0.017	0.505	0.019	0.503	0.020	0.516	0.023
$\beta_1^{(e)}$	0.844	0.083	0.616	0.155	0.705	0.179	0.800	0.060
$\beta_1^{(\mu)}$	0.496	0.003	0.512	0.002	0.483	0.001	0.526	0.003
$\alpha_1^{(\mu)}$	0.363	0.002	0.423	0.002	0.389	0.001	0.373	0.002
$\alpha_2^{(\mu)}$	0.075	0.005	0.005	0.004	0.041	0.002	0.000	0.004
$\alpha_{1,1}^{(e)}$	2.171	0.730	1.793	0.524	-2.015	0.865	3.763	0.709
$\alpha_{1,2}^{(e)}$	0.542	0.193	0.510	0.187	1.064	0.675	1.313	0.671
$\alpha_{1,3}^{(e)}$	0.730	0.929	0.765	0.980	0.567	0.214	0.258	0.162

Table 13. The estimated parameter values for Fourier-MEM. The results are shown for selected sampling windows. The standard errors are computed analytically using the asymptotic results outlined above. The parameters of ϕ_τ are excluded.

Appendix E: The baseline model

The out-of-sample performance of Spline-DCS and Fourier-MEM are benchmarked using an ARMA-style model, which we referred to as the baseline model. The construction of the baseline model is simple. The dynamics of volume, $y_{t,\tau}$, are decomposed into the periodic component $y^{p}_{t,\tau}$ and non-periodic component $y^{np}_{t,\tau}$ as

$$y_{t,\tau} = y^{p}_{t,\tau} + y^{np}_{t,\tau}. \tag{E.1}$$

The model assumes that the non-periodic component follows the dynamics given by

$$y^{np}_{t,\tau} = \sum_{i=1}^{p} \phi_i y^{np}_{t,\tau-i} + \varepsilon_{t,\tau} + \sum_{i=1}^{q} \theta_i \varepsilon_{t,\tau-i} + \boldsymbol{\varphi}^{\top} \boldsymbol{d}_{t,\tau}, \tag{E.2}$$

for $(t,\tau) \in \Psi_{T,I}$, where $\boldsymbol{d}_{t,\tau}$ is defined in (1) and $\dim(\boldsymbol{\varphi}) = m$. For simplicity, we assume that $\varepsilon_{t,\tau} \sim$ i.i.d. t_ν, where t_ν denotes Student's t with the degrees of freedom parameter ν. We choose $p = 1$ and $q = 0$.

Since we cannot observe the intra-day periodic component, $y^{p}_{t,\tau}$, its in-sample estimate is given by $\widehat{y}^{p}_{t,\tau} = \widehat{\widehat{p}}_{\tau} y_t$, where $y_t = \sum_{\tau=1}^{I} y_{t,\tau}$ is the total day volume. To obtain $\widehat{\widehat{p}}_{\tau}$, we first compute

$$\widehat{p}_{\tau} = \sum_{t=1}^{T} y_{t,\tau} / \sum_{t=1}^{T} \sum_{\tau=1}^{I} y_{t,\tau}.$$

$(\widehat{p}_{\tau})_{\tau=1}^{I}$ measures the proportion of daily volume, y_t, that is attributed to each intra-day bin. By construction, we have $\widehat{p}_{\tau} \in (0,1)$ and $\sum_{\tau=1}^{I} \widehat{p}_{\tau} = 1$. Since $\widehat{p}_{\tau}$ is noisy, we smooth it using the Fourier series to obtain $\widehat{\widehat{p}}_{\tau}$. This smoothing step means that $\sum_{\tau=1}^{I} \widehat{y}^{p}_{t,\tau} \neq y_t$. Then we set $\widehat{y}^{np}_{t,\tau} = y_{t,\tau} - \widehat{y}^{p}_{t,\tau}$ and estimate (E.2).

In order to produce out-of-sample forecasts, we need to forecast daily volume first. We assume that it follows the dynamics given by

$$y_t = c_{day} + \phi_{day} y_t + \varepsilon^{*}_{t}. \tag{E.3}$$

For simplicity, we assume that $\varepsilon^{*}_{t} \sim$ i.i.d.$N(0, \sigma^{2}_{*})$. We estimate (E.3) using in-sample observations of daily volume, and forecast y_t before forecasting $y_{t,\tau}$.

Over an out-of-sample period up to $H \in \mathbb{N}_{>0}$ days ahead, forecasts are generated as follows.

1. Forecast daily volume y_{T+h} for $h = 1, \ldots, H$ using (E.3). Denote the forecast quantity by $\widetilde{y}_{T+h}$.

2. The implied forecast of intra-day periodic component is

$$\widetilde{y}^{p}_{T+h,\tau} \equiv \widehat{\widehat{p}}_{\tau}\widetilde{y}_{T+h}$$

for each $\tau = 1, \ldots, I$ and $h = 1, \ldots, H$.

3. Forecast $y_{T+h,\tau}$. Its one-step forecasts are obtained by updating (E.2) by plugging in the errors, $\widetilde{\varepsilon}_{t,\tau} = y_{T+h,\tau} - \widetilde{y}^{p}_{T+h,\tau} - \widetilde{y}^{np}_{T+h,\tau}$, without updating in-sample parameter estimates.

4. Forecast of $y_{T+h,\tau}$ is

$$\widetilde{y}_{T+h,\tau} = \widetilde{y}^{p}_{T+h,\tau} + \widetilde{y}^{np}_{T+h,\tau}.$$

Since the baseline model has an ARMA-based structure, its out-of-sample forecasts are the first (conditional) moment forecasts of volume.

To be noted that this baseline model does not ensure that its volume forecasts are positive. Also, Step 3 means that any forecast errors in intra-day volume prediction is due to the non-periodic component, $\widetilde{y}^{np}_{t,\tau}$, which may or may not be the case. Although these aspects of the model are unintuitive, we do not think that they pose practical issues, since it serves merely as a benchmark and yielded positive volume predictions in our application.

It is also to be noted that this model does not guarantee that the forecast intra-day volume, $\widetilde{y}_{T+h,\tau}$, add up to the forecast day volume $\widetilde{y}_{T+h}$ we obtain using (E.3). This misalignment stems from our assumption regarding $y^{p}_{t,\tau}$ and the fact that we assume y_t and $y_{t,\tau}$ follow separate dynamics given by (E.2) and (E.3). That is, instead of assuming that $y^{p}_{t,\tau} = y^{p}_{t',\tau}$ for any $t \neq t'$ and all $\tau = 1, \ldots, I$, we choose to fix $(p_\tau)^{I}_{\tau}$ to be the same every day given that the day-of-the-week effect was not obvious in Figure 3. Since we use the baseline model merely as a benchmark to compare different models and specifications, we do not think that this misalignment affects the conclusions of this paper.