Quantitative Finance

ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: www.tandfonline.com/journals/rquf20

Forecasting limit order book liquidity supply–demand curves with functional autoregressive dynamics

Ying Chen, Wee Song Chua & Wolfgang Karl Härdle

To cite this article: Ying Chen, Wee Song Chua & Wolfgang Karl Härdle (2019) Forecasting limit order book liquidity supply–demand curves with functional autoregressive dynamics, Quantitative Finance, 19:9, 1473-1489, DOI: 10.1080/14697688.2019.1622290

To link to this article: https://doi.org/10.1080/14697688.2019.1622290

Quantitative Finance, 2019
Vol. 19, No. 9, 1473–1489, https://doi.org/10.1080/14697688.2019.1622290

Forecasting limit order book liquidity supply–demand curves with functional autoregressive dynamics

YING CHEN†‡§, WEE SONG CHUA*‡ and WOLFGANG KARL HÄRDLE¶‖

†Department of Mathematics, Faculty of Science, National University of Singapore, Block S17, Level 8, 2 Science Drive 2, Singapore 117543, Singapore
‡Department of Statistics & Applied Probability, Faculty of Science, National University of Singapore, Block S16, Level 7, 6 Science Drive 2, Singapore 117546, Singapore
§Risk Management Institute, National University of Singapore, 21 Heng Mui Keng Terrace, I^3 Building #04-03, Singapore 119613, Singapore
¶Ladislaus von Bortkiewicz Chair of Statistics, C.A.S.E. – Center for Applied Statistics & Economics, Humboldt-Universität zu Berlin, Unter den Linden 6, Berlin 10099, Germany
‖Sim Kee Boon Institute for Financial Economics, Singapore Management University, 81 Victoria Street, Singapore 188065, Singapore

(*Received 28 March 2018; accepted 6 May 2019; published online 9 July 2019*)

We develop a dynamic model to simultaneously characterize the liquidity demand and supply in a limit order book. The joint dynamics are modeled in a unified Vector Functional AutoRegressive (VFAR) framework. We derive a closed-form maximum likelihood estimator under sieves and establish asymptotic consistency of the proposed method under mild conditions. We find the VFAR model presents strong interpretability and accurate out-of-sample forecasts. In application to limit order book records of 12 stocks in the NASDAQ, traded from 2 January 2015 to 6 March 2015, the VFAR model yields R^2 values as high as 98.5% for in-sample estimation and 98.2% in out-of-sample forecast experiments. It produces accurate 5-, 25- and 50-min forecasts, with RMSE as low as 0.09–0.58 and MAPE as low as 0.3–4.5%. The predictive power stably reduces trading cost in the order splitting strategies and achieves excess gains of 31 basis points on average.

Keywords: Liquidity demand and supply curves; Order splitting strategy; Vector functional autoregression; Liquidity forecasting; Time series

JEL Classification: C13, C32, C53

1. Introduction

Liquidation of large orders has attracted much attention from researchers and practitioners. Markets address the large order liquidation problem in one of three ways: call auctions, dealer markets and limit order books, see Foucault *et al.* (2005). Among them, Limit Order Book (LOB) has emerged as the main source for liquidity and exhibits a growing importance worldwide. LOB records investors' orders on both the bid and ask sides with price and volume constraints. With a limit order, investors can improve the execution price, either buying or selling, according to their choice, but the execution is not immediate as a market order, nor guaranteed. Harris (1990) defines three components of liquidity as being: (1) tightness such as bid–ask spread at the best price level; (2) depth measured by quantities, e.g. eXchange Liquidity Measure (XLM); and (3) resilience referring to the recovery for deviations of spreads from their competitive level. As such, LOB contains the comprehensive information on market liquidity, not only a single-valued liquidity measurement at the best bid–ask price level, but also the queuing liquidities at deeper levels in the book. In this paper, we develop a dynamic model to simultaneously characterize the liquidity demand and supply in the LOB. The objectives are to understand the joint dynamics of liquidity at multiple levels and on both the

*Corresponding author. Email: a0054070@u.nus.edu

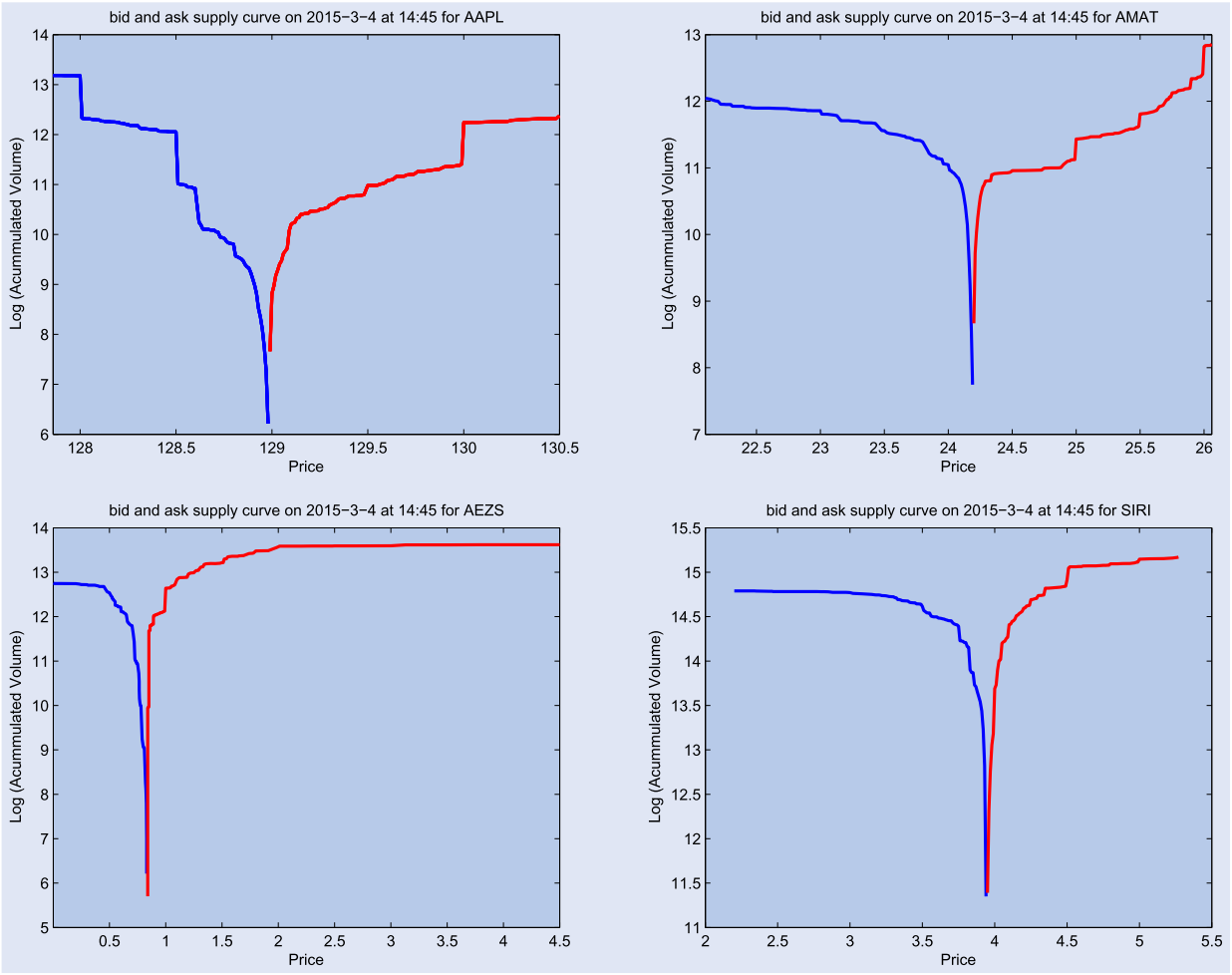

Figure 1. AAPL, AMAT, AEZS, and SIRI liquidity demand and supply curves at an arbitrarily selected time point. In our study, AAPL has the largest market value; AEZS has the smallest value and the smallest bid–ask spread on average; SIRI is the most active stock; and AMAT is the relatively less active stock among the rest.

bid and ask sides of the LOB, and explore the applicability of a dynamic model in the order splitting strategy.

Liquidity in the LOB can be well represented by a pair of demand and supply curves that are defined as the (log) accumulated volumes on the quoted prices. The demand curve corresponds to the bid side and the supply is associated with the ask side. As an illustration, figure 1 displays the liquidity curves of four stocks: Apple Inc. (AAPL); Applied Materials, Inc. (AMAT); AEterna Zentaris Inc. (AEZS); and Sirius XM Holdings Inc. (SIRI), based on the snapshots of book on 4 March 2015 at 14:45:00. Each pair of demand and supply curves forms a V-shape that is monotonically decreasing on the bid side and monotonically increasing on the ask side. The gap at the center represents the bid–ask spread, i.e. the *market tightness* at the best price level. Moreover, the *market depth* is reflected by the gradients of the liquidity curves. More precisely, the steeper the curves are, the less price impact there is for large orders, and thus the more liquidity is ready to be supplied or consumed in the market. Liquidity is concentrated on relatively few quoted prices near the best bid and ask prices, while the tails are relatively flat. This flattening out of the tail, or the gentle gradient in the tails, implies low liquidity. A buy or sell of large volumes at the extreme prices will trigger a drastic change in the price and thus increase trading cost. The dynamic dependence of the series on the liquidity demand and supply curves naturally inherits the *market resilience*, the third component of liquidity.

Liquidity is serially dependent, i.e. the current value of liquidity depends on its own past values. Though with limited information, the popular single-valued liquidity measures are found to be serially dependent; e.g. bid–ask spread (see Benston and Hagerman 1974, Stoll 1978, Fleming and Remolona 1999) and XLM (see Cooper *et al.* 1985, Gomber *et al.* 2015). These findings motivate as a first proxy the adoption of autoregressive models for liquidity in the LOB. Groß–Klußmann and Hautsch (2013) propose a long memory autoregressive conditional Poisson model for the quoted bid–ask spreads. Huberman and Halka (2001) evidence the serial dependence of bid–ask spread and market depth in an autoregressive framework. Härdle *et al.* (2015) propose a local adaptive multiplicative error model to forecast the high-frequency series of 1-min cumulative trading volumes of several NASDAQ blue chip stocks. Chordia *et al.* (2005) document in a vector autoregressive model the cross-sectional dependence among the liquidity measures of bid–ask spread, market depth and order flow, and other statistics of volatility and returns in the stock and bond markets, where the liquidity measures depend on both their own past values and the

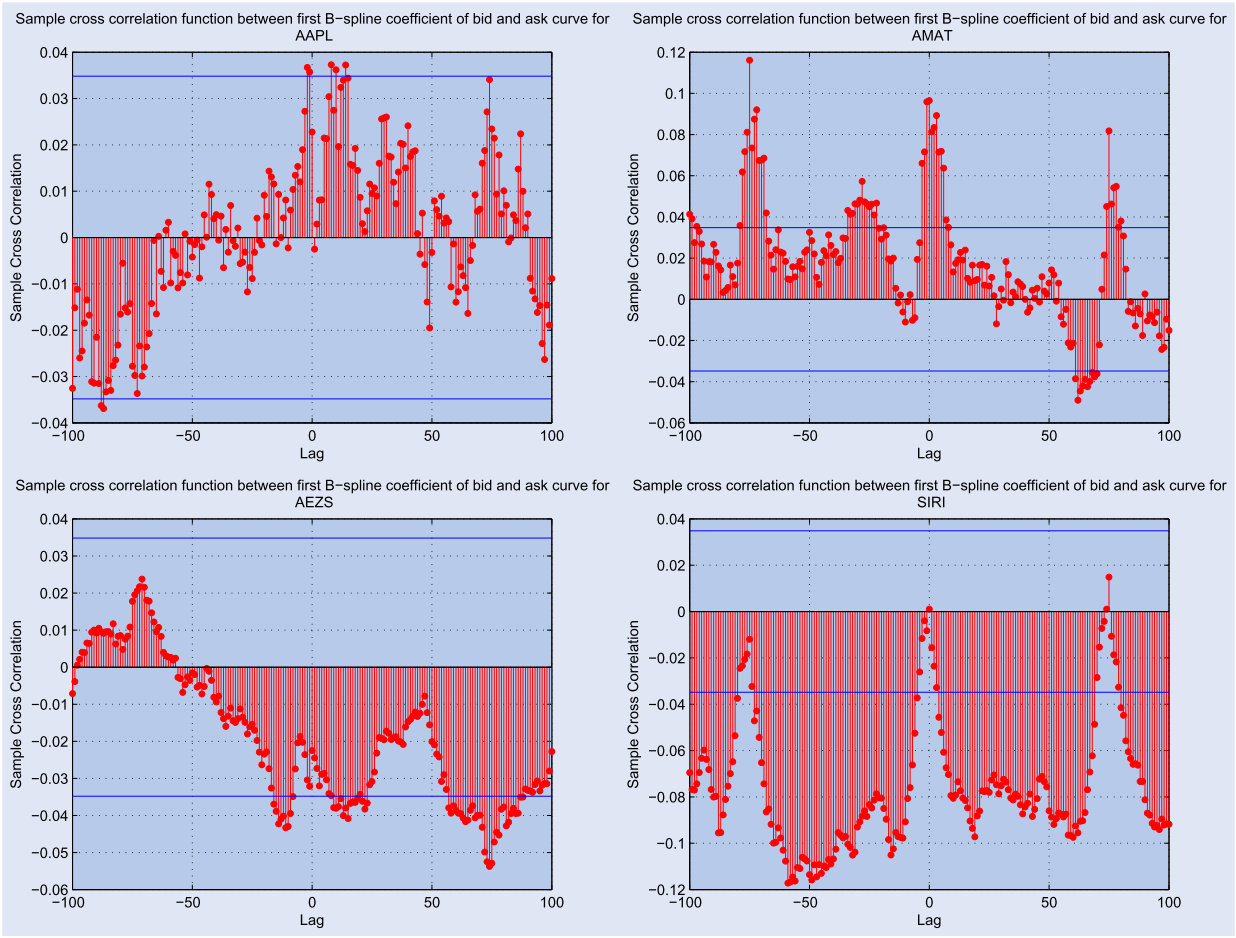

Figure 2. Sample cross-correlation function between first *B*-spline coefficient of the bid and ask curve for AAPL, AMAT, AEZS, and SIRI.

historical values of the other measures. Härdle *et al.* (2012) empirically analyze the seasonally-adjusted liquidity supply curves in the LOB using a dynamic semiparametric factor model, where the extracted factors of the curves are modeled in the Vector Error Correction (VEC) framework.

It is worth noting that there also exists serial cross-dependence in the bivariate series of liquidity demand and supply. The impact of public information on limit orders result in similar changes on both bid and ask sides, and can cause investors to switch from one side to the other. Thus it introduces lead–lag cross-dependence in both sides of liquidity. The joint serial cross-dependence suggests richer dynamics should be utilized in liquidity analysis. Building on this idea, we develop a Vector Functional AutoRegressive (VFAR) model to describe the joint dynamics of the bivariate series—liquidity demand and supply curves on the bid and ask sides of an electronic open LOB—simultaneously in a unified framework. While there is in general co-integration or a common trend in demand and supply when both variables are scalar time series, the model features the dependence of the series of curves, which is stable over time. Figure 2 displays the sample cross-correlations between the fundamental representatives of the demand and supply curves of the four stocks (the 'first *B*-spline coefficients' of the bid and ask curves). This first proxy shows that there are significant lead–lag cross-dependencies between the demand and supply curves, but the patterns deviate far from the persistence of a unit root.

In the VFAR model, we derive a closed-form maximum likelihood estimator under a sieve and establish the asymptotic consistency of the method. We investigate the finite sample performance of the proposed model along with the LOB records of 12 stocks traded in NASDAQ from 2 January 2015 to 6 March 2015, where the stocks are carefully selected to represent various types with different market capitalization and liquidity. We find the VFAR presents a strong predictability in liquidity, producing R^2 values as high as 98.5 % for in-sample estimation and 98.2 % in out-of-sample forecast experiments. Moreover, it yields accurate 5-, 25- and 50-min forecasts, with RMSE as low as 0.09–0.58 and MAPE as low as 0.3 –4.5%. Finally, the predictive power stably reduces trading cost in the order splitting strategies and achieves excess gains of 31 basis points on average.

We would like to highlight the difference between our study and the existing ones in the literature. Above all, we develop a dynamic model to estimate and forecast liquidity demand and supply curves simultaneously and demonstrate the application in order splitting execution. Härdle *et al.* (2012) implement a dynamic semiparametric factor model for liquidity curves, but the extracted factors are handled separately on each side, although it is possible to capture the joint evolution of bid and ask sides. Secondly, we develop the FAR modeling for multiple functional time series, where the continuous curves are modeled in a convolutional VFAR that is stationary. Whereas in the functional time series literature,

Bosq (2000) has proposed the FAR model for univariate functional time series and developed Yule–Walker estimation (see also Besse *et al*. 2000, Guillas 2001, Antoniadis and Sapatinas 2003, Chaudhuri *et al*. 2016). Mourid and Bensmain (2006) propose a maximum likelihood estimation with Fourier expansions, as far as we know this is the first work to model multiple functional time series. We investigate its theoretical properties, and within a maximum likelihood estimator approach based on *B*-spline expansions, we provide more flexibility in fitting beyond the Fourier expansion in Chen and Li (2017). Although the implementation focuses on the bivariate liquidity demand and supply curves in our study, the developed model is general and can be used for analyzing multiple functional time series in other research areas.

This paper is structured as follows. In Section 2, we describe the LOB data. Section 3 introduces the VFAR modeling in detail including the estimation approach and the theoretical properties. Section 4 presents the modeling setup and in-sample estimation results, reports the out-of-sample forecast results and demonstrates the application to an order execution strategy. Section 5 provides concluding remarks. All of the theoretical proofs are contained in the Appendix.

2. Data

We consider the LOB records of 12 stocks from 2 January 2015 to 6 March 2015 (44 trading days). The LOB records contain the quoted prices and volumes up to 100 price levels on each side of ask and bid. All the quotes are timestamped with decimal precision up to nanoseconds ($= 10^{-9}$ s). In total, the (buy or sell) order book contains 400 values from the best ask price, best ask volume, best bid price, and best bid volume until the 100-th best ask (bid) price and corresponding volume. The data was obtained from LOBSTER through the Research Data Center of the Collaborative Research Center 649 (https://sfb649.wiwi.hu-berlin.de/fedc/). Note that the records of the 12 stocks in LOBSTER only contain the information in the National Association of Securities Dealers Automated Quotations (NASDAQ) stock market. NASDAQ is a continuous auction trading platform where the normal continuous trading hours are between 9:30 am to 4:00 pm from Monday to Friday. During the normal trading, if an order cannot be executed immediately or completely, the remaining volumes are queued in the bid and ask sides according to a strict price-time priority order.

The stocks correspond to high variations in terms of market capitalization, liquidity tightness and depth. They are Apple Inc. (AAPL), Microsoft Corporation (MSFT), Intel Corporation (INTC), Cisco Systems, Inc. (CSCO), Sirius XM Holdings Inc. (SIRI), Applied Materials, Inc. (AMAT), Comcast Corporation (CMCSA), AEterna Zentaris Inc. (AEZS), eBay Inc. (EBAY), Micron Technology, Inc. (MU), Whole Foods Market, Inc. (WFM), and Starbucks Corporation (SBUX). The largest stock is AAPL with market value of USD737.41 billions, and the smallest is AEZS with market value of USD35.38 millions. When considering the 5-min queueing volume in the LOB, the most active stock is SIRI, with value of 3.73 millions on the bid side and 7.61 millions on the ask side. The lest active stocks are CMCSA with value of 0.02 millions on the bid side and SBUX with value of 0.03 millions on the ask side. Moreover, the average value of the bid–ask spread varies from 0.0062 (AEZS) to 0.0213 (SBUX), see table 1.

In data pre-processing, we remove the first 15 min after opening and the last 5 min before closing to eliminate the market opening and closing effect. The accumulated bid and ask volumes are log-transformed when constructing liquidity curves to reduce the impact of extraordinarily large volumes. The liquidity curves are smoothed over the 100 price levels of LOB on each side using *B*-spline basis functions. Moreover, to remove the impact of microstructure noise, the sampling frequency is set to be 5 min for a good strike between bias and variance, see AitSahalia *et al*. (2005), Zhang *et al*. (2005), and Härdle *et al*. (2018). As such, there are 75 pairs of bid and ask liquidity curves on each day for each stock. Over the whole sample period of 44 trading days, it amounts to 3300 pairs of bid and ask supply curves for each stock.

The liquidity curves exhibit significant serial dependence over time. As an illustration, figure 3 shows the sample cross-correlations between the log-accumulated volumes at best bid and ask prices for four representative stocks including AAPL with the largest market value, AEZS with the smallest value

Table 1. Summary statistics on liquidity measures for the 12 stocks traded in NASDAQ.

		Bid vol		Ask vol	
Ticker symbol	Mean spread (USD)	Min	Max	Min	Max
AAPL	0.0125	52,267	710,020	61,305	1,298,696
MSFT	0.0101	90,344	928,319	122,377	621,471
INTC	0.0102	158,900	557,251	146,959	1,142,641
CSCO	0.0101	134,790	1,316,058	266,455	4,458,672
SIRI	0.0101	1,266,528	3,725,304	3,002,680	7,605,467
AMAT	0.0102	78,944	334,794	180,749	787,983
CMCSA	0.0106	23,668	128,916	40,638	146,724
AEZS	0.0062	145,635	767,785	472,689	1,158,740
EBAY	0.0110	42,060	160,572	52,813	415,033
MU	0.0107	95,907	497,910	102,357	595,200
WFM	0.0153	34,538	114,386	41,019	159,488
SBUX	0.0213	27,467	151,022	34,914	166,932

Note: Sampling frequency is 5 min.

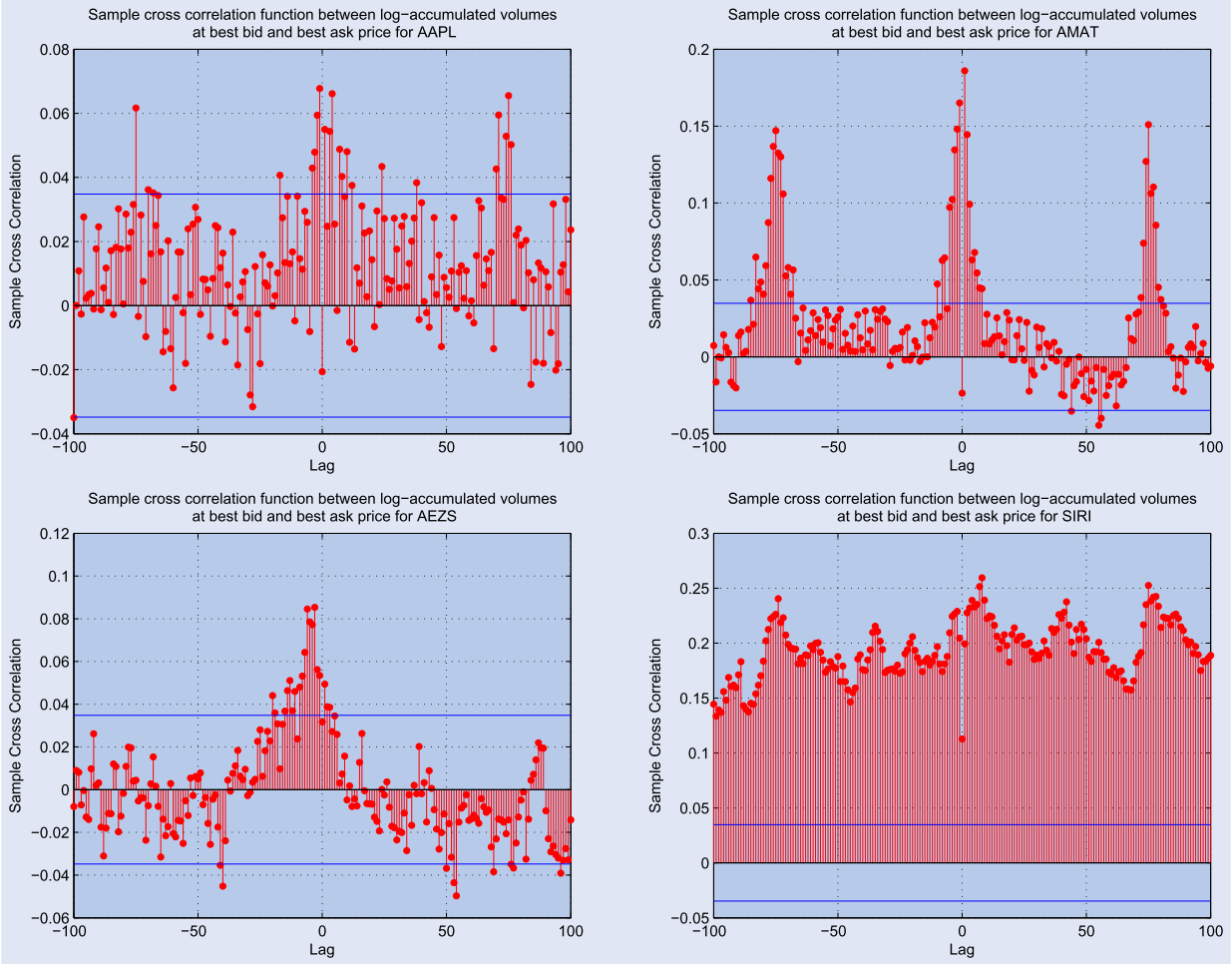

Figure 3. Sample cross-correlation function between log-accumulated volumes at best bid and ask price for AAPL, AMAT, AEZS, and SIRI.

and the smallest bid–ask spread on average, SIRI the most active stock, and AMAT the relatively mid active stock among the rest. While the simultaneous dependence between the bid and ask sides is insignificant or negatively correlated, there are significant positive serial correlations on the first lagged values of the opposite side and then decays for higher order. Similar features are observed in the other eight stocks, which are not displayed. The bid–ask cross-dependency motivates analyzing the liquidity demand and supply jointly.

In addition, the serial cross-dependence between the series of curves is reflected in figure 2, where the sample cross-correlations are computed based on the first B-spline coefficients of the bid and ask curves, which will be detailed later. It shows that there are significant lead–lag cross-dependence between the demand and supply curves, but the patterns deviate far from the persistence of unit root. We also perform the Johansen (1991) co-integration test to assess the null hypothesis of co-integration of bid and ask sides. The results support rejecting co-integration.

3. Vector functional autoregression

In this section, we present the Vector Functional AutoRegressive (VFAR) model that is to describe the joint serial cross-dependence of multiple series of continuous curves. We show how to estimate the functional parameters of the bivariate liquidity demand and supply curves, with the help of B-spline expansions and sieve method. A closed-form estimator is presented and asymptotic consistency provided.

The liquidity curves are defined on the quoted price, denoted as τ, which is assumed to be exogenous. The quoted prices are observed on a dense tick grid, that is re-scaled separately for bid and ask sides into a continuous interval $[0, 1]$, with the minimum price as 0 and maximum price as 1. Denote by $X_t^{(a)}(\tau)$ and $X_t^{(b)}(\tau)$ the liquidity supply and demand curves on the ask (a) side and the bid (b) side at $t = 1, \ldots, n$. The curves are realizations of a functional stochastic process in the space $C_{[0,1]}$ of real continuous functions on $[0, 1]$. In other words, each pair of the liquidity curves is one functional object. Over time, the curves form bivariate time series of n functional objects, each on the bid and ask sides. In our study, the curves are obtained by smoothing over the discrete log-accumulated volumes against the quoted prices at every time point.

To handle the joint dynamics of the two continuous liquidity curves, we propose a Vector Functional AutoRegressive (VFAR) model of order p:

$$\begin{bmatrix} X_t^{(a)} - \mu_a \\ X_t^{(b)} - \mu_b \end{bmatrix} = \sum_{k=1}^{p} \begin{bmatrix} \rho^{aa,k} & \rho^{ab,k} \\ \rho^{ba,k} & \rho^{bb,k} \end{bmatrix} \begin{bmatrix} X_{t-k}^{(a)} - \mu_a \\ X_{t-k}^{(b)} - \mu_b \end{bmatrix} + \begin{bmatrix} \varepsilon_t^{(a)} \\ \varepsilon_t^{(b)} \end{bmatrix}, \tag{1}$$

where the operators $\rho^{aa,k}$, $\rho^{ab,k}$, $\rho^{ba,k}$, and $\rho^{bb,k}$ measure the serial cross-dependence among the liquidity demand and supply curves on their kth lagged values. The operators are bounded linear operator from $\mathcal{H}$ to $\mathcal{H}$, a real separable Hilbert space endowed with its Borel σ-algebra $B_{\mathcal{H}}$. The mean function is denoted as $(\mu_a(\tau), \mu_b(\tau))^\top \stackrel{\text{def}}{=} (\mathsf{E}[X_t^{(a)}(\tau)], \mathsf{E}[X_t^{(b)}(\tau)])^\top$. Under stationarity, both the serial cross-dependence and the mean are constant over time. The innovations $\{\varepsilon_t^{(a)}\}_{t=1}^n$ and $\{\varepsilon_t^{(b)}\}_{t=1}^n$ are strong $\mathcal{H}$-white noise, independently and identically distributed with zero mean and finite second moment, $0 < \mathsf{E}\|\varepsilon_1^{(a)}\|^2 = \cdots = \mathsf{E}\|\varepsilon_n^{(a)}\|^2 < \infty$ and $0 < \mathsf{E}\|\varepsilon_1^{(b)}\|^2 = \cdots = \mathsf{E}\|\varepsilon_n^{(b)}\|^2 < \infty$, where the norm $\|\cdot\|$ is induced from the inner product $\langle\cdot,\cdot\rangle$ of $\mathcal{H}$. The innovation processes $\varepsilon_t^{(a)}$ and $\varepsilon_t^{(b)}$ need not be cross-independent.

In the following, we derive the estimation for the VFAR model of order 1, which can be generalized for higher order. For the notational simplification, the superscript k is dropped. We consider the convolutional VFAR, where each operator ρ is represented by a convolution kernel Hilbert–Schmidt operator,

$$\begin{aligned}
&X_t^{(a)}(\tau) - \mu_a(\tau) \\
&\quad = \int_0^1 \kappa_{ab}(\tau - s)\big\{X_{t-1}^{(b)}(s) - \mu_b(s)\big\}\,\mathrm{d}s \\
&\qquad + \int_0^1 \kappa_{aa}(\tau - s)\big\{X_{t-1}^{(a)}(s) - \mu_a(s)\big\}\,\mathrm{d}s + \varepsilon_t^{(a)}(\tau), \\
&X_t^{(b)}(\tau) - \mu_b(\tau) \\
&\quad = \int_0^1 \kappa_{bb}(\tau - s)\big\{X_{t-1}^{(b)}(s) - \mu_b(s)\big\}\,\mathrm{d}s \\
&\qquad + \int_0^1 \kappa_{ba}(\tau - s)\big\{X_{t-1}^{(a)}(s) - \mu_a(s)\big\}\,\mathrm{d}s + \varepsilon_t^{(b)}(\tau). \quad (2)
\end{aligned}$$

The kernel function $\kappa_{xy} \in L^2([0,1])$ and $\|\kappa_{xy}\|_2 < 1$ for $xy = aa$, ab, ba, and bb, where $\|\cdot\|_2$ denotes the L^2 norm in $C_{[0,1]}$. Note that a linear operator ρ on a Hilbert space $\mathcal{H}$ with norm $\|\cdot\|$ and inner product $\langle\cdot,\cdot\rangle$ is Hilbert–Schmidt if $\rho(\cdot) = \sum_j \lambda_j \langle\cdot, e_j\rangle f_j$, where $\{e_j\}$ and $\{f_j\}$ are orthonormal bases of $\mathcal{H}$ and $\{\lambda_j\}$ is a real sequence such that $\sum_j \lambda_j^2 < \infty$.

Expand the functional terms in (2) using the B-spline basis functions in $L^2([0,1])$:

$$B_{j,m}(\tau) = \frac{\tau - w_j}{w_{j+m-1} - w_j} B_{j,m-1}(\tau) + \frac{w_{j+m} - \tau}{w_{j+m} - w_{j+1}} B_{j+1,m-1}(\tau), \quad m \geq 2,$$

where m is the order, $w_1 \leq \cdots \leq w_{J+m}$ denote the sequence of knots, and

$$B_{j,1}(\tau) = \begin{cases} 1 & \text{if } w_j \leq \tau < w_{j+1}, \\ 0 & \text{otherwise.} \end{cases}$$

Plug-in the B-spline expansions to the VFAR model (2), we obtain the relationship of the B-spline coefficients in the framework of VFAR:

$$\begin{aligned}
d_{t,h}^a &= p_h^a + d_h^a(\varepsilon_t^{(a)}) + \sum_{i=1}^{\infty} \\
&\quad \times \left\{ \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{aa} - c_h^{aa} \right\} \\
&\quad \times \frac{w_{i+m} - w_i}{m} d_{t-1,i}^a + \sum_{i=1}^{\infty} \\
&\quad \times \left\{ \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{ab} - c_h^{ab} \right\} \\
&\quad \times \frac{w_{i+m} - w_i}{m} d_{t-1,i}^b, \\
d_{t,h}^b &= p_h^b + d_h^b(\varepsilon_t^{(b)}) + \sum_{i=1}^{\infty} \\
&\quad \times \left\{ \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{bb} - c_h^{bb} \right\} \\
&\quad \times \frac{w_{i+m} - w_i}{m} d_{t-1,i}^b + \sum_{i=1}^{\infty} \\
&\quad \times \left\{ \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{ba} - c_h^{ba} \right\} \\
&\quad \times \frac{w_{i+m} - w_i}{m} d_{t-1,i}^a, \quad (3)
\end{aligned}$$

where $d_{t,j}^a$ and $d_{t,j}^b$ are the B-spline coefficients for the observed functional data $X_t^{(a)}$ and $X_t^{(b)}$, respectively; p_h^a are the coefficients associated with the expansion of the mean on the ask side $\mu_a(\tau) - \int_0^1 \kappa_{ab}(\tau - s)\mu_b(s)\,\mathrm{d}s - \int_0^1 \kappa_{aa}(\tau - s)\mu_a(s)\,\mathrm{d}s$ and p_h^b are the coefficients on the bid side for $\mu_b(\tau) - \int_0^1 \kappa_{bb}(\tau - s)\mu_b(s)\,\mathrm{d}s - \int_0^1 \kappa_{ba}(\tau - s)\mu_a(s)\,\mathrm{d}s$; $d_j^a(\varepsilon_t^{(a)})$ and $d_j^b(\varepsilon_t^{(b)})$ are the B-spline coefficients for the unknown innovations $\varepsilon_t^{(a)}$ and $\varepsilon_t^{(b)}$, respectively; and c_j^{aa}, c_j^{ab}, c_j^{ba}, and c_j^{bb} are the B-spline coefficients for the unknown kernel functions κ_{aa}, κ_{ab}, κ_{ba}, and κ_{bb}, respectively. As such, the original problem of estimating the functional parameters can now be equivalently solved by the estimation of the B-spline coefficients in (3).

Given a finite sample of the functional objects, it is however impossible to estimate the infinite coefficients in (3) for $i, j = 1, \ldots, \infty$. The estimation is conducted with the help of a sieve.

3.1. *Maximum likelihood estimator under sieve*

We introduce a sequence of subsets—named sieve for the parameter space Θ, which is denoted by $\{\Theta_{J_n}\}$ with $J_n \to +\infty$ as $n \to +\infty$, see e.g. Grenander (1981) on the theory of sieves. In other words, the dimension of the subset is allowed to increase with the sample size. We have $\Theta_{J_n} \subseteq \Theta_{J_n+1}$ and the union of subsets $\bigcup \Theta_{J_n}$ is dense in the parameter space.

The sieve is defined as follows:

$$\Theta_{J_n} = \left\{ \kappa_{xy} \in L^2 \mid \kappa_{xy}(\tau) = \sum_{l=1}^{J_n} c_l^{xy} B_{l,m}(\tau), \tau \in [0,1], \times \sum_{l=1}^{J_n} l^2 (c_l^{xy})^2 \le v J_n \right\}, \tag{4}$$

where v is some known positive constant such that the constraint on c_l^{xy} can be satisfied generally without sacrifice of the growth rate of J_n. We will show the estimation in the finite subsets of the parameter space.

Under the sieve with J_n, Equation (3) can be represented in a form as follows:

$$y_t = v + C y_{t-1} + u_t, \tag{5}$$

where $y_t = (d^a_{t,1}, \ldots, d^a_{t,J_n}, d^b_{t,1}, \ldots, d^b_{t,J_n})^\top$, $v = (p^a_1, \ldots, p^a_{J_n}, p^b_1, \ldots, p^b_{J_n})^\top$, $u_t = (d^a_1(\varepsilon^{(a)}_t), \ldots, d^a_{J_n}(\varepsilon^{(a)}_t), d^b_1(\varepsilon^{(b)}_t), \ldots, d^b_{J_n}(\varepsilon^{(b)}_t))^\top$, and $C = \left[\begin{smallmatrix} R^{aa} R^{ab} \\ R^{ba} R^{bb} \end{smallmatrix}\right]$ with R^{xy} being a $J_n \times J_n$ matrix with elements $r^{xy}_{h,i} = \{\sum_{j=1}^{J_n}((w_{j+m} - w_{j+1})/(w_{j+m} - w_j) - (w_{j+m+1} - w_{j+2})(w_{j+m+1} - w_{j+1}))c_j^{xy} - c_h^{xy}\}((w_{i+m} - w_i)/m)$, for $xy = aa$, ab, ba, and bb.

We impose an assumption that the B-spline coefficients $d^a_j(\varepsilon^{(a)}_t)$ and $d^b_j(\varepsilon^{(b)}_t)$ are independently and identically Gaussian distributed with mean zero and constant variance $\sigma^2_{j,a}$ and $\sigma^2_{j,b}$, respectively. Following Geman and Hwang (1982), we define the likelihood function for (5) over the approximating subspace (4) of the original parameter space. The transition density is as follows:

$$\begin{aligned} & g\left(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \rho^{aa}, \rho^{ab}, \rho^{ba}, \rho^{bb}\right) \\ &= \frac{1}{(2\pi)^{Kn/2}} \left| I_n \otimes \Sigma_u \right|^{-1/2} \\ &\quad \times \exp\left\{ -\frac{1}{2} \left(\mathbf{y} - (Z^\top \otimes I_K)\boldsymbol{\beta} \right)^\top (I_n \otimes \Sigma_u^{-1}) \right. \\ &\quad \left. \left(\mathbf{y} - (Z^\top \otimes I_K)\boldsymbol{\beta} \right) \right\}, \end{aligned}$$

where $\mathbf{y} = \mathrm{vec}(y_1, \ldots, y_n)$, $Z = \left[\begin{smallmatrix} 1 & \cdots & 1 \\ y_0 & \cdots & y_{n-1} \end{smallmatrix}\right]$, $\boldsymbol{\beta} = \mathrm{vec}(v, C)$, $\mathbf{u} = \mathrm{vec}(u_1, \ldots, u_n)$, $K = 2J_n$, I_n is an $n \times n$ identity matrix, and vec is the column stacking operator.

The Maximum Likelihood Estimators (MLEs) are obtained with closed-form:

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= \left\{ (ZZ^\top)^{-1} Z \otimes I_K \right\} \mathbf{y} \quad \text{orequivalently,} \\ \widehat{B} &= (\widehat{v}, \widehat{C}) = YZ^\top (ZZ^\top)^{-1}, \\ \widehat{\Sigma_u} &= \frac{1}{n}(Y - BZ)(Y - BZ)^\top, \end{aligned} \tag{6}$$

where the first column of $\widehat{B}$ in (6) contains the estimators of coefficients for the mean function $v = (p^a_1, \ldots, p^a_{J_n}, p^b_1, \cdots, p^b_{J_n})^\top$. Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_1, \theta_2, \ldots, \theta_2)$, with $\theta_1 = (c^{aa}_1, \ldots, c^{aa}_{J_n}, c^{ba}_1, \ldots, c^{ba}_{J_n})^\top$ and $\theta_2 = (c^{ab}_1, \ldots, c^{ab}_{J_n}, c^{bb}_1, \ldots, c^{bb}_{J_n})^\top$, such that $\boldsymbol{\theta}$ contains J_n columns of θ_1 and J_n columns of θ_2. The estimator for c_j^{xy} for $xy = aa, ab, ba, bb$ is as follows:

$$\widehat{\boldsymbol{\theta}} = Q^{-1} YZ^\top (ZZ^\top)^{-1} (\mathbf{0}_{2J_n \times 1}, I_{2J_n})^\top W,$$

where $W = \mathrm{diag}(m/(w_{1+m} - w_1), \ldots, m/(w_{J_n+m} - w_{J_n}), m/(w_{1+m} - w_1), \ldots, m/(w_{J_n+m} - w_{J_n}))$, $Q = \left[\begin{smallmatrix} Q^1 & \mathbf{0}_{J_n \times J_n} \\ \mathbf{0}_{J_n \times J_n} & Q^1 \end{smallmatrix}\right]$, with Q^1 being a $J_n \times J_n$ matrix with elements in the jth diagonal equals $q_j - 1$ and the remaining elements in the jth column equals q_j, $q_j = (w_{j+m} - w_{j+1})/(w_{j+m} - w_j) - (w_{j+m+1} - w_{j+2})/(w_{j+m+1} - w_{j+1})$, and $\mathbf{0}$ is the zero matrix.

3.2. *Asymptotic property*

We establish the consistency property of the sieve estimators. Let $H(\boldsymbol{\rho}, \boldsymbol{\psi})$ denote the conditional entropy between a set of operators $\boldsymbol{\rho} = (\rho^{aa}, \rho^{ab}, \rho^{ba}, \rho^{bb})$ and a given set of operators $\boldsymbol{\psi}$:

$$H(\boldsymbol{\rho}, \boldsymbol{\psi}) = \mathsf{E}_{\boldsymbol{\rho}}\left[\log g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \boldsymbol{\psi}) \right].$$

THEOREM 3.1 *Assume $\{\Theta_{J_n}\}$ is chosen such that conditions **Con**1 and **Con**2 in Appendix 3 are in force. Suppose that for each $\delta > 0$, we can find subsets $\Gamma_1, \Gamma_2, \ldots, \Gamma_{l_{J_n}}$ of Θ_{J_n}, $J_n = 1, 2, \ldots$ such that*

(i) $D_{J_n} \subseteq \bigcup_{k=1}^{l_{J_n}} \Gamma_k$, *where* $D_{J_n} = \{\boldsymbol{\rho} \in \Theta_{J_n} | H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}) \le H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}_{J_n}) - \delta\}$ *for every $\delta > 0$ and every J_n.*

(ii) $\sum_{n=1}^{+\infty} l_{J_n} (\varphi_{J_n})^n < +\infty$, *where given l sets $\Gamma_1, \ldots, \Gamma_l$ in Θ_{J_n}, $\varphi_{J_n} = \sup_k \inf_{t \ge 0} \mathbf{E}_{\boldsymbol{\rho}_{0|\Theta_{J_n}}} \exp\{t \log(g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \Gamma_k)/g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \boldsymbol{\rho}_{J_n}))\}$.*

Then we have $\sup_{\widehat{\boldsymbol{\rho}}_n \in M^n_{J_n}} \| \widehat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_{0|\Theta_{J_n}} \|_{\mathscr{S}} \to 0$ a.s.

The norm $\| \cdot \|_{\mathscr{S}}$ is a Hilbert–Schmidt norm for the convolution kernel operator and its Hilbert–Schmidt norm is $\|\rho\|_{\mathscr{S}} = (\sum_j \lambda_j^2)^{1/2}$. The use of Hilbert–Schmidt norm comes from the fact that it forms a class of operators embedded in the whole space of Hilbert–Schmidt operators and for any convolution kernel operator ρ, the Hilbert–Schmidt norm of ρ is equal to the L^2 norm of its kernel function, in particular, $\|\rho\|_{\mathscr{S}} = \|\kappa\|_2$.

Note that in Theorem 3.1, $g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \Gamma_k) = \sup_{\boldsymbol{\psi} \in \Gamma_k} g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \boldsymbol{\psi})$. We define the set of all the MLEs on Θ_{J_n} given the sample size n as $M^n_{J_n} = \{\boldsymbol{\rho} \in \Theta_{J_n} | \ell(X_1^{(a)}, \ldots, X_n^{(a)}, X_1^{(b)}, \ldots, X_n^{(b)}; \boldsymbol{\rho}) = \sup_{\boldsymbol{\psi} \in \Theta_{J_n}} \ell(X_1^{(a)}, \ldots, X_n^{(a)}, X_1^{(b)}, \ldots, X_n^{(b)}; \boldsymbol{\psi})\}$. Let $\boldsymbol{\rho}_0$ denotes the true set of values for the set of parameters $(\rho_0^{aa}, \rho_0^{ab}, \rho_0^{ba}, \rho_0^{bb})$. We follow Mourid and Bensmain (2006) for the proof of Theorem 3.1 to show the convergence of the ML estimator to $\boldsymbol{\rho}_{0|\Theta_{J_n}}$, the projections of the true operators on sieve, see Appendix 3 for details. Together with the convergence of $\boldsymbol{\rho}_{0|\Theta_{J_n}}$ to the true set of operators $\boldsymbol{\rho}_0$ as the sieve dimension grows, we prove that the ML estimator converges to the true set of operators $\boldsymbol{\rho}_0$.

THEOREM 3.2 *If $J_n = \mathcal{O}(n^{1/3-\eta})$ for $\eta > 0$, then $\| \widehat{\boldsymbol{\kappa}}_{J_n} - \boldsymbol{\kappa}_{0|\Theta_{J_n}} \|_2 \to 0$ a.s. when $n \to +\infty$. $\widehat{\boldsymbol{\kappa}}_{J_n} = (\widehat{\kappa}_{aa,J_n}, \widehat{\kappa}_{ab,J_n}, \widehat{\kappa}_{ba,J_n}, \widehat{\kappa}_{bb,J_n})$ is the set of sieve estimators on Θ_{J_n} and $\boldsymbol{\kappa}_{0|\Theta_{J_n}} = (\kappa_{aa,0|\Theta_{J_n}}, \kappa_{ab,0|\Theta_{J_n}}, \kappa_{ba,0|\Theta_{J_n}}, \kappa_{bb,0|\Theta_{J_n}})$ is the projection of the set of true kernel functions $\boldsymbol{\kappa}_0$ on Θ_{J_n}. $\| \widehat{\boldsymbol{\kappa}}_{J_n} - \boldsymbol{\kappa}_{0|\Theta_{J_n}} \|_2 \to$*

Table 2. R^2, RMSE, and MAPE for in-sample estimation of the 12 stocks.

	VFAR			RW vs. VFAR		
Ticker symbol	R^2 (%)	RMSE	MAPE (%)	R^2	RMSE	MAPE
AAPL	92.03	0.34	3.61	**0.97**	**1.18**	**1.05**
MSFT	95.19	0.18	0.95	**0.98**	**1.16**	**1.07**
INTC	94.79	0.19	0.92	**0.98**	**1.15**	**1.07**
CSCO	96.16	0.19	0.86	**0.99**	**1.13**	**1.06**
SIRI	98.29	0.09	0.29	**1.00**	**1.09**	**1.00**
AMAT	95.83	0.18	0.89	**0.99**	**1.15**	**1.09**
CMCSA	93.39	0.19	1.20	**0.97**	**1.18**	**1.13**
AEZS	98.48	0.42	2.18	**0.98**	**1.45**	**1.05**
EBAY	94.88	0.23	1.55	**0.98**	**1.15**	**1.06**
MU	95.14	0.26	1.17	**0.98**	**1.16**	**1.08**
WFM	95.52	0.20	1.57	**0.98**	**1.16**	**1.01**
SBUX	94.77	0.22	2.51	**0.98**	**1.17**	**1.05**

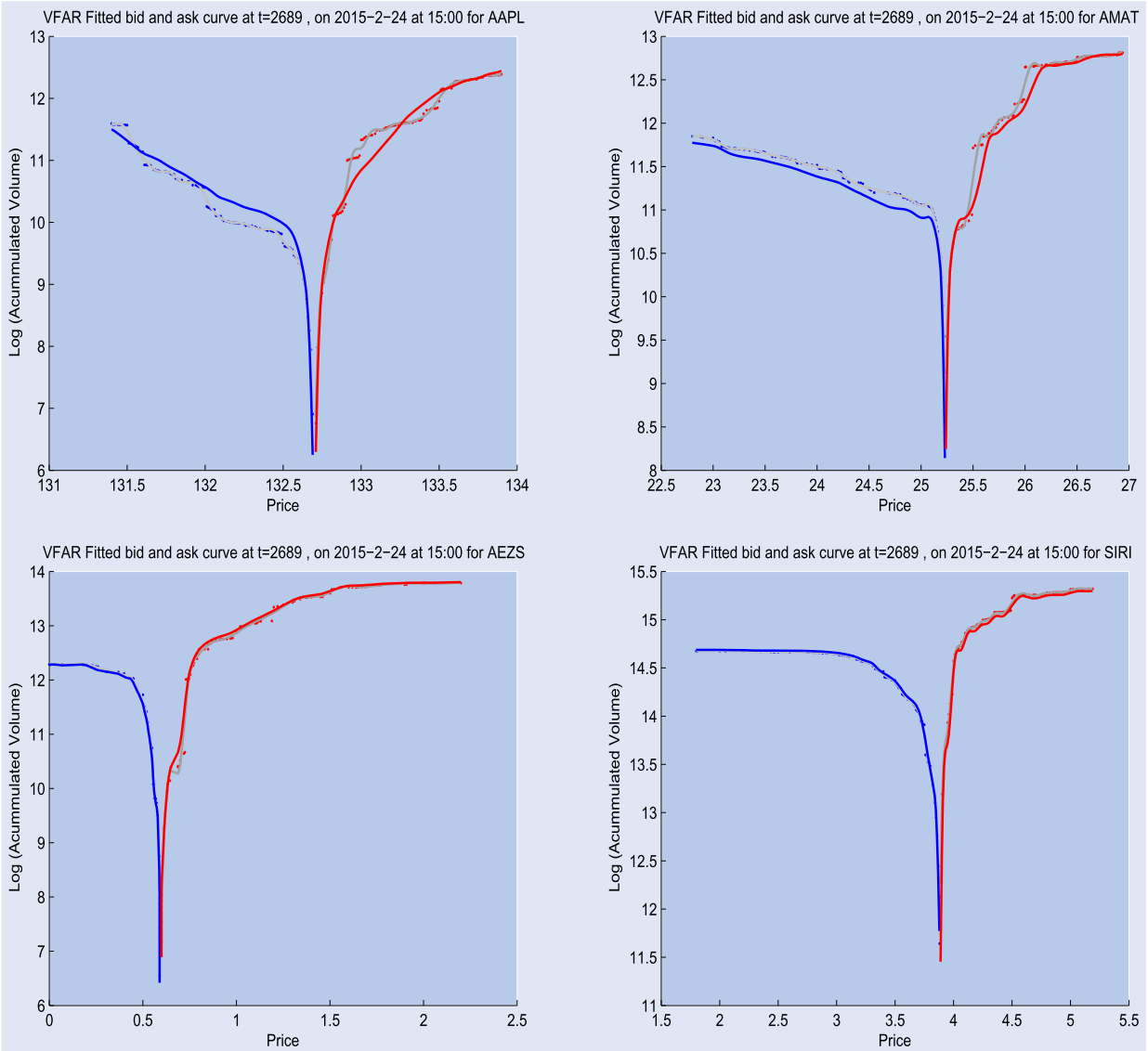

Figure 4. Estimated bid (and ask) supply curves vs. the actually observed.

0 *a.s. means that each* $\|\widehat{\kappa}_{xy,J_n} - \kappa_{xy,0|\Theta_{J_n}}\|_2 \to 0$ *a.s. for* $xy = aa, ab, ba, bb$.

By checking the conditions of Theorem 3.1, we can achieve the proof of Theorem 3.2. The proof is detailed in Appendix 4.

As $n, J_n \to \infty$, we have $\boldsymbol{\kappa}_{0|\Theta_J} \to \boldsymbol{\kappa}_0$ as $\kappa_{xy,0|\Theta_J}$ in $\boldsymbol{\kappa}_{0|\Theta_J}$ is just the B-spline truncation of the corresponding true kernel $\kappa_{xy,0}$ in $\boldsymbol{\kappa}_0$ on Θ_{J_n}. Finally we have the sieve estimator $\widehat{\boldsymbol{\kappa}}_{J_n}$ converges to the true set of kernel functions $\boldsymbol{\kappa}_0$.

4. Modeling liquidity demand and supply curves

We apply the convolutional VFAR model to study the joint dynamics of the liquidity demand and supply curves in the LOB. We investigate the in-sample and out-of-sample predictability based on the records of the 12 stocks with high variations over 44 trading days from date 2 January 2015 to 6 March 2015. We evaluate the accuracy of prediction and also demonstrate the application of the VFAR forecast in order execution strategy.

On each day, the liquidity demand and supply curves are obtained by using the B-spline expansions on the log-accumulated volumes. Throughout the analysis, we assumed prices are known exogenous variables. The equally spaced price percentiles are used as nodes and $J_n = 20$ is chosen in the sieve. The value of J_n is selected for giving on average the highest explanatory power over all the 12 stocks in our analysis. There are in total 20 coefficients on the bid side and another 20 on the ask side.

One may suspect co-integration between the bid and ask sides, though there is no empirical evidence on the existence of co-integration in figure 2, we also consider the Random Walk (RW) model as an alternative, where the liquidity curves are predicted by the most recent curves at the previous time point. The selection of random walk is also motivated by the fact the it provides a general good predictability and is hard to beat under market efficiency.

Three measures are used to evaluate the prediction performance. They are Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) for accuracy, and R^2 for the explanatory power:

$$\begin{aligned}
\mathrm{RMSE} &= \sqrt{\frac{\sum_{xy=a,b}\sum_{t=1}^{n}\sum_{\tau}\left\{X_t^{(xy)}(\tau)-\widehat{X}_t^{(xy)}(\tau)\right\}^2}{\sum_{t=1}^{n}N_t}},\\
\mathrm{MAPE} &= \frac{\sum_{xy=a,b}\sum_{t=1}^{n}\sum_{\tau}\frac{\left|X_t^{(xy)}(\tau)-\widehat{X}_t^{(xy)}(\tau)\right|}{X_t^{(xy)}(\tau)}}{\sum_{t=1}^{n}N_t},\\
R^2 &= 1-\frac{\sum_{xy=a,b}\sum_{t=1}^{n}\sum_{\tau}\left\{X_t^{(xy)}(\tau)-\widehat{X}_t^{(xy)}(\tau)\right\}^2}{\sum_{xy=a,b}\sum_{t=1}^{n}\sum_{\tau}\left\{X_t^{(xy)}(\tau)-\bar{X}\right\}^2},
\end{aligned} \tag{7}$$

where X is the actual value, $\widehat{X}$ denotes the estimate or forecast, and N_t is the total number of the observed price quotes on both sides of LOB at time point t. For each stock, we calculate these measures for the estimated/forecasted liquidity curves using the VFAR model and the alternative RW model, respectively.

4.1. *In-sample estimation*

We conduct the in-sample estimation over the whole time period of the 44 days. Table 2 reports the R^2, RMSE and MAPE of the estimated liquidity curves. It shows that VFAR provides high explanatory power for all the stocks, with R^2 ranging from 92 % (AAPL) to 98 % (AEZS), and superior prediction accuracy with RMSE smaller than 0.42 (AEZS) and MAPE lower than 3.61 % (AAPL). We compare the performance of VFAR and the alternative RW model. On the right panel, the ratio of each measure is computed for the estimates based on the RW model against those on the VFAR model. The best relative performance is marked in bold-face. Without exception, the VFAR model is better than the alternative. In terms of R^2, VFAR outperforms by up to 3 % (AAPL the largest stock and CMCSA the least active stock). As for estimation accuracy, the relative performance reaches to 13 % in MAPE (CMCSA) and at least 9 % (SIRI, the most

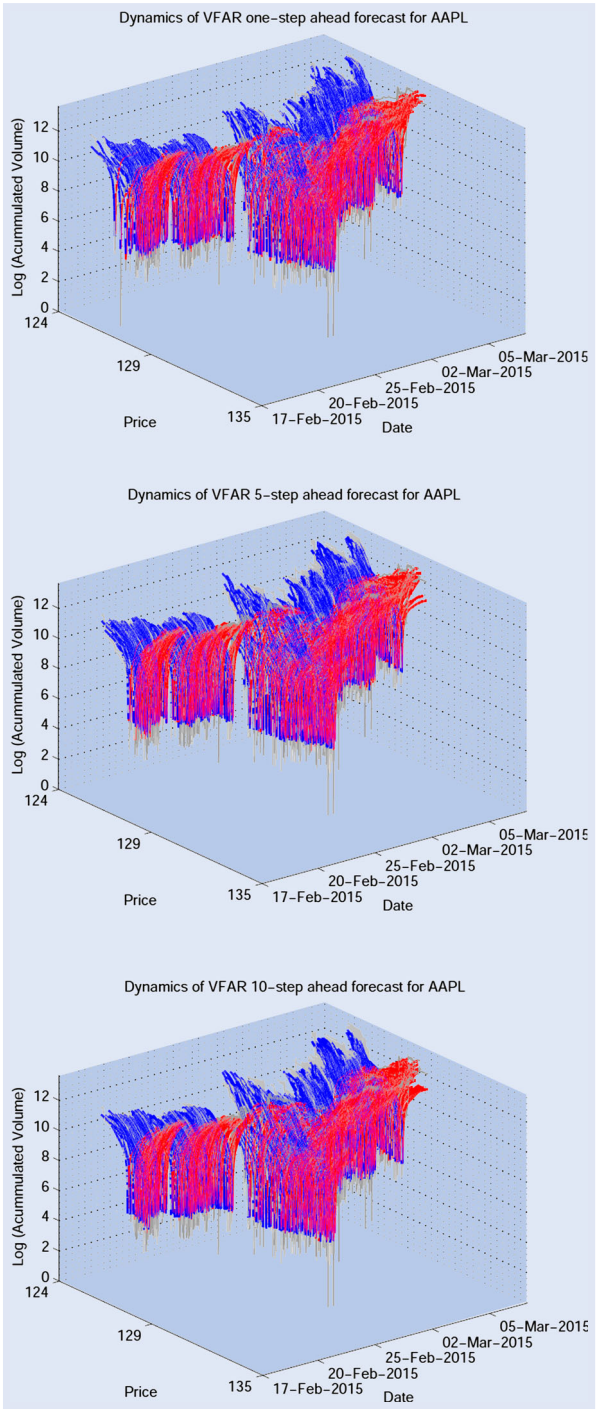

Figure 5. Dynamics of multi-step ahead forecast for AAPL. Top: 5-min ahead forecast; middle: 25-min ahead forecast; bottom: 50-min ahead forecast.

Table 3. R^2, RMSE, MAPE for multi-step ahead VFAR forecast of the 12 stocks.

	R^2 (%)			RMSE			MAPE (%)		
Ticker symbol	1-step	5-steps	10-steps	1-step	5-steps	10-steps	1-step	5-steps	10-steps
AAPL	91.13	85.64	83.74	0.37	0.48	0.51	3.61	4.21	4.49
MSFT	95.38	91.56	89.65	0.18	0.24	0.27	0.93	1.42	1.63
INTC	94.02	89.44	86.95	0.19	0.26	0.28	0.95	1.46	1.74
CSCO	96.67	93.07	90.35	0.21	0.31	0.36	0.88	1.44	1.76
SIRI	98.14	96.23	95.31	0.09	0.13	0.14	0.30	0.52	0.62
AMAT	95.47	92.17	89.99	0.19	0.25	0.29	1.00	1.47	1.72
CMCSA	92.80	89.22	88.09	0.20	0.24	0.25	1.15	1.56	1.69
AEZS	98.23	97.71	97.43	0.48	0.55	0.58	2.22	2.85	3.14
EBAY	94.64	91.54	89.74	0.23	0.29	0.32	1.31	1.79	2.03
MU	95.37	92.35	90.60	0.22	0.28	0.31	1.18	1.70	1.99
WFM	95.18	92.24	91.15	0.20	0.26	0.27	1.25	1.76	1.94
SBUX	94.49	91.82	90.63	0.23	0.28	0.30	1.81	2.27	2.48

Table 4. Ratio of R^2, RMSE, MAPE for multi-step ahead RW forecast to VFAR forecast of the 12 stocks.

	R^2			RMSE			MAPE		
Ticker symbol	1-step	5-steps	10-steps	1-step	5-steps	10-steps	1-step	5-steps	10-steps
AAPL	**0.97**	**0.93**	**0.90**	**1.15**	**1.19**	**1.22**	**1.03**	**1.11**	**1.13**
MSFT	**0.99**	**0.98**	**0.97**	**1.10**	**1.12**	**1.14**	0.97	0.98	**1.02**
INTC	**0.98**	**0.96**	**0.94**	**1.12**	**1.16**	**1.19**	**1.02**	**1.05**	**1.07**
CSCO	**0.99**	**0.98**	**0.97**	**1.09**	**1.13**	**1.14**	**1.01**	**1.01**	**1.05**
SIRI	**1.00**	**0.99**	**0.99**	**1.04**	**1.10**	**1.09**	0.90	0.90	0.90
AMAT	**0.99**	**0.97**	**0.96**	**1.11**	**1.16**	**1.18**	**1.01**	**1.05**	**1.09**
CMCSA	**0.98**	**0.94**	**0.91**	**1.14**	**1.23**	**1.28**	**1.06**	**1.13**	**1.19**
AEZS	**0.98**	**0.98**	**0.98**	**1.36**	**1.40**	**1.39**	**1.05**	**1.12**	**1.12**
EBAY	**0.99**	**0.96**	**0.94**	**1.12**	**1.19**	**1.23**	**1.05**	**1.11**	**1.16**
MU	**0.98**	**0.96**	**0.95**	**1.15**	**1.20**	**1.22**	**1.08**	**1.12**	**1.15**
WFM	**0.99**	**0.96**	**0.94**	**1.13**	**1.23**	**1.29**	**1.05**	**1.12**	**1.20**
SBUX	**0.98**	**0.96**	**0.94**	**1.15**	**1.21**	**1.24**	**1.03**	**1.11**	**1.15**

active stock) and up to 45 % (AEZS that has the smallest bid–ask spread on average) in RMSE. We find the superior performance of the VFAR is robust with respect to market capitalization, market tightness and depth.

Figure 4 visualizes the fitted liquidity demand and supply curves and the actual values at an arbitrarily selected date, 24 February 2015 at 3 pm, of the four representative stocks, AAPL, AMAT, AEZS and SIRI. The estimated curves reasonably trace the queuing orders displayed as discrete dots as well as the smoothed liquidity curves in gray color. The accuracy is quite stable, especially in the middle around the best quotes as well as the extreme in the tails.

4.2. Out-of-sample forecast

In this section, we analyze the model's forecasting performance in a realistic setup. In particular, a trader is assumed to observe the LOB at 5-min snapshots, with the information over the past 30 trading days. The trader can only submit orders every 5 min and thus asks for multi-step ahead out-of-sample forecasts for the liquidity curves, starting from the 31st trading day onwards. Among others, the trader is interested in 1-, 5- and 10-step ahead forecasts that correspond to 5-, 25- and 50-min ahead liquidity curves, respectively. As such, the first pair of the forecasted curves is for time $t = 2251$, based on the past 30 trading days of $30 \times 75 = 2250$ functional objects. Each time, he moves forward one period, i.e. 5 min and performs re-estimation and forecast until reaching the end of the sample at $t = 3300$.

Figure 5 gives graphical illustrations of the forecasted liquidity curves for AAPL with the VFAR model. The forecasts closely trace the realized liquidity curves. It is remarkable that the VFAR model is able to catch the dynamic movements of the liquidity curves over the period from 17 February to 06 March 2015 for different forecast horizon from 5- to 50-min.

Table 3 reports the RMSE, MAPE and predictive power of the liquidity curves forecast for the 12 stocks. Even if in the 'worst' case, the VFAR approach is able to achieve high R^2 ranging from 91.13 % (1-step AAPL) to 83.74 % (10-step AAPL), low RMSE of 0.48 (1-step AEZS) to 0.58 (10-step AEZS), and low MAPE of 3.61 % (1-step AAPL) to 4.49 % (10-step AAPL). The relative performance of the alternative RW model is summarized in table 4. Again, the VFAR model dominates the RW model across forecast horizons and forecast measures. Though the improvement in R^2 is weak, the advantage is obvious in terms of the forecast error reduction. In terms of RMSE, the VFAR model reaches about 4 % (1-step SIRI) in the worst case and 36 % (1-step AEZS) and 40 % (5-step AEZS) in the best case. On the other hand, the VFAR model does not always yield improvement in the MAPE comparison. However, the RW performs better than VFAR only in 5 out of 36 instances. In other cases, VFAR outperforms the RW by up to 20 %. The relative superior performance grows as the forecast horizon increases, indicating that the

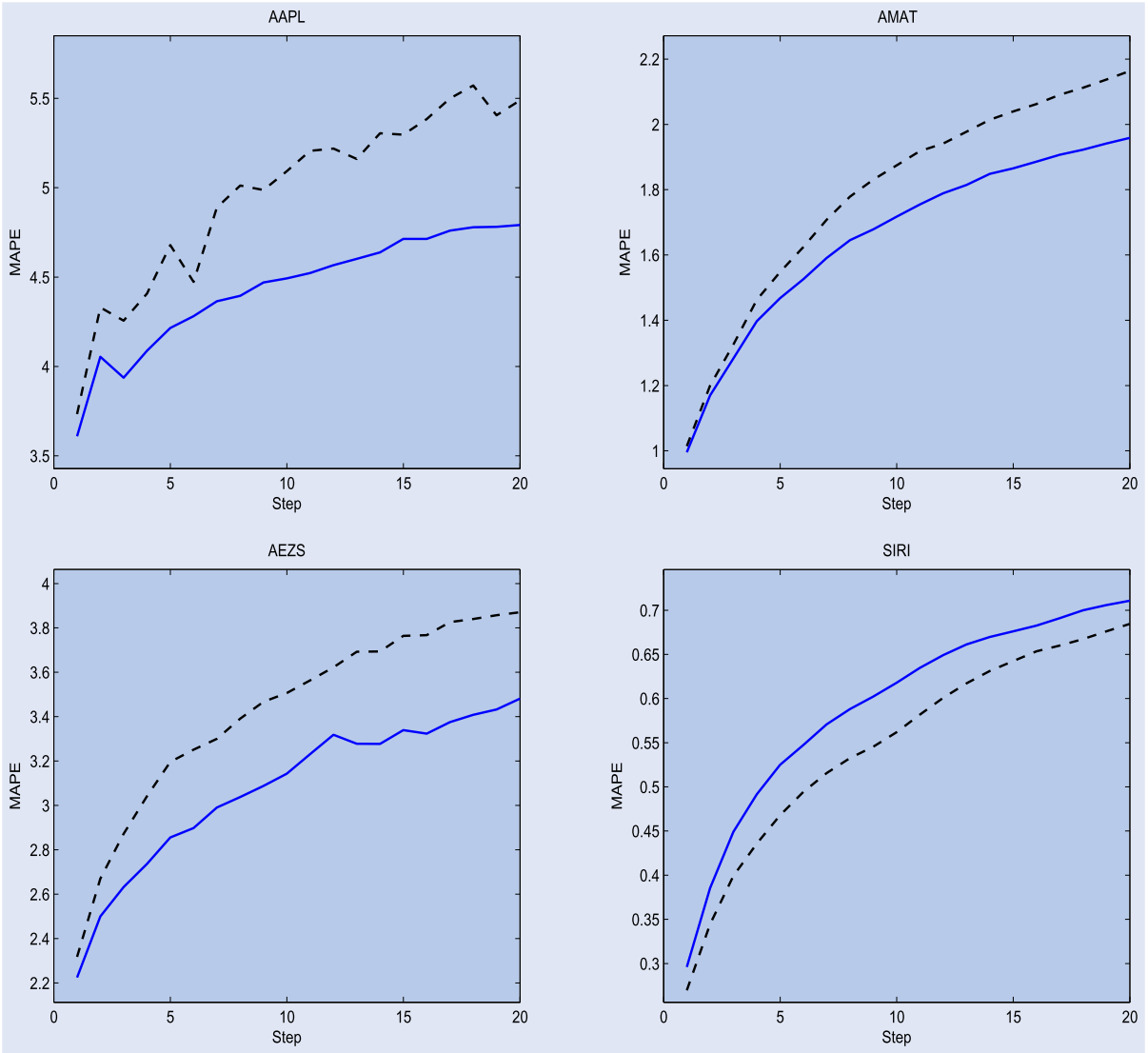

Figure 6. Mean absolute percentage errors (MAPEs) implied by the VFAR approach (blue) and the RW approach (black dashed) for different number of steps ahead forecasts, corresponding to 5–100 min.

utilization of cross-dependence in liquidity curves helps to improve out-of-sample prediction.

We also find that the strong predictability of the VFAR model is stable with respect to forecast horizons. Figure 6 displays the MAPEs for various multi-steps ahead forecasts ranging from 1 to 20 steps ahead for the four representative stocks AAPL, AMAT, AEZS, and SIRI. Except SIRI, the VFAR forecasts outperform the RW with lower MAPEs. As we forecast further into the future, the advantage of VFAR over RW increases for AAPL, AMAT and AEZS. As the most active stock considered in our data analysis, SIRI has weaker predictive power in the out-of-sample forecasting. It is the only asset for which MAPE is greater than the random walk specification alternative, although this difference shrinks as the steps forward increase. This weaker predictive power could be due to the fact that SIRI is the most active stock, and its dynamics is difficult to characterize, even by the VFAR.

To summarize, the proposed VFAR model is able to successfully predict the liquidity curves over various forecasting periods. These results can be applied to various financial and economics applications, and we will show next an application to order execution strategy as an example.

4.3. Application to order execution strategy

In this section, we show how to utilize the forecasting results in the previous section to a practical application on order execution. Assume that an investor decides to buy (sell) v number of shares over a course of a trading day, starting from 9:45 to 15:55. The volume v to be traded is chosen to be 5 or 10 times the average pending volume at the best bid (ask) price, yielding the following buy (sell) quantities in the respective two cases of (a) high and (b) very high liquidity demand:

(a) AAPL-4000 (4000); MSFT-26 000 (29 000); INTC-27 000 (31 000); CSCO-67 000 (53 000); SIRI-675 000 (654 000); AMAT-17 000 (18 000); CMCSA-7000 (8000); AEZS-12 000 (17 000); EBAY-6000 (6000); MU-8000 (8000); WFM-2000 (3000); SBUX-2000 (2000).

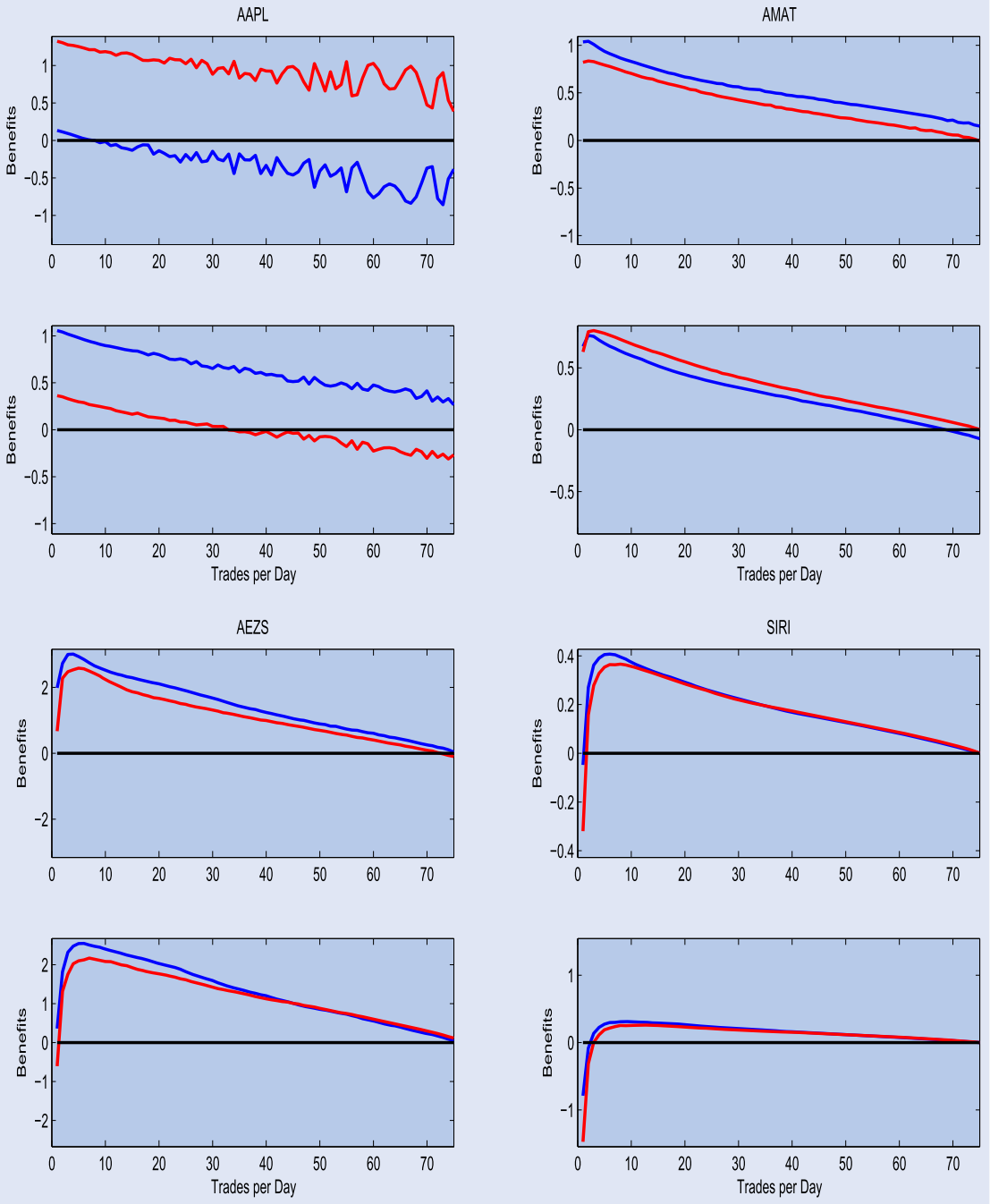

Figure 7. Average percentage gains by strategy (ii) in excess of the equal-splitting strategy (i) when buying (blue) and selling (red) shares based on m VFAR-predicted time points per day. For each stock, top: high liquidity demand corresponding to 5 times the average first level market depth; bottom: very high liquidity demand corresponding to 10 times the average first level market depth.

(b) AAPL-8000 (8000); MSFT-52 000 (58 000); INTC-54 000 (62 000); CSCO-134 000 (106 000); SIRI-1 350 000 (1 308 000); AMAT-34 000 (36 000); CMCSA-14 000 (16 000); AEZS-24 000 (34 000); EBAY-12 000 (12 000); MU-16 000 (16 000); WFM-4000 (6000); SBUX-4000 (4000).

Suppose that trading is only performed on a 5 min grid throughout the day corresponding to 75 possible trading time points, and investor can only make decision at 9:45 but not monitor the market anymore during the day. The forecasting horizon covers $h = 75$ periods on each trading day. The following are two execution strategies the investor can use:

(i) Splitting the buy (sell) order of size v equally in 5-min frequency over the trading day, resulting into 75 trades of size $v/75$ each; see Almgren and Chriss (2000).

(ii) Placing different orders at m (5-min interval) time points throughout the day where the VFAR predicted implied trading costs c of volume v are smallest. The volume is split over the m time points according to the relative proportion of expected trading costs, i.e. at time point i, $w_i \cdot v$ shares are traded, with $w_i = c_i / \sum_{j=1}^{m} c_j$ for $i = 1, \ldots, m$; see Härdle *et al.* (2012).

Note that strategy (i) is a special case of strategy (ii) with $m = 75$ and the volume v is equally split. For strategy (ii),

if $m = 1$, it is the extreme case where the whole quantity is traded only once. The VFAR prediction of the trading costs involves predicting the whole bid and ask curve at each time point and compute the effective cost of trading using prevailing bid and ask quotes assumed to be known at the respective time points.

We implement these strategies to 14 forecasting days, from 17 February to 06 March 2015. Figure 7 shows the average percentage reduction in trading costs of strategy (ii) in excess of the equal-splitting strategy (i) for different values of $m \in [1, 75]$ for AAPL, AMAT, AEZS, and SIRI. Overall, we observe that strategic placement of orders according to VFAR predictions achieve excess gains of 31 basis points on average. Generally, the behavior of the curves is similar, increasing as m increases from 1, and converge to zero as m reaches the upper limit of 75. The pattern shows that making small number of large market orders is superior to an equal-splitting strategy, while for the extreme case where $m = 1$, there is lesser benefit or even loss since the transactions have to walk up the book too severely and cause large price impacts; and where $m = 75$, the relative benefit only results from the strategic non-equal weighting scheme. All in all, the VFAR model is successful in predicting times when the market is sufficiently deep to execute large orders.

5. Conclusion

Predictions of future liquidity supply and demand in the limit order book (LOB) help in analyzing optimal splitting strategies for large orders to reduce cost. To capture not only the volume around the best bid and ask price in the LOB, but also the pending volumes more deeply in the book, it becomes an ultra-high dimensional problem. Motivated by the significant cross-dependency of the bid and ask side of the market, we proposed a Vector Functional AutoRegressive (VFAR) model to estimate and forecast the liquidity supply and demand curves in the functional domain.

The model is applied to 12 stocks traded in the National Association of Securities Dealers Automated Quotations (NASDAQ) stock market. It is shown that the VFAR model gives R^2 values as high as 98.5 % for in-sample estimation. In out-of-sample forecast experiments, it produces accurate 5-, 25- and 50-min forecasts, with MAPE as low as 0.3 –4.5%. The predictive power of the VFAR model can be further used to improve order execution strategies at lower trading cost.

Our results contribute to the finance domain in helping both practitioners and academics to better understand the dynamics of available liquidity of a LOB and aiding to construct a forward-looking trading strategy. In the area of financial econometrics, we extend the VAR framework to the functional domain, develop the method, and derive the theoretical results. For future work, one can consider adding exogenous variables to the VFAR model, extending the VARX framework in the functional domain, to develop the VFARX model. In application, one can jointly forecast the liquidity supply and demand of a basket of stocks, rather than just an individual stock that was presented in this paper, and include other exogenous variables (in VFARX) such as oil prices, or electricity prices, that may affect the liquidity of certain stocks. Most important, the VFAR model is general, has high interpretability, and can be used for other multiple functional time series modeling and forecasting.

Acknowledgments

We would like to thank the editor and two anonymous referees for their constructive comments to help improve the quality of this paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The research of Ying Chen is supported by the Academic Research Funding R-155-000-178-114 and IDS Funding R-155-000-185-64 at the National University of Singapore. Support from IRTG 1792 ‘High Dimensional Non Stationary Time Series’, Humboldt-Universität zu Berlin, is gratefully acknowledged.

References

Aït-Sahalia, Y., Mykland, P. and Zhang, L., How often to sample a continuous-time process in the presence of market microstructure noise. *Rev. Financ. Stud.*, 2005, **18**, 351–416.

Almgren, R. and Chriss, N., Optimal execution of portfolio transactions. *J. Risk.*, 2000, **3**, 5–39.

Antoniadis, A. and Sapatinas, T., Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *J. Multivar. Anal.*, 2003, **87**, 135–158.

Benston, G. and Hagerman, R., Determinants of bid-asked spreads in the over-the-counter market. *J. Financ. Econ.*, 1974, **1**, 353–364.

Besse, P., Cardot, H. and Stephenson, D., Autoregressive forecasting of some functional climatic variations. *Scand. J. Stat.*, 2000, **27**, 673–687.

Bosq, D., *Linear Processes in Function Spaces: Theory and Applications*, 2000 (Springer: New York).

Chaudhuri, K., Kim, M. and Shin, Y., Forecasting distributions of inflation rates: A functional Autoregressive Approach. *J. R. Stat. Soc.: Ser. A (Stat. Soc.)*, 2016, **179**, 65–102.

Chen, Y. and Li, B., An adaptive functional autoregressive forecast model to predict electricity price curves. *J. Bus. Econ. Stat.*, 2017, **35**(3), 371–388.

Chordia, T., Sarkar, A. and Subrahmanyam, A., An empirical analysis of stock and bond market liquidity. *Rev. Financial Stud.*, 2005, **18**(1), 85–129.

Cooper, K., Groth, J. and Avera, W., Liquidity, exchange listing, and common stock performance. *J. Econ. Bus.*, 1985, **37**, 19–33.

Fleming, M. and Remolona, E., Price formation and liquidity in the U.S. treasury market: The response to public information. *J. Financ.*, 1999, **54**, 1901–1915.

Foucault, T., Kadan, O. and Kandel, E., Limit order book as a market for liquidity. *Rev. Financ. Stud.*, 2005, **18**, 1171–1217.

Geman, S. and Hwang, C.R., Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Stat.*, 1982, **10**, 401–414.
Gomber, P., Schweickert, U. and Theissen, E., Liquidity dynamics in an electronic open limit order book: An event study approach. *Eur. Financ. Manage.*, 2015, **21**, 52–78.
Grenander, U., *Abstract Inference*, 1981 (Wiley: New York).
Groß-Klußmann, A. and Hautsch, N., Predicting bid ask spreads using long-memory autoregressive conditional Poisson models. *J. Forecast.*, 2013, **32**, 724–742.
Guillas, S., Rates of convergence of autocorrelation estimates for autoregressive Hilbertian processes. *Stat. Probab. Lett.*, 2001, **55**, 281–291.
Härdle, W., Hautsch, N. and Mihoci, A., Modelling and forecasting liquidity supply using semiparametric factor dynamics. *J. Empir. Financ.*, 2012, **19**, 610–625.
Härdle, W., Hautsch, N. and Mihoci, A., Local adaptive multiplicative error models for high-frequency forecasts. *J. Appl. Economet.*, 2015, **30**, 529–550.
Härdle, W. K., Chen, S., Liang, C. and Schienle, M., Time-varying limit order book networks. IRTG 1792 Discussion Paper 2018-016, IRTG 1792, Humboldt Universität zu Berlin, Germany, 2018.
Harris, L., Liquidity, trading rules and electronic trading systems. Technical report, NYU Salomon Center Series in Finance and Economics, 1990.
Huberman, G. and Halka, D., Systematic liquidity. *J. Financ. Res.*, 2001, **24**, 161–178.
Hwang, C., Gaussian measure of large balls in a Hilbert space. *Proc. Am. Math. Soc.*, 1980, **78**, 107–110.
Johansen, S., Estimating and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 1991, **59**, 1551–1580.
Mourid, T. and Bensmain, N., Sieves estimator of the operator of a functional autoregressive process. *Stat. Probab. Lett.*, 2006, **76**, 93–108.
Stoll, H., The pricing of security dealer services: An empirical study of NASDAQ stocks. *J. Financ.*, 1978, **33**, 1153–1172.
Zhang, L., Mykland, P. and Aït-Sahalia, Y., A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Am. Stat. Assoc.*, 2005, **100**, 1394–1411.

Appendix 1. Derivation of the *B*-spline coefficient relationship as shown in Section 3

$$
\begin{aligned}
X_t^{(a)}(\tau) &= \sum_{h=1}^{\infty} p_h^a B_{h,m}(\tau) \\
&\quad + \int_0^1 \left\{ \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} c_j^{aa} d_{t-1,i}^a B_{j,m}(\tau - s) B_{i,m}(s) \right\} \mathrm{d}s \\
&\quad + \int_0^1 \left\{ \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} c_j^{ab} d_{t-1,i}^b B_{j,m}(\tau - s) B_{i,m}(s) \right\} \mathrm{d}s \\
&\quad + \sum_{j=1}^{\infty} d_j^a(\varepsilon_t^{(a)}) B_{j,m}(\tau) \\
&= \sum_{h=1}^{\infty} p_h^a B_{h,m}(\tau) + \sum_{j=1}^{\infty} d_j^a(\varepsilon_t^{(a)}) B_{j,m}(\tau) \\
&\quad + \sum_{h=1}^{\infty}\sum_{i=1}^{\infty} \\
&\quad \times \left\{ \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{aa} - c_h^{aa} \right\} \\
&\quad \times \frac{w_{i+m} - w_i}{m} d_{t-1,i}^a B_{h,m}(\tau) \\
&\quad + \sum_{h=1}^{\infty}\sum_{i=1}^{\infty} \\
&\quad \times \left\{ \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{ab} - c_h^{ab} \right\} \\
&\quad \times \frac{w_{i+m} - w_i}{m} d_{t-1,i}^b B_{h,m}(\tau), \\
X_t^{(b)}(\tau) &= \sum_{h=1}^{\infty} p_h^b B_{h,m}(\tau) \\
&\quad + \int_0^1 \left\{ \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} c_j^{bb} d_{t-1,i}^b B_{j,m}(\tau - s) B_{i,m}(s) \right\} \mathrm{d}s \\
&\quad + \int_0^1 \left\{ \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} c_j^{ba} d_{t-1,i}^a B_{j,m}(\tau - s) B_{i,m}(s) \right\} \mathrm{d}s \\
&\quad + \sum_{j=1}^{\infty} d_j^b(\varepsilon_t^{(b)}) B_{j,m}(\tau) \\
&= \sum_{h=1}^{\infty} p_h^b B_{h,m}(\tau) + \sum_{j=1}^{\infty} d_j^b(\varepsilon_t^{(b)}) B_{j,m}(\tau) \\
&\quad + \sum_{h=1}^{\infty}\sum_{i=1}^{\infty} \\
&\quad \times \left\{ \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{bb} - c_h^{bb} \right\} \\
&\quad \times \frac{w_{i+m} - w_i}{m} d_{t-1,i}^b B_{h,m}(\tau) \\
&\quad + \sum_{h=1}^{\infty}\sum_{i=1}^{\infty} \\
&\quad \times \left\{ \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{ba} - c_h^{ba} \right\} \\
&\quad \times \frac{w_{i+m} - w_i}{m} d_{t-1,i}^a B_{h,m}(\tau). \qquad \text{(A1)}
\end{aligned}
$$

Rearranging the above equations gives the relationship of the *B*-spline coefficients in (3).

Next we show how the expansion was obtained in (A1). We only show for the first integral in (A1) as the expansion other integrals can be obtained similarly.

$$
\begin{aligned}
&\int_0^1 \left\{ \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} c_j^{aa} d_{t-1,i}^a B_{j,m}(\tau - s) B_{i,m}(s) \right\} \mathrm{d}s \\
&= \sum_{j=1}^{\infty} c_j^{aa} \int_0^1 B_{j,m}(\tau - s) \left\{ \sum_{i=1}^{\infty} d_{t-1,i}^a B_{i,m}(s) \right\} \mathrm{d}s \\
&= \sum_{j=1}^{\infty} c_j^{aa} \left\{ \frac{1}{m} \sum_{i=1}^{\infty} d_{t-1,i}^a (w_{i+m} - w_i) B_{j,m}(\tau - s) \right\} \Bigg|_{s=0}^{s=1}
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{j=1}^{\infty} c_j^{aa} \int_0^1 \left[\frac{1}{m} \sum_{i=1}^{\infty} d_{t-1,i}^a (w_{i+m} - w_i) \right. \\
&\left\{ \frac{m}{w_{j+m} - w_j} B_{j,m-1}(\tau - s) \right. \\
&\left. \left. - \frac{m}{w_{j+m+1} - w_{j+1}} B_{j+1,m-1}(\tau - s) \right\} \right] \mathrm{d}s \\
&= -\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \frac{d_{t-1,i}^a (w_{i+m} - w_i)}{m} c_j^{aa} B_{j,m}(\tau) \\
&\quad + \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \frac{d_{t-1,i}^a (w_{i+m} - w_i)}{m} c_j^{aa} \int_{\tau-1}^{\tau} \left\{ \frac{m}{w_{j+m} - w_j} B_{j,m-1}(z) \right. \\
&\quad \left. - \frac{m}{w_{j+m+1} - w_{j+1}} B_{j+1,m-1}(z) \right\} \mathrm{d}z \\
&= -\sum_{h=1}^{\infty} \sum_{i=1}^{\infty} \frac{d_{t-1,i}^a (w_{i+m} - w_i)}{m} c_h^{aa} B_{h,m}(\tau) \\
&\quad + \sum_{h=1}^{\infty} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) \\
&\times c_j^{aa} \frac{w_{i+m} - w_i}{m} d_{t-1,i}^a B_{h,m}(\tau) \\
&= \sum_{h=1}^{\infty} \sum_{i=1}^{\infty} \left[\sum_{j=1}^{\infty} \left\{ \frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right\} c_j^{aa} - c_h^{aa} \right] \\
&\times \frac{w_{i+m} - w_i}{m} d_{t-1,i}^a B_{h,m}(\tau).
\end{aligned}
$$

The second equality made use of integration by parts, with $(\mathrm{d}/\mathrm{d}s) B_{j,m}(\tau - s) = -((m/(w_{j+m} - w_j)) B_{j,m-1}(\tau - s) - (m/(w_{j+m+1} - w_{j+1})) B_{j+1,m-1}(\tau - s))$ and $\int \sum_{i=1}^{\infty} d_{t-1,i}^a B_{i,m}(s)\, \mathrm{d}s = (1/m) \sum_{i=1}^{\infty} d_{t-1,i}^a (w_{i+m} - w_i)$. In the third equality, we made the substitution of $z = \tau - s$. For the fourth equality, we made use of the formula: $\int_{-\infty}^{\tau} B_{j,m}(z)\, \mathrm{d}z = ((w_{j+m+1} - w_{j+1})/(m+1)) \sum_{h=1}^{\infty} B_{h,m+1}(\tau)$, and truncating the sum up till the Jth term. We also swapped the notation j for the first summation with h in the fourth equality.

Appendix 2. Derivation of the ML estimator in (eqn6)

For $t = 1, \ldots, n$, we write (5) compactly as the following:

$$
Y = BZ + U. \tag{A2}
$$

By applying vec operator to (A2) yields

$$
\begin{aligned}
\mathrm{vec}(Y) &= \mathrm{vec}(BZ) + \mathrm{vec}(U) \\
&= (Z^\top \otimes I_K)\mathrm{vec}(B) + \mathrm{vec}(U)
\end{aligned}
$$

or equivalently,

$$
\mathbf{y} = (Z^\top \otimes I_K)\boldsymbol{\beta} + \mathbf{u},
$$

where $\otimes$ is the Kronecker product.

Assuming

$$
\mathbf{u} = \mathrm{vec}(U) = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \sim \mathcal{N}(0, I_n \otimes \Sigma_u),
$$

the probabilistic density of $\mathbf{u}$ is

$$
f_{\mathbf{u}}(\mathbf{u}) = \frac{1}{(2\pi)^{Kn/2}} |I_n \otimes \Sigma_u|^{-1/2} \exp\left\{ -\frac{1}{2} \mathbf{u}^\top (I_n \otimes \Sigma_u^{-1}) \mathbf{u} \right\}.
$$

In addition,

$$
\mathbf{u} = \begin{bmatrix} I_K & & & 0 \\ -C & \ddots & & \\ & \ddots & \ddots & \\ 0 & & -C & I_K \end{bmatrix} (\mathbf{y} - \mathbf{v}) + \begin{bmatrix} -C \\ 0 \\ \vdots \\ 0 \end{bmatrix} y_0,
$$

where $\mathbf{v} = (v, \ldots, v)^\top$ is a $(Kn \times 1)$ vector. Consequently, $\partial \mathbf{u} / \partial \mathbf{y}^\top$ is a lower triangular matrix with unit diagonal which has unit determinant. Therefore using $\mathbf{u} = \mathbf{y} - (Z^\top \otimes I_K)\boldsymbol{\beta}$, the transition density is as follows:

$$
\begin{aligned}
&g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \rho^{aa}, \rho^{ab}, \rho^{ba}, \rho^{bb}) = f_{\mathbf{y}}(\mathbf{y}) = \left| \frac{\partial \mathbf{u}}{\partial \mathbf{y}^\top} \right| f_{\mathbf{u}}(\mathbf{u}) \\
&= \frac{1}{(2\pi)^{Kn/2}} |I_n \otimes \Sigma_u|^{-1/2} \\
&\exp\left\{ -\frac{1}{2} \Big(\mathbf{y} - (Z^\top \otimes I_K)\boldsymbol{\beta} \Big)^\top (I_n \otimes \Sigma_u^{-1}) \Big(\mathbf{y} - (Z^\top \otimes I_K)\boldsymbol{\beta} \Big) \right\}.
\end{aligned}
$$

The (approximated) log-likelihood function is

$$
\begin{aligned}
&\ell\Big(X_1^{(a)}, \ldots, X_n^{(a)}, X_1^{(b)}, \ldots, X_n^{(b)}; \rho^{aa}, \rho^{ab}, \rho^{ba}, \rho^{bb} \Big) = \ell(\boldsymbol{\beta}, \Sigma_u) \\
&= -\frac{Kn}{2} \log 2\pi - \frac{n}{2} \log |\Sigma_u| - \frac{1}{2} \Big(\mathbf{y} - (Z^\top \otimes I_K)\boldsymbol{\beta} \Big)^\top \\
&\times (I_n \otimes \Sigma_u^{-1}) \Big(\mathbf{y} - (Z^\top \otimes I_K)\boldsymbol{\beta} \Big) \\
&= -\frac{Kn}{2} \log 2\pi - \frac{n}{2} \log |\Sigma_u| - \frac{1}{2} \sum_{t=1}^{n} \Big(y_t - v - C y_{t-1} \Big)^\top \\
&\times \Sigma_u^{-1} \Big(y_t - v - C y_{t-1} \Big) \\
&= -\frac{Kn}{2} \log 2\pi - \frac{n}{2} \log |\Sigma_u| - \frac{1}{2} \sum_{t=1}^{n} \Big(y_t - C y_{t-1} \Big)^\top \\
&\times \Sigma_u^{-1} \Big(y_t - C y_{t-1} \Big) + v^\top \Sigma_u^{-1} \sum_{t=1}^{n} \Big(y_t - C y_{t-1} \Big) \\
&\quad - \frac{n}{2} v^\top \Sigma_u^{-1} v \\
&= -\frac{Kn}{2} \log 2\pi - \frac{n}{2} \log |\Sigma_u| - \frac{1}{2} \mathrm{tr}\Big((Y - BZ)^\top \Sigma_u^{-1} (Y - BZ) \Big)
\end{aligned}
$$

and the first order partial differentiations are as follows:

$$
\begin{aligned}
\frac{\partial \ell}{\partial \boldsymbol{\beta}} &= (Z \otimes I_K)(I_n \otimes \Sigma_u^{-1}) \Big(\mathbf{y} - (Z^\top \otimes I_K)\boldsymbol{\beta} \Big) \\
&= (Z \otimes \Sigma_u^{-1})\mathbf{y} - (ZZ^\top \otimes \Sigma_u^{-1})\boldsymbol{\beta}, \qquad \text{(A3)} \\
\frac{\partial \ell}{\partial \Sigma_u} &= -\frac{n}{2} \Sigma_u^{-1} + \frac{1}{2} \Sigma_u^{-1} (Y - BZ)(Y - BZ)^\top \Sigma_u^{-1}.
\end{aligned}
$$

By equating the first order partial derivatives in (A3) to zero, we obtain the maximum likelihood estimators in (6).

Appendix 3. Proof of Theorem 3.1

The growth of J_n is determined by the following two conditions:

Con1: If there exists a sequence $\{\boldsymbol{\rho}_{J_n}\}$ such that $\boldsymbol{\rho}_{J_n} \in \Theta_{J_n} \forall n$ and $H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}_{J_n}) \to H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}_{0|\Theta_{J_n}})$, then $\|\boldsymbol{\rho}_{J_n} - \boldsymbol{\rho}_{0|\Theta_{J_n}}\|_{\mathscr{S}} \to 0$; meaning each $\|\rho_{J_n}^{xy} - \rho_{0|\Theta_{J_n}}^{xy}\|_{\mathscr{S}} \to 0$, for $xy = aa, ab, ba, bb$. Here $\boldsymbol{\rho}_{0|\Theta_{J_n}}$ denotes the projection of the set of true operators $\boldsymbol{\rho}_0$ on the sieve Θ_{J_n}.

Con2: There exists a sequence $\{\boldsymbol{\rho}_{J_n}\}$ described in **Con1** such that $H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}_{J_n}) \to H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}_{0|\Theta_{J_n}})$.

Fix $\delta > 0$. Following Mourid and Bensmain (2006), we only need to show that

$$P(D_{J_n} \cap M_{J_n}^n \neq \emptyset) = 0, \tag{A4}$$

because if (A4) holds, then with probability 1

$$\inf_{\varphi \in M_{J_n}^n} H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}) \geq H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}_{J_n}) - \delta,$$

for all n sufficiently large. Since δ is arbitrary, and

$$H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}_{J_n}) \to H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}_{0|\Theta_{J_n}}),$$

by condition **Con2** we deduce

$$\liminf \inf_{\boldsymbol{\rho} \in M_{J_n}^n} H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}) \geq H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}_{0|\Theta_{J_n}}) \quad \text{a.s.}$$

Combining with

$$H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}) \leq H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}_{0|\Theta_{J_n}}),$$

we have

$$\lim_{n \to +\infty} \sup_{\boldsymbol{\rho} \in M_{J_n}^n} |H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}) - H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}_{0|\Theta_{J_n}})| = 0 \quad \text{a.s.} \tag{A5}$$

Fix $\varepsilon > 0$, and for each n choose $\boldsymbol{\psi}_n \in M_{J_n}^n$ such that

$$\frac{d(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\psi}_n)}{1 + d(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\psi}_n)} > \sup_{\boldsymbol{\rho} \in M_{J_n}^n} \frac{d(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}))}{1 + d(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}))} - \varepsilon.$$

Condition **Con1** combined with (A5) imply that

$$d(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\psi}_n) \to 0 \quad \text{a.s.}$$

Hence,

$$\limsup \sup_{\boldsymbol{\rho} \in M_{J_n}^n} \frac{d(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}))}{1 + d(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}))} \leq \varepsilon.$$

Since ε is arbitrary, we deduce that $M_{J_n}^n \to \boldsymbol{\rho}_{0|\Theta_{J_n}}$, which is the desired result. Therefore, it suffices to prove (A4).

For now, n and J_n are fixed. Then

$$\begin{aligned}
&(D_{J_n} \cap M_{J_n}^n \neq \emptyset) \\
&\quad \subseteq \left\{ \sup_{\boldsymbol{\rho} \in D_{J_n}} \ell(X_1^{(a)}, \ldots, X_n^{(a)}, X_1^{(b)}, \ldots, X_n^{(b)}; \boldsymbol{\rho}) \right. \\
&\quad \left. \geq \ell(X_1^{(a)}, \ldots, X_n^{(a)}, X_1^{(b)}, \ldots, X_n^{(b)}; \boldsymbol{\rho}_{J_n}) \right\} \\
&\quad \subseteq \bigcup_{k=1}^{l_{J_n}} \left\{ \sup_{\boldsymbol{\rho} \in \Gamma_k} \prod_{i=1}^n g\left(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \boldsymbol{\rho}\right) \right. \\
&\quad \left. \leq \prod_{i=1}^n g\left(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \boldsymbol{\rho}_{J_n}\right) \right\} \\
&\quad \subseteq \bigcup_{k=1}^{l_{J_n}} \left\{ \prod_{i=1}^n g\left(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \Gamma_k\right) \right. \\
&\quad \left. \leq \prod_{i=1}^n g\left(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \boldsymbol{\rho}_{J_n}\right) \right\}.
\end{aligned}$$

Next we bound the probability of this latter set and called it π.

$$\begin{aligned}
\pi &\leq \sum_{k=1}^{l_{J_n}} P\left[\prod_{i=1}^n g\left(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \Gamma_k\right) \right. \\
&\quad \left. \leq \prod_{i=1}^n g\left(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \boldsymbol{\rho}_{J_n}\right) \right] \\
&= \sum_{k=1}^{l_{J_n}} P\left[\exp \sum_{i=1}^n \left\{ t_k \log \frac{g(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \Gamma_k)}{g(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \boldsymbol{\rho}_{J_n})} \right\} \geq 1 \right] \\
&\leq \sum_{k=1}^{l_{J_n}} \mathsf{E}_{\boldsymbol{\rho}_{0|\Theta_{J_n}}} \left[\exp \left\{ t_k \log \frac{g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \Gamma_k)}{g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \boldsymbol{\rho}_{J_n})} \right\} \right]^n
\end{aligned}$$

for any nonnegative arbitrary $t_1, \ldots, t_k$ and conditionally to $X_{i-1}^{(a)}$ and $X_{i-1}^{(b)}$, the laws of the real r.v. $g(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \Gamma_k)$ and $g(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \boldsymbol{\rho}_{J_n})$ are images of g by the translations of the laws ε_i which are i.i.d. Hence, we get

$$\pi \leq l_{J_n} (\varphi_{J_n})^n.$$

Finally, result (A4) is deduced by condition (ii) of Theorem 3.1 and by the Borel–Cantelli lemma.

Appendix 4. Proof of consistency result in Theorem 3.2

Without loss of generality, we assume that p_j^a and p_j^b are all zeros. For non-zero cases, the same consistency results can be obtained. We check the condition **Con1**. We replace J_n by J in the remaining of this section for notational simplicity, and let all summation be from 1 to J. Using the definition of the entropy, we have

$$\begin{aligned}
&H(\boldsymbol{\rho}_{0|\Theta_J}, \boldsymbol{\rho}_{0|\Theta_J}) - H(\boldsymbol{\rho}_{0|\Theta_J}, \boldsymbol{\rho}_{\Theta_J}) \\
&\quad = H(\boldsymbol{\kappa}_{0|\Theta_J}, \boldsymbol{\kappa}_{0|\Theta_J}) - H(\boldsymbol{\kappa}_{0|\Theta_J}, \boldsymbol{\kappa}_{\Theta_J}) \\
&\quad = -\frac{1}{2} \log |\Sigma_u| + \frac{1}{2} \log |\Sigma_{u,J}| + \mathsf{E}\left(-\frac{1}{2} x^\top \Sigma_u^{-1} x + \frac{1}{2} x_J^\top \Sigma_{u,J}^{-1} x_J \right),
\end{aligned}$$

where

$$x = \begin{bmatrix}
d_{t,1}^a - \sum_i \left(\sum_j q_j c_j^{aa} - c_1^{aa} \right) \frac{w_{i+m} - wi}{m} d_{t-1,i}^a \\
\quad - \sum_i \left(\sum_j q_j c_j^{ab} - c_1^{ab} \right) \frac{w_{i+m} - wi}{m} d_{t-1,i}^b \\
\vdots \\
d_{t,J}^a - \sum_i \left(\sum_j q_j c_j^{aa} - c_J^{aa} \right) \frac{w_{i+m} - wi}{m} d_{t-1,i}^a \\
\quad - \sum_i \left(\sum_j q_j c_j^{ab} - c_J^{ab} \right) \frac{w_{i+m} - wi}{m} d_{t-1,i}^b \\
d_{t,1}^b - \sum_i \left(\sum_j q_j c_j^{ba} - c_1^{ba} \right) \frac{w_{i+m} - wi}{m} d_{t-1,i}^a \\
\quad - \sum_i \left(\sum_j q_j c_j^{bb} - c_1^{bb} \right) \frac{w_{i+m} - wi}{m} d_{t-1,i}^b \\
\vdots \\
d_{t,J}^b - \sum_i \left(\sum_j q_j c_j^{ba} - c_J^{ba} \right) \frac{w_{i+m} - wi}{m} d_{t-1,i}^a \\
\quad - \sum_i \left(\sum_j q_j c_j^{bb} - c_J^{bb} \right) \frac{w_{i+m} - wi}{m} d_{t-1,i}^b
\end{bmatrix},$$

$$x_J = \begin{bmatrix} d^a_{t,1} - \sum_i \left(\sum_j q_j c^{aa}_{j,J} - c^{aa}_{1,J} \right) \frac{w_{i+m} - wi}{m} d^a_{t-1,i} \\ \quad - \sum_i \left(\sum_j q_j c^{ab}_{j,J} - c^{ab}_{1,J} \right) \frac{w_{i+m} - wi}{m} d^b_{t-1,i} \\ \vdots \\ d^a_{t,J} - \sum_i \left(\sum_j q_j c^{aa}_{j,J} - c^{aa}_{J,J} \right) \frac{w_{i+m} - wi}{m} d^a_{t-1,i} \\ \quad - \sum_i \left(\sum_j q_j c^{ab}_{j,J} - c^{ab}_{J,J} \right) \frac{w_{i+m} - wi}{m} d^b_{t-1,i} \\ d^b_{t,1} - \sum_i \left(\sum_j q_j c^{ba}_{j,J} - c^{ba}_{1,J} \right) \frac{w_{i+m} - wi}{m} d^a_{t-1,i} \\ \quad - \sum_i \left(\sum_j q_j c^{bb}_{j,J} - c^{bb}_{1,J} \right) \frac{w_{i+m} - wi}{m} d^b_{t-1,i} \\ \vdots \\ d^b_{t,J} - \sum_i \left(\sum_j q_j c^{ba}_{j,J} - c^{ba}_{J,J} \right) \frac{w_{i+m} - wi}{m} d^a_{t-1,i} \\ \quad - \sum_i \left(\sum_j q_j c^{bb}_{j,J} - c^{bb}_{J,J} \right) \frac{w_{i+m} - wi}{m} d^b_{t-1,i} \end{bmatrix}.$$

Here Σ_u, $c^{aa}_j, c^{ab}_j, c^{ba}_j$, and c^{bb}_j denote the covariance matrix and B-spline coefficients for the kernel $\boldsymbol{\kappa}_{0|\Theta_J}$; and $\Sigma_{u,J}$, $c^{aa}_{j,J}, c^{ab}_{j,J}, c^{ba}_{j,J}$, and $c^{bb}_{j,J}$ denote the covariance matrix and B-spline coefficients for the kernel $\boldsymbol{\kappa}_J$. $\boldsymbol{\kappa}_J$ is the set of kernel functions for $\boldsymbol{\rho}_J$ with $\boldsymbol{\rho}_J \in \Theta_J$; and $\boldsymbol{\kappa}_{0|\Theta_J}$ is the projection of the set of true kernel functions $\boldsymbol{\kappa}_0$ on Θ_J. Assuming $\Sigma_u = \Sigma_{u,J}$, we have

$$\begin{aligned} &H(\boldsymbol{\rho}_{0|\Theta_J}, \boldsymbol{\rho}_{0|\Theta_J}) - H(\boldsymbol{\rho}_{0|\Theta_J}, \boldsymbol{\rho}_{\Theta_J}) \\ &\quad = \mathsf{E}\left(-\frac{1}{2} x^\top \Sigma_u^{-1} x + \frac{1}{2} x_J^\top \Sigma_u^{-1} x_J \right) \\ &\quad = \frac{1}{2} \sum_{r,s} (\Sigma_u^{-1})_{r,s} \mathsf{E}\Big\{ (x_J)_r (x_J)_s - (x)_r (x)_s \Big\}, \end{aligned}$$

where $(\Sigma_u^{-1})_{r,s}$ is the rth row, sth column of Σ_u^{-1}, $(x_J)_r$ is the rth element of x_J, and $(x)_r$ is the rth element of x.

Since the only difference between $(x_J)_r(x_J)_s$ and $(x)_r(x)_s$ are the different B-spline coefficients, we can group the individual terms of the expansion of $(x_J)_r(x_J)_s$ and the expansion $(x)_r(x)_s$ together. After canceling out the common terms not containing the B-spline coefficients, each of the grouped terms will contain a product of some common terms and the subtraction between the B-spline coefficients (of the same index) of the two kernels or the subtraction between the product of B-spline coefficients of one kernel and that of the other kernel (of the same combination of indices). Hence, if $H(\boldsymbol{\kappa}_{0|\Theta_J}, \boldsymbol{\kappa}_{\Theta_J}) \to H(\boldsymbol{\kappa}_{0|\Theta_J}, \boldsymbol{\kappa}_{0|\Theta_J})$ as $n, J \to \infty$, we have $c^{aa}_{j,J} \to c^{aa}_j$, $c^{ab}_{j,J} \to c^{ab}_j$, $c^{ba}_{j,J} \to c^{ba}_j$, $c^{bb}_{j,J} \to c^{bb}_j$ and consequently $\boldsymbol{\rho}_J \to \boldsymbol{\rho}_{0|\Theta_J}$.

For the condition **Con2** and (i) of Theorem 3.1, we follow similar arguments as in Mourid and Bensmain (2006). To verify Theorem 3.1 (ii), we define

$$\varphi(t) = \mathsf{E}_{\boldsymbol{\kappa}_{0|\Theta_J}} \left\{ \exp\left(t \log \frac{g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \Gamma_k)}{g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \boldsymbol{\kappa}_J)} \right) \right\},$$

where $g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \Gamma_k) = \sup_{\boldsymbol{\psi} \in \Gamma_k} g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \boldsymbol{\psi})$. Furthermore, we have $\varphi(0) = 1$ and $\varphi' = \mathsf{E}_{\boldsymbol{\kappa}_{0|\Theta_J}} \log(g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \Gamma_k)/g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \boldsymbol{\kappa}_J))$.

For a fixed $\kappa \in \Gamma_k$, we have

$$\begin{aligned} A &= \mathsf{E}_{\boldsymbol{\kappa}_{0|\Theta_J}} \log g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \Gamma_k) \\ &\quad - \mathsf{E} \log g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \boldsymbol{\kappa}) \\ &= \mathsf{E}_{\boldsymbol{\kappa}_{0|\Theta_J}} \sup_{\boldsymbol{\psi} \in \Gamma_k} \Big\{ \log g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \boldsymbol{\psi}) \\ &\quad - \log g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \boldsymbol{\kappa}) \Big\} \\ &= \mathsf{E}_{\boldsymbol{\kappa}_{0|\Theta_J}} \sup_{\boldsymbol{\psi} \in \Gamma_k} \Big\{ -\frac{1}{2} \log |\Sigma_{u,\boldsymbol{\psi}}| + \frac{1}{2} \log |\Sigma_{u,\boldsymbol{\kappa}}| \\ &\quad - \frac{1}{2} x_{\boldsymbol{\psi}}^\top \Sigma_{u,\boldsymbol{\psi}}^{-1} x_{\boldsymbol{\psi}} + \frac{1}{2} x_{\boldsymbol{\kappa}}^\top \Sigma_{u,\boldsymbol{\kappa}}^{-1} x_{\boldsymbol{\kappa}} \Big\}, \end{aligned}$$

where $x_{\boldsymbol{\psi}}$ and $x_{\boldsymbol{\kappa}}$ have the same form as x_J, with J replaced by $\boldsymbol{\psi}$ and $\boldsymbol{\kappa}$, respectively. $\Sigma_{u,\boldsymbol{\psi}}$, $c^{aa}_{j,\boldsymbol{\psi}}, c^{ab}_{j,\boldsymbol{\psi}}, c^{ba}_{j,\boldsymbol{\psi}}$, and $c^{bb}_{j,\boldsymbol{\psi}}$ denote the covariance matrix and B-spline coefficients for the kernel $\boldsymbol{\psi}$, while $\Sigma_{u,\boldsymbol{\kappa}}$, $c^{aa}_{j,\boldsymbol{\kappa}}, c^{ab}_{j,\boldsymbol{\kappa}}, c^{ba}_{j,\boldsymbol{\kappa}}$, and $c^{bb}_{j,\boldsymbol{\kappa}}$ denote that for the kernel $\boldsymbol{\kappa}$.

Assuming $\Sigma_{u,\boldsymbol{\psi}} = \Sigma_{u,\boldsymbol{\kappa}} = \Sigma_u$, we have

$$A = \mathsf{E}_{\boldsymbol{\kappa}_{0|\Theta_J}} \sup_{\boldsymbol{\psi} \in \Gamma_k} \left\{ \frac{1}{2} \sum_{r,s} (\Sigma_u^{-1})_{r,s} \Big((x_{\boldsymbol{\psi}})_r (x_{\boldsymbol{\psi}})_s - (x_{\boldsymbol{\kappa}})_r (x_{\boldsymbol{\kappa}})_s \Big) \right\},$$

where $(\Sigma_u^{-1})_{r,s}$ is the rth row, sth column of Σ_u^{-1}, $(x_{\boldsymbol{\psi}})_r$ is the rth element of $x_{\boldsymbol{\psi}}$, and $(x_{\boldsymbol{\kappa}})_r$ is the rth element of $x_{\boldsymbol{\kappa}}$.

We follow the similar conditions and arguments in Mourid and Bensmain (2006) and obtain $A \le C_1 / J^{\eta/2}$, where C_1 is a constant. In addition, for $\delta > 0$,

$$\varphi'(0) = H(\boldsymbol{\kappa}_{0|\Theta_J}, \boldsymbol{\kappa}) - H(\boldsymbol{\kappa}_{0|\Theta_J}, \boldsymbol{\kappa}_J) + A \le C_2 J^{-\eta/2} - \delta.$$

Using Taylor expansion and the results from Hwang (1980) such that $\varphi''(t) \le C_3 J^2$, we have $\varphi(1/J^2) \le 1 - \delta / C_4 J^2$, where C_2, C_3, and C_4 are constants. Since $\varphi_J = \sup_k \inf_{t \ge 0} \varphi(t)$, we can deduce that for sufficiently large J, we have

$$l_J (\varphi_J)^n \le C J^{C J^{1+\eta}} \left(1 - \frac{\delta}{C J^2} \right)^n,$$

which is summable if $J = \mathcal{O}(n^{1/3-\delta})$ for $\delta > 0$ (see Hwang 1980). Note that C is a constant. Finally, we can apply Theorem 3.1 to obtain the result that the ML estimator $\widehat{\boldsymbol{\kappa}}$ obtained on Θ_{J_n} converges to the projected true set of kernel functions $\boldsymbol{\kappa}_{0|\Theta_J}$. As $n, J_n \to \infty$, $\boldsymbol{\kappa}_{0|\Theta_J} \to \boldsymbol{\kappa}_0$ because each $\kappa_{xy,0|\Theta_J}$ in $\boldsymbol{\kappa}_{0|\Theta_J}$ is just the B-spline truncation of the corresponding true kernel $\kappa_{xy,0}$ in $\boldsymbol{\kappa}_0$ on Θ_{J_n}.