Forecasting Limit Order Book Liquidity Supply-Demand Curves with Functional AutoRegressive Dynamics

Ying Chen*
Wee Song Chua*²
Wolfgang K. Härdle*²

* National University of Singapore, Singapore
*² Humboldt-Universität zu Berlin, Germany

BERLIN

ECONOMIC RISK

SFB 649

Forecasting Limit Order Book Liquidity Supply-Demand Curves with Functional AutoRegressive Dynamics

Ying Chen[1,3], Wee Song Chua[*1], and Wolfgang Karl Härdle[2]

[1]Department of Statistics and Applied Probability, National University of Singapore, Singapore
[2]Ladislaus von Bortkiewicz Chair of Statistics, C.A.S.E.- Center for Applied Statistics & Economics, Humboldt-Universität zu Berlin, Spandauer Str. 1, 10178 Berlin, Germany; Sim Kee Boon Institute for Financial Economics, Singapore Management University, 50 Stamford Road, 178899 Singapore, Singapore
[3]Risk Management Institute, National University of Singapore, Singapore

Abstract

Limit order book contains comprehensive information of liquidity on bid and ask sides. We propose a Vector Functional AutoRegressive (VFAR) model to describe the dynamics of the limit order book and demand curves and utilize the fitted model to predict the joint evolution of the liquidity demand and supply curves. In the VFAR framework, we derive a closed-form maximum likelihood estimator under sieves and provide the asymptotic consistency of the estimator. In application to limit order book records of 12 stocks in NASDAQ

[*]**Corresponding author**: Wee Song Chua
Email: a0054070@u.nus.edu   Phone: +65-6516 3470   Fax: +65-6872 3919

traded from 2 Jan 2015 to 6 Mar 2015, it shows the VAR model presents a strong predictability in liquidity curves, with R^2 values as high as 98.5 percent for in-sample estimation and 98.2 percent in out-of-sample forecast experiments. It produces accurate $5-, 25-$ and $50-$minute forecasts, with root mean squared error as low as 0.09 to 0.58 and mean absolute percentage error as low as 0.3 to 4.5 percent.

Keywords: Limit order book, Liquidity risk, multiple functional time series
JEL Codes: C13, C32, C53

1   Introduction

Liquidity is a fundamental determinant of market quality. It is important for regulators, market makers and traders to understand the dynamics of liquidity. An imbalance in market liquidity creates challenges not just for market participants but also for the financing structure of the economy in long term. While regulators need to monitor market liquidity to ensure trade transparency and market stability, market participants are motivated to forecast liquidity for e.g. optimal execution strategies on order splitting and submissions.

Liquidity is traditionally measured by some single-valued statistics such as market tightness of bid-ask spread that is computed with the best bid (buy) and ask (sell) prices and market depth based on the volumes at the best quotes or related. As a comparison, Limit Order Book (LOB) contains much more comprehensive information on liquidity, which matches investors' orders on bid and ask sides based on the price-time priority. LOB tells not only the bid-ask spread and the volumes at the best quotes, but also the queuing orders at various sizes and prices.

The information contained in LOB can be well represented by liquidity curves. The liquidity curves display accumulated volumes against quoted prices on both bid and ask sides. Figure 1 gives a graphical illustration, which displays the snapshots of the liquidity curves of two stocks, Sirius XM Holdings Inc. (SIRI) and Comcast Corporation (CMCSA), traded on March 4, 2015 at 14:45. The liquidity curves have V-shape that are monotonically decreasing on the bid side and monotonically increasing on the ask side. In most cases, there is no crossing of the curves and the gap at the center represents the bid-ask spread. The gradient of the liquidity curves reflects the market depth that 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
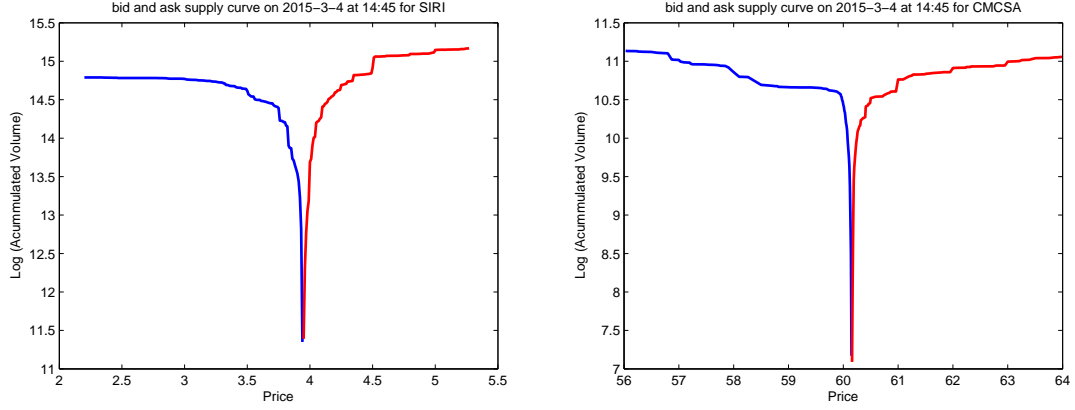
2

Figure 1: SIRI and CMCSA bid and ask supply curve at an arbitrary selected time point. SIRI and CMCSA are the most actively and least actively traded stock in our sample respectively.

VFARrandBidAskCurvePlot

in market. It observes that liquidity is concentrated on relatively few 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. If a trader buys or sells in large volumes at the extreme prices, a drastic change is triggered in the price.

Though with limited information, the single-valued liquidity measures are found to be serially dependent in e.g. Bid-ask spread (e.g. Benston and Hagerman, 1974; Stoll, 1978; Fleming and Remolona, 1999) and Exchange Liquidity Measure (XLM) (see Cooper, Groth and Avera, 1985; Gomber, Schweickert and Theissen, 2015). AutoRegressive models have been employed to describe the dynamics of the liquidity measures. Groß-Klußmann and Hautsch (2013) proposed a long memory AutoRegressive conditional Poisson model for the quoted bid-ask spreads. Huberman and Halka (2001) evidenced the serial dependence of bid-ask spread and depth in the AutoRegressive model. Härdle, Hautsch and Mihoci (2015) proposed a local adaptive multiplicative error model to forecast the high-frequency series of one-minute cumulative trading volumes of several NASDAQ blue chip stocks.

Serial dependence also exists in limit order demand and supply, see Dierker, Kim, Lee and Morck (2014). Chordia, Sarkar and Subrahmanyam (2003) documented the cross-sectional dependence among multiple liquidity measures using a Vector AutoRegressive model for bid-ask spreads, depth, volatility, returns, and order flow in the stock and bond markets, where a liquidity measure not only depends on its own past

3

values, but also those of other measures. Çetin, Jarrow and Protter (2004) introduced liquidity supply curve for robust arbitrage pricing theory. Härdle, Hautsch and Mihoci (2012) studied the de-seasonalized liquidity supply curves in a limit order book market using a dynamic semiparametric factor model.

To understand the dynamics of LOB, it is of high relevance to simultaneously consider the pending quantities deeply queuing on both sides, besides the lead-lag dependence among the single-valued liquidity measures of each curve separately. Public or private information can cause investors to switch from one side to the other, and simultaneously market-wide events can result in similar changes to both bid and ask sides of the limit order book. The joint serial dependence suggests richer dynamics in limit order book and should be utilized in liquidity analysis. In our study, we employ a Vector Functional AutoRegressive (VFAR) model to describe the dynamics of two liquidity curves – demand and supply on bid and ask sides of an electronic open LOB – simultaneously in a unified framework. We derive a closed-form maximum likelihood estimator under sieve and provide asymptotic consistency of the VFAR estimator. The proposed VFAR model is general and can be used for modeling other multiple functional time series.

We investigate the finite sample performance of the proposed forecast model. In the application to the LOB records of 12 stocks traded in NASDAQ from 2 Jan 2015 to 6 Mar 2015, we find the VFAR presents a strong predictability in liquidity curves, with R^2 values as high as 98.5 percent for in-sample estimation and 98.2 percent in out-of-sample forecast experiments. It also produces accurate $5-, 25-$ and $50-$minute forecasts, with root mean squared error as low as 0.09 to 0.58 and mean absolute percentage error as low as 0.3 to 4.5 percent.

This paper is structured as follows. In Section 2, we describe the LOB data. Section 3 presents the VFAR model including estimation and asymptotic property. Section 4 reports the analytical results for both in-sample and out-of-sample in real data analysis. Section 5 provides concluding remarks. All of the theoretical proofs are contained in the Appendix.

2 Data

We consider 12 stocks traded in the National Association of Securities Dealers Automated Quotations (NASDAQ) stock market from 2 Jan 2015 to 6 Mar 2015 (44 trading days). The limit order book (LOB) records were obtained from LOBSTER

through the Research Data Center of the Collaborative Research Center 649 (). NASDAQ is a continuous auction trading platform where the normal continuous trading hours are between 9:30 a.m. to 4:00 p.m. 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 12 stocks 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). These stocks cover a wide range in terms of market capitalization, liquidity tightness and depth. The market value of AAPL is USD737.41 billions the largest compared to USD35.38 millions for the smallest sample stock AEZS. The 5-minute queueing volume in the LOB ranges from 3.73 millions for the most active stock (SIRI) to 0.02 millions for the least active stock (CMCSA) on the bid side and 7.61 millions (SIRI) to 0.03 millions (SBUX) 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.

Ticker Symbol	Mean spread (USD)	Bid vol		Ask vol	
		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

Table 1: Summary statistics on liquidity measures for the 12 stocks traded in NASDAQ. Sampling frequency is 5 minutes.

The LOB records contain the quoted prices and volumes up to 100 price levels on each side. All the quotes are timestamped with decimal precision up to nanoseconds

$(= 10^{-9}$ seconds). 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. For unoccupied price levels, the variables are filled with 9999999999 for ask and -9999999999 for bid, with volumes being 0.

To remove the impact of microstructure noise, the sampling frequency is set to be 5 minutes for a good strike between bias and variance, see Aït-Sahalia, Mykland and Zhang (2005) and Zhang and Aït-Sahalia (2005). The first 15 minutes after opening and the last 5 minutes before closing are discarded to eliminate the market opening and closing effect. Moreover, the accumulated bid and ask volumes are log-transformed when constructing liquidity curves to reduce the impact of extraordinarily large volumes. After the data processing, there are 75 pairs of bid and ask liquidity curves for each stock on each trading day. Over the whole sample period of 44 trading days, it amounts to $3,300$ pairs of bid and ask supply curves for each stock.

The liquidity curves, containing the complete information in LOB, exhibit significant serial dependence. As an illustration, Figure 2 shows the sample cross correlations between the log-accumulated volumes at best bid and ask prices for 6 stocks including AAPL with the largest market value, AEZS with the smallest value and the smallest bid-ask spread on average, CMCSA the least active stock and three well-known MSFT, INTL and EBAY. While the simultaneous dependence between the bid and ask sides is insignificant or negatively correlated, there is positive dependence on the lagged values of the opposite side. Similar features are observed in the other 6 stocks. The bid-ask cross dependency motivates analysing the two liquidity curves jointly.

3   Vector Functional AutoRegressive Model

In this section, we present the Vector Functional AutoRegressive (VFAR) setup that is directly applicable to multiple (e.g. bivariate) continuous curves over time. We show how to estimate the functional parameters, with the help of B-spline expansion and sieve, and provide the asymptotic consistency of the estimator. In functional domain, Bosq (2000) has proposed Functional AutoRegressive (FAR) model for univariate functional time series and developed Yule-Walker estimation (see also Besse, Cardot and Stephenson, 2000; Kim, Chaudhuri and Shin, 2015; Guillas, 2001; Antoniadis
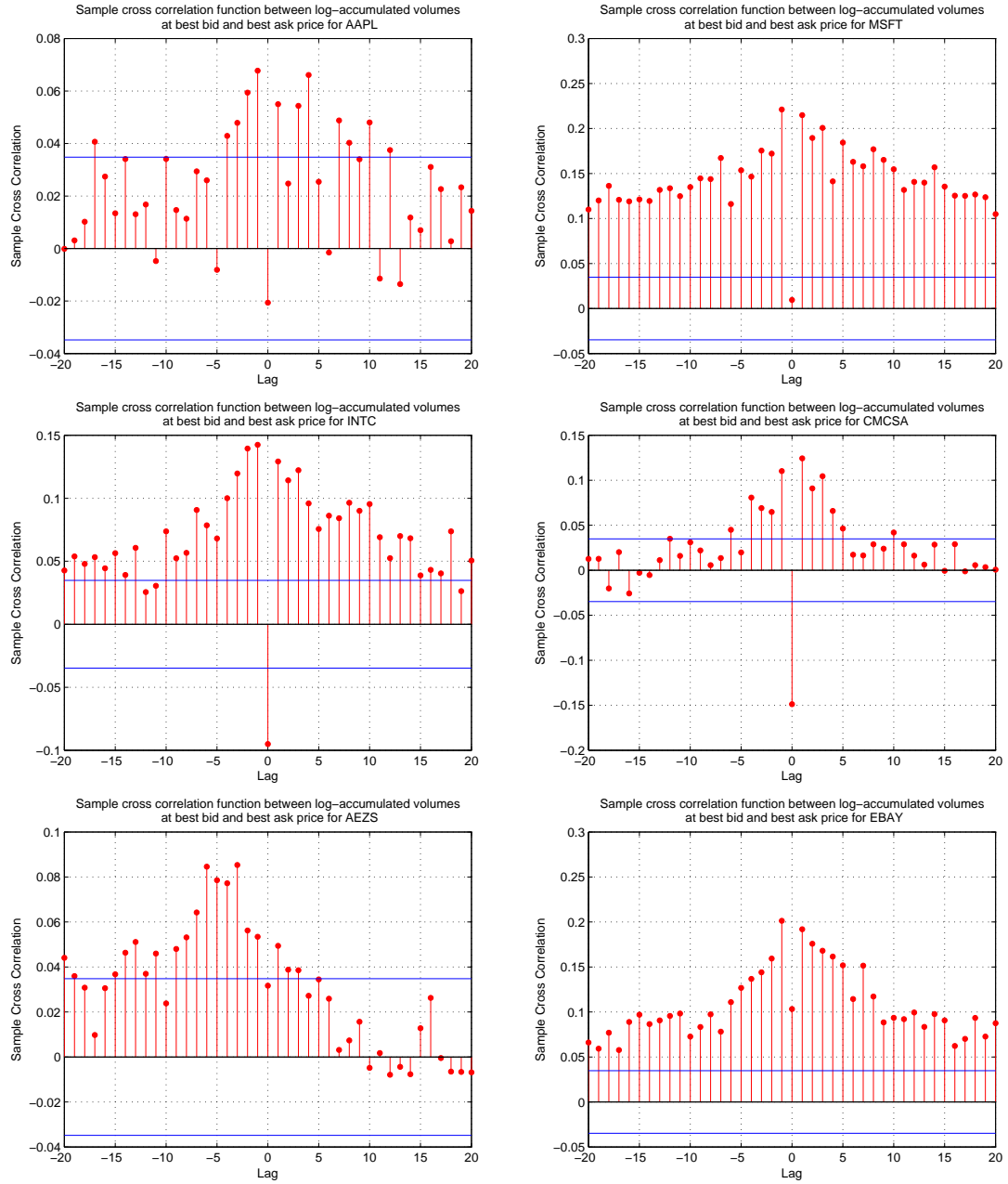
Figure 2: Sample cross correlation function between log-accumulated volumes at best bid and ask price for AAPL, MSFT, INTC, CMCSA, AEZS, and EBAY

VFARcrossCorrPlot

and Sapatinas, 2003; Kokoszka and Zhang, 2010). Mourid and Bensmain (2006) proposed a maximum likelihood estimation with Fourier expansions. Chen and Li (2015) adopted an adaptive approach to extend the applicability of the FAR model in both stationary and non-stationary situations. It is worth noting that the proposed VFAR model is able to analyze multiple functional time series jointly. Furthermore, the maximum likelihood estimator is derived with B-spline expansions that provides more flexibility in fit than e.g. the Fourier expansion.

Our interest is to model the joint dynamic dependence of liquidity curves on the bid and ask sides. Let $X_t^{(a)}(\tau)$ and $X_t^{(b)}(\tau)$ for $\tau \in [0,1]$ be the two processes in the function space $C_{[0,1]}$ of real continuous functions on $[0,1]$. The superscripts (b) and (a) represent bid and ask respectively. Each pair of the liquidity curves can be thought as a data object at time $t = 1, \cdots, n$, and together, they form a time series of n functional objects each on the bid and the ask sides. At each time t, the liquidity curves $X_t^{(b)}$ and $X_t^{(a)}$ are observed containing the quoted prices as well as the corresponding log-accumulated volumes. To handle the two continuous liquidity curves simultaneously, we propose a Vector Functional AutoRegressive (VFAR) model:

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

where $(\mu_a, \mu_b)^\top$ are the mean functions and the operators ρ^{aa}, ρ^{ab}, ρ^{ba}, and ρ^{bb} measure the cross-dependence among the liquidity demand and supply curves on their 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 innovations $\{\varepsilon_t^{(a)}\}_{t=1}^n$ and $\{\varepsilon_t^{(b)}\}_{t=1}^n$ are strong $\mathcal{H}$-white noise, i.i.d. with zero mean and $0 < \mathbb{E}\|\varepsilon_1^{(a)}\|^2 = \cdots = \mathbb{E}\|\varepsilon_n^{(a)}\|^2 < \infty$ and $0 < \mathbb{E}\|\varepsilon_1^{(b)}\|^2 = \cdots = \mathbb{E}\|\varepsilon_n^{(b)}\|^2 < \infty$, 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.

The operators ρ can be represented by a convolution kernel Hilbert-Schmidt operator, which gives

$$X_t^{(a)}(\tau) - \mu_a(\tau) = \int_0^1 \kappa_{ab}(\tau - s)\{X_{t-1}^{(b)}(s) - \mu_b(s)\}ds$$

$$+ \int_0^1 \kappa_{aa}(\tau - s)\{X_{t-1}^{(a)}(s) - \mu_a(s)\}ds + \varepsilon_t^{(a)}(\tau)$$

$$X_t^{(b)}(\tau) - \mu_b(\tau) = \int_0^1 \kappa_{bb}(\tau - s)\{X_{t-1}^{(b)}(s) - \mu_b(s)\}ds$$

$$+ \int_0^1 \kappa_{ba}(\tau - s)\{X_{t-1}^{(a)}(s) - \mu_a(s)\}ds + \varepsilon_t^{(b)}(\tau) \tag{2}$$

where 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 $C_{[0,1]}$.

We expand the functional terms in (2) using B-spline basis function 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

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

We obtain:

$$X_t^{(a)}(\tau) = \sum_{j=1}^{\infty} d_{t,j}^a B_{j,m}(\tau), \quad X_t^{(b)}(\tau) = \sum_{j=1}^{\infty} d_{t,j}^b B_{j,m}(\tau),$$

$$\varepsilon_t^{(a)}(\tau) = \sum_{j=1}^{\infty} d_j^a(\varepsilon_t^{(a)})B_{j,m}(\tau), \quad \varepsilon_t^{(b)}(\tau) = \sum_{j=1}^{\infty} d_j^b(\varepsilon_t^{(b)})B_{j,m}(\tau),$$

$$\kappa_{aa}(\tau) = \sum_{j=1}^{\infty} c_j^{aa} B_{j,m}(\tau), \quad \kappa_{bb}(\tau) = \sum_{j=1}^{\infty} c_j^{bb} B_{j,m}(\tau),$$

$$\kappa_{ab}(\tau) = \sum_{j=1}^{\infty} c_j^{ab} B_{j,m}(\tau), \quad \kappa_{ba}(\tau) = \sum_{j=1}^{\infty} c_j^{ba} B_{j,m}(\tau).$$

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; $d_j^a(\varepsilon_t^{(a)})$ 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.

Plug-in the B-spline expansions to the VFAR model (2), and let p_h^a be the coefficients associated with the expansion of $\mu_a(\tau) - \int_0^1 \kappa_{ab}(\tau - s)\mu_b(s)ds - \int_0^1 \kappa_{aa}(\tau - s)\mu_a(s)ds$ while p_h^b be the coefficients for $\mu_b(\tau) - \int_0^1 \kappa_{bb}(\tau - s)\mu_b(s)ds - \int_0^1 \kappa_{ba}(\tau - s)\mu_a(s)ds$, we have:

$$X_t^{(a)}(\tau) = \sum_{h=1}^{\infty} p_h^a B_{h,m}(\tau) + \int_0^1 \left\{ \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} c_j^{aa} d_{t-1,i}^a B_{j,m}(\tau - s)B_{i,m}(s) \right\}ds$$

$$+ \int_0^1 \left\{ \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} c_j^{ab} d_{t-1,i}^b B_{j,m}(\tau - s) B_{i,m}(s) \right\} ds + \sum_{j=1}^{\infty} d_j^a(\varepsilon_t^{(a)}) B_{j,m}(\tau)$$

$$= \sum_{h=1}^{\infty} p_h^a B_{h,m}(\tau)$$

$$+ \sum_{h=1}^{\infty} \sum_{i=1}^{\infty} \left\{ \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{aa} - c_h^{aa} \right\} \frac{w_{i+m} - w_i}{m} d_{t-1,i}^a B_{h,m}(\tau)$$

$$+ \sum_{h=1}^{\infty} \sum_{i=1}^{\infty} \left\{ \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{ab} - c_h^{ab} \right\} \frac{w_{i+m} - w_i}{m} d_{t-1,i}^b B_{h,m}(\tau)$$

$$+ \sum_{j=1}^{\infty} d_j^a(\varepsilon_t^{(a)}) B_{j,m}(\tau)$$

$$X_t^{(b)}(\tau) = \sum_{h=1}^{\infty} p_h^b B_{h,m}(\tau) + \int_0^1 \left\{ \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} c_j^{bb} d_{t-1,i}^b B_{j,m}(\tau - s) B_{i,m}(s) \right\} ds$$

$$+ \int_0^1 \left\{ \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} c_j^{ba} d_{t-1,i}^a B_{j,m}(\tau - s) B_{i,m}(s) \right\} ds + \sum_{j=1}^{\infty} d_j^b(\varepsilon_t^{(b)}) B_{j,m}(\tau)$$

$$= \sum_{h=1}^{\infty} p_h^b B_{h,m}(\tau)$$

$$+ \sum_{h=1}^{\infty} \sum_{i=1}^{\infty} \left\{ \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{bb} - c_h^{bb} \right\} \frac{w_{i+m} - w_i}{m} d_{t-1,i}^b B_{h,m}(\tau)$$

$$+ \sum_{h=1}^{\infty} \sum_{i=1}^{\infty} \left\{ \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{ba} - c_h^{ba} \right\} \frac{w_{i+m} - w_i}{m} d_{t-1,i}^a B_{h,m}(\tau)$$

$$+ \sum_{j=1}^{\infty} d_j^b(\varepsilon_t^{(b)}) B_{j,m}(\tau) \tag{3}$$

Rearranging the above equations, we obtain the relationship of the B-spline coefficients in the framework of VFAR:

$$d_{t,h}^a = p_h^a + \sum_{i=1}^{\infty} \left\{ \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{aa} - c_h^{aa} \right\} \frac{w_{i+m} - w_i}{m} d_{t-1,i}^a$$

$$+ \sum_{i=1}^{\infty} \left\{ \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{ab} - c_h^{ab} \right\} \frac{w_{i+m} - w_i}{m} d_{t-1,i}^b + d_h^a(\varepsilon_t^{(a)})$$

$$d_{t,h}^b = p_h^b + \sum_{i=1}^{\infty} \left\{ \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{bb} - c_h^{bb} \right\} \frac{w_{i+m} - w_i}{m} d_{t-1,i}^b$$

$$+ \sum_{i=1}^{\infty} \left\{ \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{ba} - c_h^{ba} \right\} \frac{w_{i+m} - w_i}{m} d_{t-1,i}^a + d_h^b(\varepsilon_t^{(b)}) \tag{4}$$

The original problem of estimating the functional parameters is converted to the

estimation of the B-spline coefficients. It is however impossible to estimate infinite coefficients given finite sample.

3.1 Sieve estimator

We introduce a sequence of subsets - a sieve for a parameter space Θ, is denoted by $\{\Theta_{J_n}\}$ where $\Theta_{J_n} \subseteq \Theta_{J_n+1}$ and the union of subsets $\bigcup \Theta_{J_n}$ is dense in the parameter space. While allowing the dimension of the subset to increase when sample size gets larger, we will estimate the unknown parameters on the finite subset of the parameter space. The sieve is defined as follows:

$$\Theta_{J_n} = \left\{ \kappa_{xy} \in L^2 \mid \kappa_{xy}(\tau) = \sum_{l=1}^{J_n} c_l^{xy} B_{l,m}(\tau), \quad \tau \in [0,1], \quad \sum_{l=1}^{J_n} l^2 (c_l^{xy})^2 \leq v J_n \right\} \tag{5}$$

where $J_n \to +\infty$ as $n \to +\infty$ and v is some known positive constant such that without any sacrifice of the growth rate of J_n, the constraint for c_l^{xy} can be satisfied generally, see e.g. Grenander (1981) on the theory of sieves.

Under the sieve with J_n, Equation (4) can be represented in matrix form, which yields a form of Vector AutoRegressive (VAR) of order 1:

$$\begin{bmatrix} d_{t,1}^a \\ \vdots \\ d_{t,J_n}^a \\ d_{t,1}^b \\ \vdots \\ d_{t,J_n}^b \end{bmatrix} = \begin{bmatrix} p_1^a \\ \vdots \\ p_{J_n}^a \\ p_1^b \\ \vdots \\ p_{J_n}^b \end{bmatrix} + \begin{bmatrix} r_{1,1}^{aa} & \cdots & r_{1,J_n}^{aa} & r_{1,1}^{ab} & \cdots & r_{1,J_n}^{ab} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ r_{J_n,1}^{aa} & \cdots & r_{J_n,J_n}^{aa} & r_{J_n,1}^{ab} & \cdots & r_{J_n,J_n}^{ab} \\ r_{1,1}^{ba} & \cdots & r_{1,J_n}^{ba} & r_{1,1}^{bb} & \cdots & r_{1,J_n}^{bb} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ r_{J_n,1}^{ba} & \cdots & r_{J_n,J_n}^{ba} & r_{J_n,1}^{bb} & \cdots & r_{J_n,J_n}^{bb} \end{bmatrix} \begin{bmatrix} d_{t-1,1}^a \\ \vdots \\ d_{t-1,J_n}^a \\ d_{t-1,1}^b \\ \vdots \\ d_{t-1,J_n}^b \end{bmatrix} + \begin{bmatrix} d_1^a(\varepsilon_t^{(a)}) \\ \vdots \\ d_{J_n}^a(\varepsilon_t^{(a)}) \\ d_1^b(\varepsilon_t^{(b)}) \\ \vdots \\ d_{J_n}^b(\varepsilon_t^{(b)}) \end{bmatrix} \tag{6}$$

where $r_{h,i}^{xy}$ denotes $\left\{ \sum_{j=1}^{J_n} \left(\frac{w_{j+m}-w_{j+1}}{w_{j+m}-w_j} - \frac{w_{j+m+1}-w_{j+2}}{w_{j+m+1}-w_{j+1}} \right) c_j^{xy} - c_h^{xy} \right\} \frac{w_{i+m}-w_i}{m}$, for $xy = aa$, ab, ba, and bb. Equation (6) can be also represented as:

$$y_t = v + C y_{t-1} + u_t \tag{7}$$

where $y_t = \left(d_{t,1}^a, \cdots, d_{t,J_n}^a, d_{t,1}^b, \cdots, d_{t,J_n}^b \right)^\top$, $v = \left(p_1^a, \cdots, p_{J_n}^a, p_1^b, \cdots, p_{J_n}^b \right)^\top$, $u_t = \left(d_1^a(\varepsilon_t^{(a)}), \cdots, d_{J_n}^a(\varepsilon_t^{(a)}), d_1^b(\varepsilon_t^{(b)}), \cdots, d_{J_n}^b(\varepsilon_t^{(b)}) \right)^\top$, and C be the matrix with elements

$r_{h,i}^{xy}$ in (6).

Assuming the presample value y_0 is available, define:

$$
\begin{aligned}
Y &= (y_1, \cdots, y_n), \\
B &= (v, C), \\
Z_t &= \begin{bmatrix} 1 \\ y_t \end{bmatrix}, \\
Z &= (Z_0, \cdots, Z_{n-1}), \\
U &= (u_1, \cdots, u_n), \\
\mathbf{y} &= \mathrm{vec}(Y), \\
\boldsymbol{\beta} &= \mathrm{vec}(B), \\
\mathbf{u} &= \mathrm{vec}(U), \\
K &= 2J_n.
\end{aligned}
$$

where vec is the column stacking operator. Using the notations, for $t = 1, \cdots, n$, we can write (7) compactly as the following:

$$Y = BZ + U \tag{8}$$

By applying vec operator to (8) 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.

We impose an assumption that the B-spline coefficients $d_j^a(\varepsilon_t^{(a)})$ are independently and identically Gaussian distributed with mean zero and constant variance $\sigma_{j,a}^2$. The same applies for $d_j^b(\varepsilon_t^{(b)})$ with $\sigma_{j,b}^2$. Following Geman and Hwang (1982), we define the likelihood function for VFAR over the approximating subspace (5) of the original parameter space. Assuming

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

12

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

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

In addition,

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

where $\mathbf{v} = (v, \cdots, v)^{\top}$ is a $(Kn \times 1)$ vector. Consequently, $\frac{\partial \mathbf{u}}{\partial \mathbf{y}^{\top}}$ 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:

$$g\left(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \rho^{aa}, \rho^{ab}, \rho^{ba}, \rho^{bb}\right) = f_{\mathbf{y}}(\mathbf{y}) = \left|\frac{\partial \mathbf{u}}{\partial \mathbf{y}^{\top}}\right| f_{\mathbf{u}}(\mathbf{u})$$

$$= \frac{1}{(2\pi)^{Kn/2}}\left|I_n \otimes \Sigma_u\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\mathbf{y} - (Z^{\top} \otimes I_K)\boldsymbol{\beta}\right)^{\top}(I_n \otimes \Sigma_u^{-1})\left(\mathbf{y} - (Z^{\top} \otimes I_K)\boldsymbol{\beta}\right)\right\}.$$

The (approximated) log-likelihood function is:

$$\ell\left(X_1^{(a)}, \cdots, X_n^{(a)}, X_1^{(b)}, \cdots, X_n^{(b)}; \rho^{aa}, \rho^{ab}, \rho^{ba}, \rho^{bb}\right) = \ell(\boldsymbol{\beta}, \Sigma_u)$$

$$= -\frac{Kn}{2}\log 2\pi - \frac{n}{2}\log\left|\Sigma_u\right| - \frac{1}{2}\left(\mathbf{y} - (Z^{\top} \otimes I_K)\boldsymbol{\beta}\right)^{\top}(I_n \otimes \Sigma_u^{-1})\left(\mathbf{y} - (Z^{\top} \otimes I_K)\boldsymbol{\beta}\right)$$

$$= -\frac{Kn}{2}\log 2\pi - \frac{n}{2}\log\left|\Sigma_u\right| - \frac{1}{2}\sum_{t=1}^{n}\left(y_t - v - Cy_{t-1}\right)^{\top}\Sigma_u^{-1}\left(y_t - v - Cy_{t-1}\right)$$

$$= -\frac{Kn}{2}\log 2\pi - \frac{n}{2}\log\left|\Sigma_u\right| - \frac{1}{2}\sum_{t=1}^{n}\left(y_t - Cy_{t-1}\right)^{\top}\Sigma_u^{-1}\left(y_t - Cy_{t-1}\right)$$

$$\quad + v^{\top}\Sigma_u^{-1}\sum_{t=1}^{n}\left(y_t - Cy_{t-1}\right) - \frac{n}{2}v^{\top}\Sigma_u^{-1}v$$

$$= -\frac{Kn}{2}\log 2\pi - \frac{n}{2}\log\left|\Sigma_u\right| - \frac{1}{2}\operatorname{tr}\left((Y - BZ)^{\top}\Sigma_u^{-1}(Y - BZ)\right)$$

and the first order partial differentiations are as follows:

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = (Z \otimes I_K)(I_n \otimes \Sigma_u^{-1})\Big(\mathbf{y} - (Z^\top \otimes I_K)\boldsymbol{\beta}\Big)$$

$$= (Z \otimes \Sigma_u^{-1})\mathbf{y} - (ZZ^\top \otimes \Sigma_u^{-1})\boldsymbol{\beta} \tag{9}$$

$$\frac{\partial \ell}{\partial \Sigma_u} = -\frac{n}{2}\Sigma_u^{-1} + \frac{1}{2}\Sigma_u^{-1}(Y - BZ)(Y - BZ)^\top \Sigma_u^{-1}$$

By equating the first order partial derivatives in (9) to zero, we obtain the maximum likelihood estimators:

$$\widehat{\boldsymbol{\beta}} = \Big\{(ZZ^\top)^{-1}Z \otimes I_K\Big\}\mathbf{y}, \text{ or equivalently,}$$

$$\widehat{B} = (\widehat{v}, \widehat{C}) = YZ^\top(ZZ^\top)^{-1} \tag{10}$$

$$\widehat{\Sigma_u} = \frac{1}{n}(Y - BZ)(Y - BZ)^\top$$

The first column of $YZ^\top(ZZ^\top)^{-1}$ in (10) is the estimator for $v = \Big(p_1^a, \cdots, p_{J_n}^a, p_1^b,$

$\cdots, p_{J_n}^b\Big)^\top$. To show the estimator for c_j^{xy} for $xy = aa, ab, ba, bb$ as in (2), we further define the following notations:

$$W = \text{diag}\Big(\frac{m}{w_{1+m} - w_1}, \cdots, \frac{m}{w_{J_n+m} - w_{J_n}}, \frac{m}{w_{1+m} - w_1}, \cdots, \frac{m}{w_{J_n+m} - w_{J_n}}\Big),$$

$$q_j = \frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}},$$

$$\theta_1 = (c_1^{aa}, \cdots, c_{J_n}^{aa}, c_1^{ba}, \cdots, c_{J_n}^{ba})^\top,$$

$$\theta_2 = (c_1^{ab}, \cdots, c_{J_n}^{ab}, c_1^{bb}, \cdots, c_{J_n}^{bb})^\top,$$

$$\boldsymbol{\theta} = (\theta_1, \cdots, \theta_1, \theta_2, \cdots, \theta_2),$$

$$Q = \begin{bmatrix} q_1 - 1 & q_2 & \cdots & q_{J_n} & & & & & \mathbf{0} \\ q_1 & q_2 - 1 & \cdots & q_{J_n} & & & & & \\ \vdots & \vdots & \ddots & \vdots & & & & & \\ q_1 & q_2 & \cdots & q_{J_n} - 1 & & & & & \\ & & & & q_1 - 1 & q_2 & \cdots & q_{J_n} \\ & & & & q_1 & q_2 - 1 & \cdots & q_{J_n} \\ & & & & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & & & & q_1 & q_2 & \cdots & q_{J_n} - 1 \end{bmatrix},$$

where $\boldsymbol{\theta}$ contains J_n columns of θ_1 and J_n columns of θ_2. Therefore we have the

14

estimator for c_j^{xy} for $xy = aa, ab, ba, bb$ as follows:

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

3.2 Asymptotic property

We now derive the consistency results 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 $\boldsymbol{\psi}$:

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

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

C1: 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 $\left\|\boldsymbol{\rho}_{J_n} - \boldsymbol{\rho}_{0|\Theta_{J_n}}\right\|_{HS} \to 0$; meaning each $\left\|\rho_{J_n}^{xy} - \rho_{0|\Theta_{J_n}}^{xy}\right\|_{HS} \to 0$, 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}.

C2: There exists a sequence $\{\boldsymbol{\rho}_{J_n}\}$ described in **C1** such that $H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}_{J_n}) \to H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}_{0|\Theta_{J_n}})$.

The norm $\|\cdot\|_{\mathscr{S}}$ is a Hilbert-Schmidt norm for the convolution kernel operator. Recall 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 $\sum_j \lambda_j^2 < \infty$. The convolution kernel operator satisfies the definition and its Hilbert-Schmidt norm is $\|\rho\|_{\mathscr{S}} = (\sum_j \lambda_j^2)^{1/2}$. The Hilbert-Schmidt norm is chosen for our study because of the fact that the convolution kernel operator defined in our paper forms a class of operators embedded in the whole space of Hilbert-Schmidt operators and for any convolution kernel operator ρ, we have the Hilbert-Schmidt norm of ρ equal to the L^2 norm of its kernel function, that is $\|\rho\|_{HS} = \|\kappa\|_2$.

Theorem 3.1 *Assume $\{\Theta_{J_n}\}$ is chosen such that conditions $C1$ and $C2$ are in force. Suppose that for each $\delta > 0$, we can find subsets $\Gamma_1, \Gamma_2, \cdots, \Gamma_{l_{J_n}}$ of Θ_{J_n}, $J_n = 1, 2, \cdots$ such that*

(i) $D_{J_n} \subseteq \bigcup_{k=1}^{l_{J_n}} \Gamma_k$, where $D_{J_n} = \{\boldsymbol{\rho} \in \Theta_{J_n} | H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}) \leq H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}_{J_n}) - \delta\}$ 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 $\Gamma_1, \cdots, \Gamma_l$ in Θ_{J_n}, where $\varphi_{J_n} = \sup_k \inf_{t \geq 0} \mathbb{E}_{\boldsymbol{\rho}_{0|\Theta_{J_n}}} \exp \left\{ t \log \frac{g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \Gamma_k)}{g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \boldsymbol{\rho}_{J_n})} \right\}$.*

Then we have $\sup_{\widehat{\boldsymbol{\rho}}_n \in M_{J_n}^n} \|\widehat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_{0|\Theta_{J_n}}\|_{HS} \to 0$ a.s.

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 ML estimators on Θ_{J_n} given the sample size n as $M_{J_n}^n = \{\boldsymbol{\rho} \in \Theta_{J_n} | \ell(X_1^{(a)}, \cdots, X_n^{(a)}, X_1^{(b)}, \cdots, X_n^{(b)}; \boldsymbol{\rho}) = \sup_{\boldsymbol{\psi} \in \Theta_{J_n}} \ell(X_1^{(a)}, \cdots, X_n^{(a)}, X_1^{(b)}, \cdots, X_n^{(b)}; \boldsymbol{\psi})\}$. The proof of Theorem 3.1 shows the convergence of the ML estimator to $\boldsymbol{\rho}_{0|\Theta_{J_n}}$, the projections of the true operators on sieve, see Appendix. 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 $n \to +\infty$ and $\|\cdot\|_2$ is the L^2 norm in $C_{[0,1]}$.*
$\widehat{\boldsymbol{\kappa}}_{J_n} = (\widehat{\kappa}_{aa,J_n}, \widehat{\kappa}_{ab,J_n}, \widehat{\kappa}_{ba,J_n}, \widehat{\kappa}_{bb,J_n})$ *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}. $\|\widehat{\boldsymbol{\kappa}}_{J_n} - \boldsymbol{\kappa}_{0|\Theta_{J_n}}\|_2 \to 0$ 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 the Appendix. 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 Empirical applications of the VFAR model

In this section, we apply the VFAR model to estimate the joint dynamics of the liquidity demand and supply curves and investigate its in-sample and out-of-sample predictability.

4.1 In-sample estimation

We conduct in-sample estimation based on the liquidity demand and supply curves over 44 trading days from date 2 Jan 2015 to 6 Mar 2015. We employs B-spline

expansions with equally-spaced price percentile as nodes and $J_n = 20$ in the sieve. There are in total 20 coefficients for the bid and another 20 for the ask liquidity curves. Moreover, we perform estimation with the Random Walk (RW) model of no drift, where the liquidity curves are directly estimated by their most recent curves at the previous time point. Though simple, random walk provides a general good predictability and hard to beat under market efficiency.

We use three measures as indicators of predict performance, the root mean squared estimation error (RMSE) and the mean absolute percentage error (MAPE) for accuracy, and R^2 for the explanatory power:

$$RMSE = \sqrt{\frac{\sum_{xy=a,b} \sum_{t=1}^{n} \sum_{\tau} \left\{ X_t^{(xy)}(\tau) - \widehat{X}_t^{(xy)}(\tau) \right\}^2}{N}}$$

$$MAPE = \frac{\sum_{xy=a,b} \sum_{t=1}^{n} \sum_{\tau} \frac{\left| X_t^{(xy)}(\tau) - \widehat{X}_t^{(xy)}(\tau) \right|}{X_t^{(xy)}(\tau)}}{N}$$

$$R^2 = 1 - \frac{\sum_{xy=a,b} \sum_{t=1}^{n} \sum_{\tau} \left\{ X_t^{(xy)}(\tau) - \widehat{X}_t^{(xy)}(\tau) \right\}^2}{\sum_{xy=a,b} \sum_{t=1}^{n} \sum_{\tau} \left\{ X_t^{(xy)}(\tau) - \bar{X} \right\}^2} \tag{11}$$

We calculate these measures for the estimated liquidity curves in the VFAR models and the alternative RW model.

Table 2 reports the R^2, RMSE and MAPE of the estimated liquidity curves in the VFAR model. It shows that VFAR provides high explanatory power for all the stocks, with R^2 ranging from 92 percent (AAPL) to 98 percent (AEZS), RMSE smaller than 0.34 (AAPL) and MAPE lower than 3.61 percent. On the right panel, the alternative RW model is compared with the VFAR model by calculating the ratio of each measure. In each column, the number in bold-face indicates the best relative performance of VFAR for each stock and performance measure.

Table 2 shows that, without exception, the VFAR model is always better than the RW model. In terms of R^2, VFAR outperforms by up to 3 percent (AAPL and CMCSA). As for estimation accuracy, the relative performance reaches to 13 percent in MAPE (CSCO) and at least 9 percent (SIRI, the most active stock) and up to 45 percent (AEZS that has the smallest bid-ask spread on average).

To visualize the in-sample fit, Figure 3 depicts the estimated bid and ask supply curves vs. the observed ones at an arbitrarily selected date, 24 February 2015 at

Ticker Symbol	VFAR			RW vs VFAR		
	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**

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

3p.m. for four stocks, AAPL, AMAT, AEZS and SIRI, representing heterogeneous stocks in terms of market capitalization and liquidity. The estimated curves display V-shape, and reasonably trace both the actual queuing orders displayed as discrete dots as well as the smoothed liquidity curves in grey colour. Moreover, the accuracy is stable in the middle around the best quotes and also the tails.

4.2 Forecast

We make an out-of-sample forecast for the liquidity curves starting from the 31st trading day onwards and predict $1-, 5-$ and $10-$step ahead forecasts that correspond to $5-, 25-$ and $50-$minute ahead liquidity curves respectively. The first pair of forecasted curves is for time $t = 2251$, based on the past 30 trading days of $30 \times 75 = 2250$ functional objects. Each time, we move forward one period, i.e. 5 minutes at a time and perform re-estimation and forecast until reaching the end of the sample at $t = 3300$.

Figure 4 gives graphical illustrations of the forecasted liquidity curves for AAPL with the VFAR model. The forecasts closely trace the realized liquidity curves. What is dramatic is its capacity 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 $50-$minute.

Table 3 reports the forecast RMSE, MAPE and predict power for liquidity curves of the 12 stocks. Even if in the worst case, the VFAR approach in forecasting is
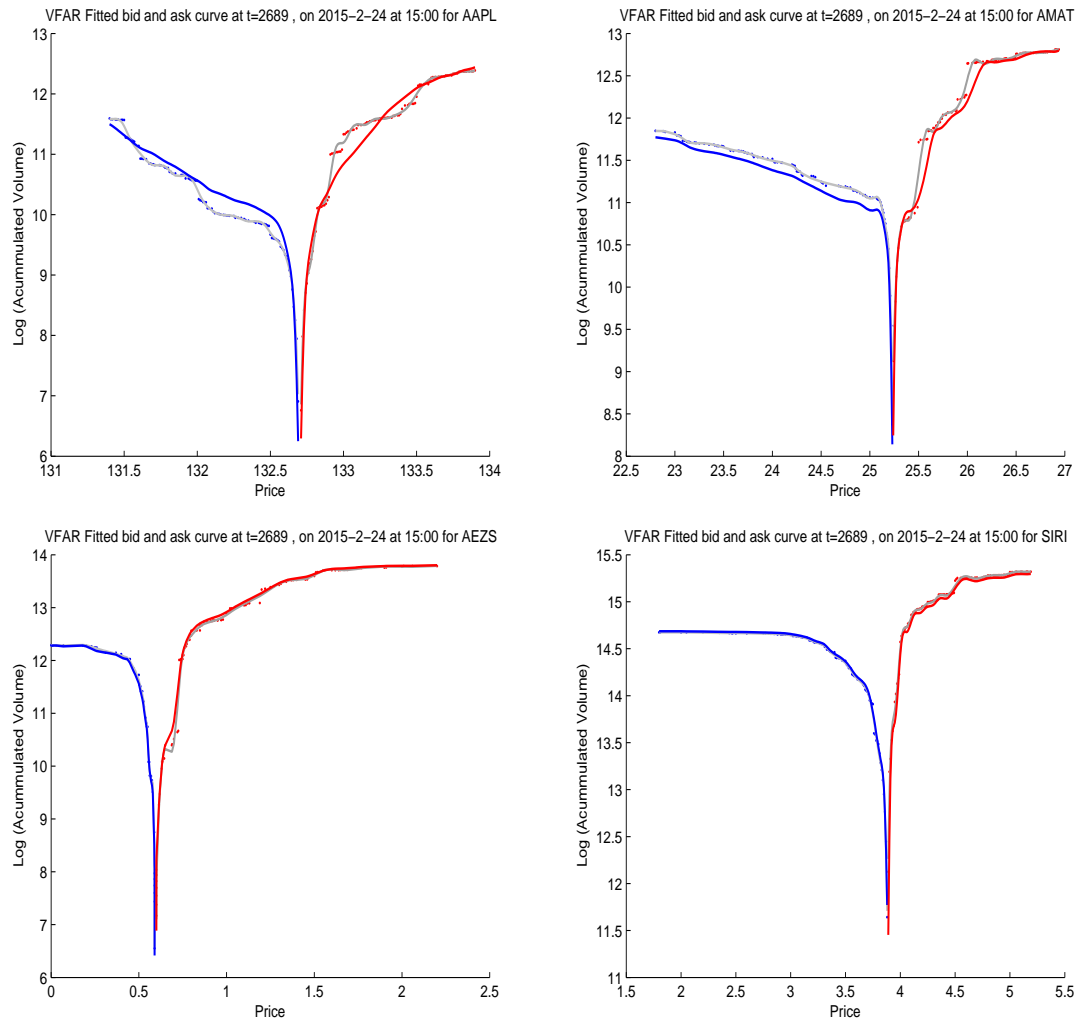
Figure 3: Estimated bid (and ask) supply curves vs. the actually observed

VFARrandVfarPlot

able to achieve high R^2 ranging from 91.13 percent (1-step AAPL) to 83.74 percent (10-step AAPL), low RMSE of 0.48 (1-step AEZS) to 0.58 (10-step AEZS), and low MAPE of 3.61 percent (1-step AAPL) to 4.49 percent (10-step AAPL). In addition to the forecasts from the VFAR model, we also compute forecasts from the random walk model, see 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 tremendous in terms of the reduction in the RMSE of the VFAR model reaches about 4 percent (1-step SIRI) in the worst case and 36 percent (1-step AEZS) and 40 percent (5-step AEZS) in the best case. Compared with the random walk model, the VFAR model does not always have an absolute advantage for the MAPE comparison. As for MAPE, only in 5 out of 36 instances, we have the RW performing better than VFAR. In the other 31 cases, VFAR outperforms the RW by up to 20 percent. The relative superior performance grows as the forecast horizon increases, indicating that the utilization of cross-dependence in liquidity curves helps to improve out-of-sample prediction.

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, for example, deriving an optimal trading strategy and forecasting of the demand and supply elasticities.

5   Conclusion

Predictions of future liquidity supply and demand in the limit order book (LOB) helps in analyzing optimal splitting strategies for large orders to reduce cost, see Härdle et al. (2012). 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 a high-dimensional problem. In addition, we see significant cross-dependency of the bid and ask side of the market in Section 3. We proposed a Vector Functional AutoRegressive (VFAR) model to model and forecast LOB liquidity supply-demand curves, taking into consideration of the bid-ask cross dependency.

The model is applied to 12 stocks in the National Association of Securities Dealers Automated Quotations (NASDAQ) stock market. It is shown that our model gives R^2 values as high as 98.5 percent for in-sample estimation. In out-of-sample forecast experiments, it produces accurate $5-, 25-$ and $50-$minutes forecasts, with mean absolute percentage error as low as 0.3 to 4.5 percent. Most important, the VFAR
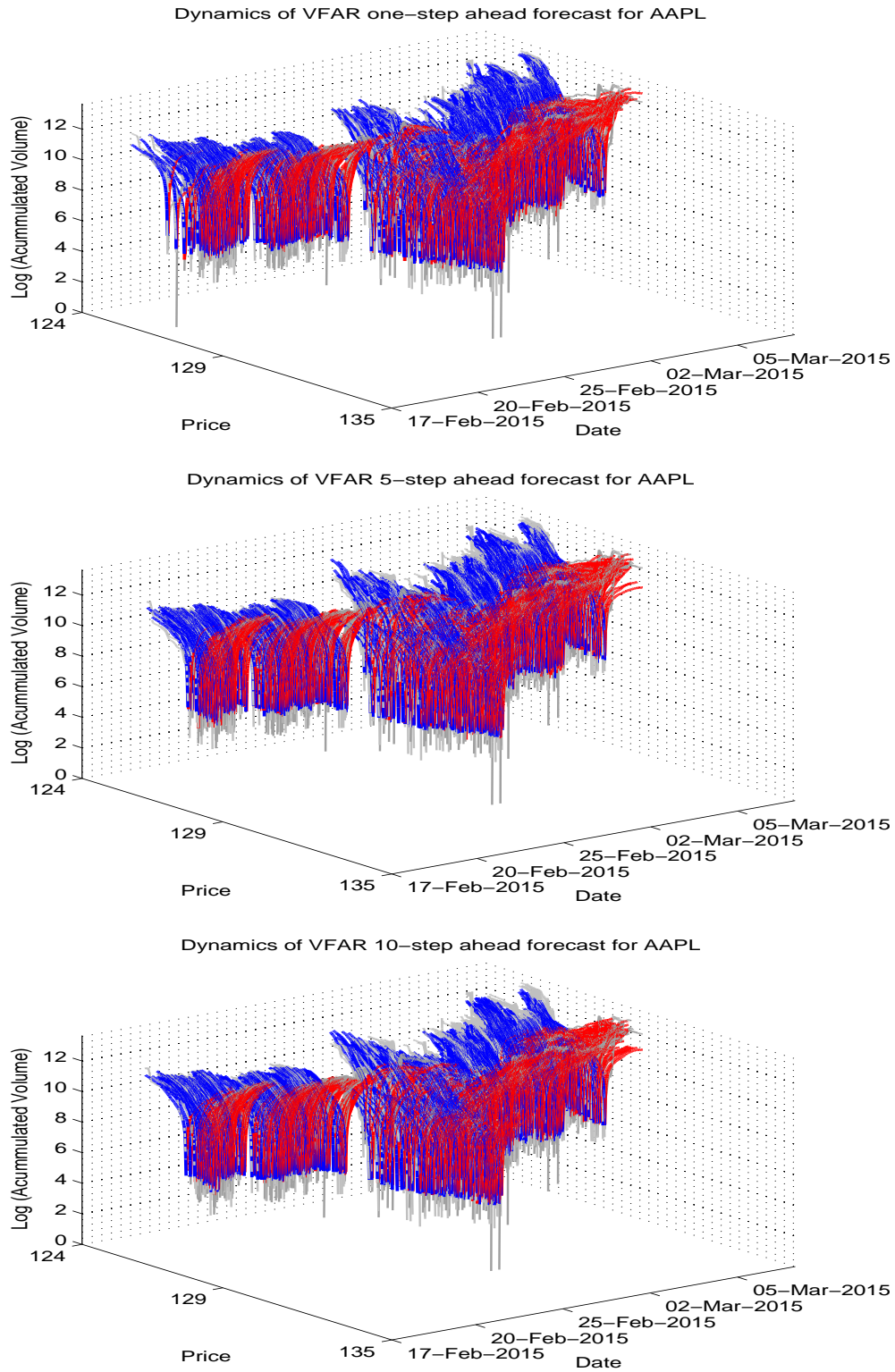
Figure 4: Dynamics of multi-step ahead forecast for AAPL. Top: 5−minute ahead forecast; Middle: 25−minute ahead forecast; Bottom: 50−minute ahead forecast.

VFARdynamicPlot

Ticker Symbol	R^2			RMSE			MAPE		
	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 3: R^2, RMSE, MAPE for multi-step ahead VFAR forecast of the 12 stocks

Ticker Symbol	R^2			RMSE			MAPE		
	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

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

model is general and can be used for other multiple functional time series modeling and forecasting.

A   Appendix

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

First we show how the expansion was obtained in (3). We only show for the first integral in (3) as the expansion other integrals can be obtained similarly.

$$\int_0^1 \left\{ \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} c_j^{aa} d_{t-1,i}^a B_{j,m}(\tau - s) B_{i,m}(s) \right\} ds$$

$$= \sum_{j=1}^{\infty} c_j^{aa} \int_0^1 B_{j,m}(\tau - s) \left\{ \sum_{i=1}^{\infty} d_{t-1,i}^a B_{i,m}(s) \right\} ds$$

$$= \sum_{j=1}^{\infty} c_j^{aa} \left\{ \frac{1}{m} \sum_{i=1}^{\infty} d_{t-1,i}^a (w_{i+m} - w_i) B_{j,m}(\tau - s) \right\} \Big|_{s=0}^{s=1}$$

$$+ \sum_{j=1}^{\infty} c_j^{aa} \int_0^1 \left\{ \frac{1}{m} \sum_{i=1}^{\infty} d_{t-1,i}^a (w_{i+m} - w_i) \left\{ \frac{m}{w_{j+m} - w_j} B_{j,m-1}(\tau - s) \right. \right.$$

$$\left. \left. - \frac{m}{w_{j+m+1} - w_{j+1}} B_{j+1,m-1}(\tau - s) \right\} \right\} ds$$

$$= - \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \frac{d_{t-1,i}^a (w_{i+m} - w_i)}{m} c_j^{aa} B_{j,m}(\tau)$$

$$+ \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \frac{d_{t-1,i}^a (w_{i+m} - w_i)}{m} c_j^{aa} \int_{\tau-1}^{\tau} \left\{ \frac{m}{w_{j+m} - w_j} B_{j,m-1}(z) \right.$$

$$\left. - \frac{m}{w_{j+m+1} - w_{j+1}} B_{j+1,m-1}(z) \right\} dz$$

$$= - \sum_{h=1}^{\infty} \sum_{i=1}^{\infty} \frac{d_{t-1,i}^a (w_{i+m} - w_i)}{m} c_h^{aa} B_{h,m}(\tau)$$

$$+ \sum_{h=1}^{\infty} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{aa} \frac{w_{i+m} - w_i}{m} d_{t-1,i}^a B_{h,m}(\tau)$$

$$= \sum_{h=1}^{\infty} \sum_{i=1}^{\infty} \left[\sum_{j=1}^{\infty} \left\{ \frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right\} c_j^{aa} - c_h^{aa} \right] \frac{w_{i+m} - w_i}{m} d_{t-1,i}^a B_{h,m}(\tau)$$

The second equality made use of integration by parts, with $\frac{d}{ds} B_{j,m}(\tau - s) = - \left(\frac{m}{w_{j+m} - w_j} \right.$ $B_{j,m-1}(\tau - s) - \frac{m}{w_{j+m+1} - w_{j+1}} B_{j+1,m-1}(\tau - s) \right)$ and $\int \sum_{i=1}^{\infty} d_{t-1,i}^a B_{i,m}(s) ds = \frac{1}{m} \sum_{i=1}^{\infty}$

25

$d_{t-1,i}^a(w_{i+m} - w_i)$. In the third equality, we made the substitution of $z = \tau - s$. For the fourth equality, we made use of the formula: $\int_{-\infty}^{\tau} B_{j,m}(z)dz = \frac{w_{j+m+1} - w_{j+1}}{m+1} \sum_{h=1}^{\infty} B_{h,m+1}(\tau)$, and truncating the sum up till the J-th term. We also swapped the notation j for the first summation with h in the fourth equality.

A.2   Proof of Theorem 3.1

Fix $\delta > 0$. We only need to show that

$$P(D_{J_n} \cap M_{J_n}^n \neq \emptyset) = 0, \tag{12}$$

because if (12) 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 **C2** we deduce

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

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,

$$\lim_{n \to +\infty} \sup_{\boldsymbol{\rho} \in M_{J_n}^n} |H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}) - H(\boldsymbol{\rho}_{0|\Theta_{J_n}}, \boldsymbol{\rho}_{0|\Theta_{J_n}})| = 0 \quad a.s. \tag{13}$$

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 **C1** combined with (13) imply that

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

Hence,

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

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 (12).

For now, n and J_n are fixed. Then

$$(D_{J_n} \cap M_{J_n}^n \ne \emptyset)$$
$$\subseteq \left\{ \sup_{\boldsymbol{\rho} \in D_{J_n}} \ell(X_1^{(a)}, \cdots, X_n^{(a)}, X_1^{(b)}, \cdots, X_n^{(b)}; \boldsymbol{\rho}) \ge \ell(X_1^{(a)}, \cdots, X_n^{(a)}, X_1^{(b)}, \cdots, X_n^{(b)}; \boldsymbol{\rho}_{J_n}) \right\}$$
$$\subseteq \bigcup_{k=1}^{l_{J_n}} \left\{ \sup_{\boldsymbol{\rho} \in \Gamma_k} \prod_{i=1}^{n} g\left(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \boldsymbol{\rho}\right) \le \prod_{i=1}^{n} g\left(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \boldsymbol{\rho}_{J_n}\right) \right\}$$
$$\subseteq \bigcup_{k=1}^{l_{J_n}} \left\{ \prod_{i=1}^{n} g\left(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \Gamma_k\right) \le \prod_{i=1}^{n} g\left(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \boldsymbol{\rho}_{J_n}\right) \right\}.$$

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

$$\pi \le \sum_{k=1}^{l_{J_n}} P\left[\prod_{i=1}^{n} g\left(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \Gamma_k\right) \le \prod_{i=1}^{n} g\left(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \boldsymbol{\rho}_{J_n}\right) \right]$$
$$= \sum_{k=1}^{l_{J_n}} P\left[\exp \sum_{i=1}^{n} \left\{ t_k \log \frac{g(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \Gamma_k)}{g(X_i^{(a)}, X_i^{(b)}, X_{i-1}^{(a)}, X_{i-1}^{(b)}, \boldsymbol{\rho}_{J_n})} \right\} \ge 1 \right]$$
$$\le \sum_{k=1}^{l_{J_n}} \mathbb{E}_{\boldsymbol{\rho}_{0|\Theta_{J_n}}} \left[\exp \left\{ t_k \log \frac{g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \Gamma_k)}{g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \boldsymbol{\rho}_{J_n})} \right\} \right]^{n}$$

for any nonnegative arbitrary $t_1, \cdots, t_k$ 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

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

27

Finally, result (12) is deduced by condition (ii) of Theorem 3.1 and by the Borel-Cantelli lemma.

A.3 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 **C1**. 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

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

where

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

$$
x_J = \begin{bmatrix}
d_{t,1}^a - \sum_i(\sum_j q_j c_{j,J}^{aa} - c_{1,J}^{aa})\frac{w_{i+m}-wi}{m}d_{t-1,i}^a - \sum_i(\sum_j q_j c_{j,J}^{ab} - c_{1,J}^{ab})\frac{w_{i+m}-wi}{m}d_{t-1,i}^b \\
\vdots \\
d_{t,J}^a - \sum_i(\sum_j q_j c_{j,J}^{aa} - c_{J,J}^{aa})\frac{w_{i+m}-wi}{m}d_{t-1,i}^a - \sum_i(\sum_j q_j c_{j,J}^{ab} - c_{J,J}^{ab})\frac{w_{i+m}-wi}{m}d_{t-1,i}^b \\
d_{t,1}^b - \sum_i(\sum_j q_j c_{j,J}^{ba} - c_{1,J}^{ba})\frac{w_{i+m}-wi}{m}d_{t-1,i}^a - \sum_i(\sum_j q_j c_{j,J}^{bb} - c_{1,J}^{bb})\frac{w_{i+m}-wi}{m}d_{t-1,i}^b \\
\vdots \\
d_{t,J}^b - \sum_i(\sum_j q_j c_{j,J}^{ba} - c_{J,J}^{ba})\frac{w_{i+m}-wi}{m}d_{t-1,i}^a - \sum_i(\sum_j q_j c_{j,J}^{bb} - c_{J,J}^{bb})\frac{w_{i+m}-wi}{m}d_{t-1,i}^b
\end{bmatrix}.
$$

Here $\Sigma_u, c_j^{aa}, c_j^{ab}, c_j^{ba}$, and c_j^{bb} denote the covariance matrix and B-spline coefficients for the kernel $\boldsymbol{\kappa}_{0|\Theta_J}$; and $\Sigma_{u,J}, c_{j,J}^{aa}, c_{j,J}^{ab}, c_{j,J}^{ba}$, and $c_{j,J}^{bb}$ 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

$$H(\boldsymbol{\rho}_{0|\Theta_J}, \boldsymbol{\rho}_{0|\Theta_J}) - H(\boldsymbol{\rho}_{0|\Theta_J}, \boldsymbol{\rho}_{\Theta_J}) = \mathbb{E}\Big(-\frac{1}{2}x^\top \Sigma_u^{-1} x + \frac{1}{2}x_J^\top \Sigma_u^{-1} x_J\Big)$$
$$= \frac{1}{2}\sum_{r,s}(\Sigma_u^{-1})_{r,s}\mathbb{E}\Big\{(x_J)_r(x_J)_s - (x)_r(x)_s\Big\},$$

where $(\Sigma_u^{-1})_{r,s}$ is the r-th row, s-th column of Σ_u^{-1}, $(x_J)_r$ is the r-th element of x_J, and $(x)_r$ is the r-th 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 cancelling 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 $c_{j,J}^{aa} \to c_j^{aa}$, $c_{j,J}^{ab} \to c_j^{ab}$, $c_{j,J}^{ba} \to c_j^{ba}$, $c_{j,J}^{bb} \to c_j^{bb}$ and consequently $\boldsymbol{\rho}_J \to \boldsymbol{\rho}_{0|\Theta_J}$.

For the condition **C2** and (i) of Theorem 3.1, we follow similar arguments as in Mourid and Bensmain (2006). To verify Theorem 3.1 (ii), we define

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

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 $\varphi' = \mathbb{E}_{\boldsymbol{\kappa}_{0|\Theta_J}}\log\frac{g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \Gamma_k)}{g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \boldsymbol{\kappa}_J)}$.

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

$$A = \mathbb{E}_{\boldsymbol{\kappa}_{0|\Theta_J}}\log g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \Gamma_k) - \mathbb{E}\log g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \boldsymbol{\kappa})$$
$$= \mathbb{E}_{\boldsymbol{\kappa}_{0|\Theta_J}}\sup_{\boldsymbol{\psi}\in\Gamma_k}\left\{\log g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \boldsymbol{\psi}) - \log g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \boldsymbol{\kappa})\right\}$$
$$= \mathbb{E}_{\boldsymbol{\kappa}_{0|\Theta_J}}\sup_{\boldsymbol{\psi}\in\Gamma_k}\left\{-\frac{1}{2}\log|\Sigma_{u,\boldsymbol{\psi}}| + \frac{1}{2}\log|\Sigma_{u,\boldsymbol{\kappa}}| - \frac{1}{2}x_{\boldsymbol{\psi}}^\top \Sigma_{u,\boldsymbol{\psi}}^{-1} x_{\boldsymbol{\psi}} + \frac{1}{2}x_{\boldsymbol{\kappa}}^\top \Sigma_{u,\boldsymbol{\kappa}}^{-1} x_{\boldsymbol{\kappa}}\right\},$$

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. $\Sigma_{u,\boldsymbol{\psi}}$, $c_{j,\boldsymbol{\psi}}^{aa}, c_{j,\boldsymbol{\psi}}^{ab}, c_{j,\boldsymbol{\psi}}^{ba}$, and $c_{j,\boldsymbol{\psi}}^{bb}$ denote the covariance matrix and B-spline coefficients for the kernel $\boldsymbol{\psi}$, while $\Sigma_{u,\boldsymbol{\kappa}}$, $c_{j,\boldsymbol{\kappa}}^{aa}, c_{j,\boldsymbol{\kappa}}^{ab}, c_{j,\boldsymbol{\kappa}}^{ba}$, and $c_{j,\boldsymbol{\kappa}}^{bb}$ denote that for the kernel $\boldsymbol{\kappa}$.

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

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

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

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

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

Using Taylor expansion and the results from Hwang (1980) such that $\varphi''(t) \le C_3 J^2$, we have $\varphi(\frac{1}{J^2}) \le 1 - \frac{\delta}{C_4 J^2}$, 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

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

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 $\widehat{\kappa}$ obtained on Θ_{J_n} converges to the projected true set of kernel functions $\kappa_{0|\Theta_J}$. As $n, J_n \to \infty$, $\kappa_{0|\Theta_J} \to \kappa_0$ because each $\kappa_{xy,0|\Theta_J}$ in $\kappa_{0|\Theta_J}$ is just the B-spline truncation of the corresponding true kernel $\kappa_{xy,0}$ in κ_0 on Θ_{J_n}.

References

Aït-Sahalia, Y., Mykland, P. and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise, *Review of Financial Studies* **18**: 351–416.

Antoniadis, A. and Sapatinas, T. (2003). Wavelet methods for continuous-time prediction using hilbert-valued autoregressive processes, *Journal of Multivariate Analysis* **87**: 135–158.

Benston, G. and Hagerman, R. (1974). Determinants of bid-asked spreads in the over-the-counter market, *Journal of Financial Economics* **1**: 353–364.

Besse, P., Cardot, H. and Stephenson, D. (2000). Autoregressive forecasting of some functional climatic variations, *Scandinavian Journal of Statistics* **27**: 673–687.

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

Çetin, U., Jarrow, R. and Protter, P. (2004). Liquidity risk and arbitrage pricing theory, *Finance and Stochastics* **8**: 311–341.

Chen, Y. and Li, B. (2015). An adaptive functional autoregressive forecast model to predict electricity price curves, *Journal of Business and Economic Statistics*, (Accepted).

Chordia, T., Sarkar, A. and Subrahmanyam, A. (2003). An empirical analysis of stock and bond market liquidity, *Federal Reserve Bank of New York Staff Reports* **164**.

Cooper, K., Groth, J. and Avera, W. (1985). Liquidity, exchange listing, and common stock performance, *Journal of Economics and Business* **37**: 19–33.

Dierker, M., Kim, J.-W., Lee, J. and Morck, R. (2014). Interacting limit order demand and supply curves, *Alberta School of Business, working paper* .

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

Geman, S. and Hwang, C.-R. (1982). Nonparametric maximum likelihood estimation by the method of sieves, *Annals of Statistics* **10**: 401–414.

Gomber, P., Schweickert, U. and Theissen, E. (2015). Liquidity dynamics in an electronic open limit order book: An event study approach, *European Financial Management* **21**: 52–78.

Grenander, U. (1981). *Abstract inference*, Wiley, New York.

Groß-Klußmann, A. and Hautsch, N. (2013). Predicting bid ask spreads using long-memory autoregressive conditional Poisson models, *Journal of Forecasting* **32**(8): 724–742.

Guillas, S. (2001). Rates of convergence of autocorrelation estimates for autoregressive hilbertian processes, *Statistics & Probability Letters* **55**: 281–291.

Härdle, W., Hautsch, N. and Mihoci, A. (2012). Modelling and forecasting liquidity supply using semiparametric factor dynamics, *Journal of Empirical Finance* **19**: 610–625.

Härdle, W., Hautsch, N. and Mihoci, A. (2015). Local adaptive multiplicative error models for high-frequency forecasts, *Journal of Applied Econometrics* **30**(4): 529–550.

Huberman, G. and Halka, D. (2001). Systematic liquidity, *Journal of Financial Research* **24**: 161–178.

Hwang, C. (1980). Gaussian measure of large balls in a hilbert space, *Proceedings of the American Mathematical Society* **78**: 107–110.

Kim, M., Chaudhuri, K. and Shin, Y. (2015). Forecasting distribution of inflation rates: A functional autoregressive approach, *Journal of the Royal Statistical Society Series A, Statistics in Society,* (Early online publication).

Kokoszka, P. and Zhang, X. (2010). Improved estimation of the kernel of the functional autore-gressive process, *Technical report, Utah State University* .

Mourid, T. and Bensmain, N. (2006). Sieves estimator of the operator of a functional autoregressive process, *Statistics & Probability Letters* **76**: 93–108.

Stoll, H. (1978). The pricing of security dealer services: An empirical study of nasdaq stocks, *Journal of Finance* **33**: 1153–1172.

Zhang, L., M. P. and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data, *Journal of the American Statistical Association* **100**(472): 1394–1411.

SFB 649 Discussion Paper Series 2016

001 "Downside risk and stock returns: An empirical analysis of the long-run and short-run dynamics from the G-7 Countries" by Cathy Yi-Hsuan Chen, Thomas C. Chiang and Wolfgang Karl Härdle, January 2016.

002 "Uncertainty and Employment Dynamics in the Euro Area and the US" by Aleksei Netsunajev and Katharina Glass, January 2016.

003 "College Admissions with Entrance Exams: Centralized versus Decentralized" by Isa E. Hafalir, Rustamdjan Hakimov, Dorothea Kübler and Morimitsu Kurino, January 2016.

004 "Leveraged ETF options implied volatility paradox: a statistical study" by Wolfgang Karl Härdle, Sergey Nasekin and Zhiwu Hong, February 2016.

005 "The German Labor Market Miracle, 2003 -2015: An Assessment" by Michael C. Burda, February 2016.

006 "What Derives the Bond Portfolio Value-at-Risk: Information Roles of Macroeconomic and Financial Stress Factors" by Anthony H. Tu and Cathy Yi-Hsuan Chen, February 2016.

007 "Budget-neutral fiscal rules targeting inflation differentials" by Maren Brede, February 2016.

008 "Measuring the benefit from reducing income inequality in terms of GDP" by Simon Voigts, February 2016.

009 "Solving DSGE Portfolio Choice Models with Asymmetric Countries" by Grzegorz R. Dlugoszek, February 2016.

010 "No Role for the Hartz Reforms? Demand and Supply Factors in the German Labor Market, 1993-2014" by Michael C. Burda and Stefanie Seele, February 2016.

011 "Cognitive Load Increases Risk Aversion" by Holger Gerhardt, Guido P. Biele, Hauke R. Heekeren, and Harald Uhlig, March 2016.

012 "Neighborhood Effects in Wind Farm Performance: An Econometric Approach" by Matthias Ritter, Simone Pieralli and Martin Odening, March 2016.

013 "The importance of time-varying parameters in new Keynesian models with zero lower bound" by Julien Albertini and Hong Lan, March 2016.

014 "Aggregate Employment, Job Polarization and Inequalities: A Transatlantic Perspective" by Julien Albertini and Jean Olivier Hairault, March 2016.

015 "The Anchoring of Inflation Expectations in the Short and in the Long Run" by Dieter Nautz, Aleksei Netsunajev and Till Strohsal, March 2016.

016 "Irrational Exuberance and Herding in Financial Markets" by Christopher Boortz, March 2016.

017 "Calculating Joint Confidence Bands for Impulse Response Functions using Highest Density Regions" by Helmut Lütkepohl, Anna Staszewska-Bystrova and Peter Winker, March 2016.

018 "Factorisable Sparse Tail Event Curves with Expectiles" by Wolfgang K. Härdle, Chen Huang and Shih-Kang Chao, March 2016.

019 "International dynamics of inflation expectations" by Aleksei Netšunajev and Lars Winkelmann, May 2016.

020 "Academic Ranking Scales in Economics: Prediction and Imdputation" by Alona Zharova, Andrija Mihoci and Wolfgang Karl Härdle, May 2016.

SFB 649 Discussion Paper Series 2016

For a complete list of Discussion Papers published by the SFB 649,
please visit http://sfb649.wiwi.hu-berlin.de.