# Forecasting Limit Order Book Liquidity with Functional AutoRegressive Dynamics

Ying Chen

Wee Song Chua

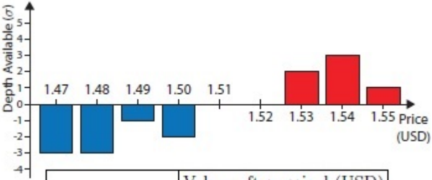
Wolfgang Karl Härdle

Department of Statistics and Applied Probability
National University of Singapore

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
and Economics
Humboldt-Universität zu Berlin

Sim Kee Boon Institute for Financial Economics
Singapore Management University

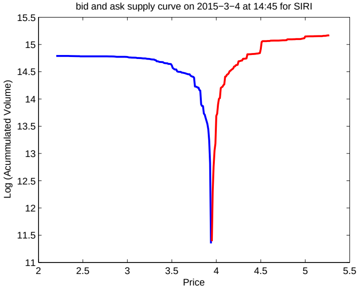
# Limit order book

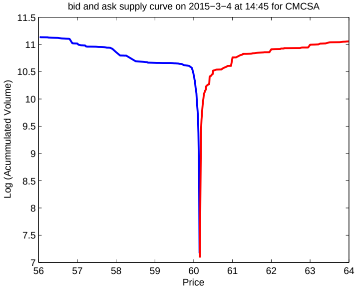



| Arriving order $x$ | Values after arrival (USD) | | | |
|---|---|---|---|---|
| | $b(t_x)$ | $a(t_x)$ | $m(t_x)$ | $s(t_x)$ |
| Initial Values | 1.50 | 1.53 | 1.515 | 0.03 |
| $(\$1.48, -3\sigma, t_x)$ | 1.50 | 1.53 | 1.515 | 0.03 |
| $(\$1.51, -3\sigma, t_x)$ | 1.51 | 1.53 | 1.52 | 0.02 |
| $(\$1.55, -3\sigma, t_x)$ | 1.50 | 1.54 | 1.52 | 0.04 |
| $(\$1.55, -5\sigma, t_x)$ | 1.50 | 1.55 | 1.525 | 0.05 |
| $(\$1.54, 4\sigma, t_x)$ | 1.50 | 1.53 | 1.515 | 0.03 |
| $(\$1.52, 4\sigma, t_x)$ | 1.50 | 1.52 | 1.51 | 0.02 |
| $(\$1.47, 4\sigma, t_x)$ | 1.48 | 1.53 | 1.505 | 0.05 |
| $(\$1.50, 4\sigma, t_x)$ | 1.49 | 1.50 | 1.495 | 0.01 |

# Liquidity demand and supply curve

SIRI and CMCSA bid $X_t^{(b)}$ and ask $X_t^{(a)}$ supply curve on March 4, 2015 at 14:45pm. Liquidity curves Data source




Bid ask curve

# Economic implications

- Liquidity demand and supply curves provide information on traders' expectations of the price
- Improve order execution strategy
- Smaller transaction cost
- Robust arbitrage pricing theory (Çetin, Jarrow and Protter, 2004)
- Forecasting with DSFM (Härdle, Hautsch and Mihoci, 2012)

# Liquidity measures

- Bid-ask spread (Benston and Hagerman, 1974; Stoll, 1978; Fleming and Remolona, 1999)
- Liquidity depth based on volumes at the best quotes or some particular price quotes, e.g. XLM (Cooper, Groth and Avera, 1985; Gomber, Schweickert and Theissen, 2015)

# Autoregressive models

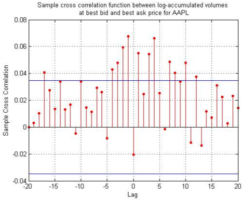
- Long memory autoregressive conditional Poisson model (Groß-Klußmann and Hautsch, 2013)
- Autoregressive model (Huberman and Halka, 2001)
- Vector autoregressive model (Chordia, Sarkar and Subrahmanyam, 2003)
- Local adaptive multiplicative error model (Härdle, Hautsch and Mihoci, 2015; Härdle, Mihoci and Ting, 2016)

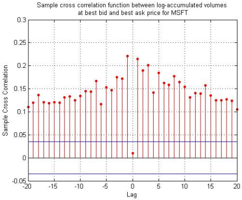
# Cross-correlations

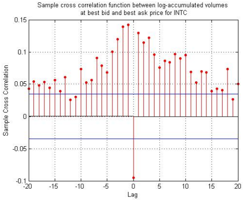
- Limit order demand and supply elasticities are cross-related (Dierker, Kim, Lee and Morck, 2014)
- Public or private information: switch sides
- Market-wide events: similar changes on both sides

# Cross-correlations

Sample cross correlation function between log-accumulated volumes at best bid and best ask price for AAPL, MSFT, and INTC.



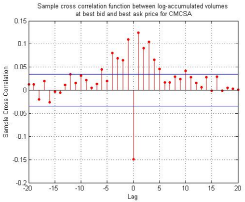



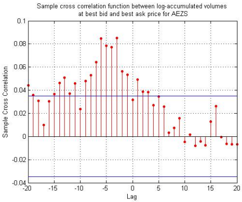



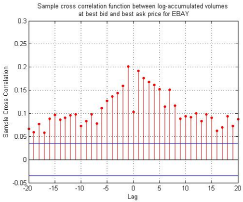

Cross Correlation

# Cross-correlations

Sample cross correlation function between log-accumulated volumes at best bid and best ask price for CMCSA, AEZS, and EBAY.




Cross Correlation

- Cross correlations suggest richer dynamics
- Richer dynamics of liquidity allows for more precise forecasts
- The bid/ask cross-dependency motivates to analyze two liquidity curves simultaneously

# Objectives

- Employ a Vector Functional AutoRegressive (VFAR) model:

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

  where $X_t^{(a)}$ and $X_t^{(b)}$ are the curves on the (b)id and (a)sk side, while $\rho$ is a bounded linear operator.

- Asymptotic consistency of the VFAR estimator
- Finite sample performance in real data analysis

# Outline

1. Motivation ✓
2. Data
3. Vector Functional AutoRegression (VFAR)
4. Empirical Applications
5. Conclusion

# Data

LOB records of 12 stocks traded in NASDAQ stock market from 2 Jan 2015 to 6 Mar 2015 (44 trading days)

- 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)
- Starbucks Corporation (SBUX)

CRC649 LOBSTER

# Summary statistics

5-minutes snapshots of the LOB data.

| Ticker Symbol | Spread (USD) | | Bid vol | | Ask vol | |
|---|---|---|---|---|---|---|
| | min | max | min | max | min | max |
| AAPL | 0.01 | 0.07 | 52,267 | 710,020 | 61,305 | 1,298,696 |
| MSFT | 0.01 | 0.02 | 90,344 | 928,319 | 122,377 | 621,471 |
| INTC | 0.01 | 0.02 | 158,900 | 557,251 | 146,959 | 1,142,641 |
| CSCO | 0.01 | 0.02 | 134,790 | 1,316,058 | 266,455 | 4,458,672 |
| SIRI | 0.01 | 0.02 | 1,266,528 | 3,725,304 | 3,002,680 | 7,605,467 |
| AMAT | 0.01 | 0.03 | 78,944 | 334,794 | 180,749 | 787,983 |
| CMCSA | 0.01 | 0.06 | 23,668 | 128,916 | 40,638 | 146,724 |
| AEZS | 0.0001 | 0.05 | 145,635 | 767,785 | 472,689 | 1,158,740 |
| EBAY | 0.01 | 0.04 | 42,060 | 160,572 | 52,813 | 415,033 |
| MU | 0.01 | 0.04 | 95,907 | 497,910 | 102,357 | 595,200 |
| WFM | 0.01 | 0.12 | 34,538 | 114,386 | 41,019 | 159,488 |
| SBUX | 0.01 | 0.12 | 27,467 | 151,022 | 34,914 | 166,932 |

# Data pre-processing

- Sampling frequency: 5 minutes (Aït-Sahalia, Mykland and Zhang, 2005; Zhang and Aït-Sahalia, 2005)
- Discarded the first 15 minutes and the last 5 minutes (Härdle et al., 2012)
- Log-transformed the accumulated volumes (Potters and Bouchaud, 2003)
- Obtain 75 pairs of bid and ask liquidity curves, for each stock, at each trading day
- Total: 3300 pairs of bid and ask supply curves over the whole sample period of 44 trading days for each stock

# Vector Functional AutoRegression (VFAR)

$X_t^{(a)}(\tau),\ X_t^{(b)}(\tau) \in C_{(-\infty,\infty)}$. ▸ $X_t^{(a)}$

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

where $(\mu_a, \mu_b)^\top$ are the mean functions. The operators $\rho$ is bounded linear operator from $\mathcal{H}$ to $\mathcal{H}$. 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
$0 < \mathsf{E}\,\|\varepsilon_1^{(a)}\|^2 = \cdots = \mathsf{E}\,\|\varepsilon_n^{(a)}\|^2 < \infty$ and
$0 < \mathsf{E}\,\|\varepsilon_1^{(b)}\|^2 = \cdots = \mathsf{E}\,\|\varepsilon_n^{(b)}\|^2 < \infty$, $\varepsilon_t^{(a)}$ and $\varepsilon_t^{(b)}$ need not be cross-independent.

# Convolution kernel operator

The operators $\rho$ in (1):

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

where the kernel function $\kappa_{xy} \in L^2((-\infty, \infty))$ and $\|\kappa_{xy}\|_2 < 1$ for $xy = aa$, $ab$, $ba$, and $bb$.

# $B$-spline expansion

Expand the functions in (2) using $B$-spline basis function in $L^2((-\infty, \infty))$:

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

where $m$ is the order, and $w_1 \leq \cdots \leq w_{J+m}$ denotes the knot sequence. Note that

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

$B$-spline expansions

# Sieve

Introduce a sequence of subsets $\{\Theta_{J_n}\}$ - a sieve, where $\Theta_{J_n} \subseteq \Theta_{J_n+1}$ and the union of subsets $\bigcup \Theta_{J_n}$ is dense in the parameter space (Grenander, 1981).

The sieve is defined as follows:

$$\Theta_{J_n} = \left\{ \kappa_{xy} \in L^2 \mid \kappa_{xy}(\tau) = \sum_{l=1}^{J_n} c_l^{xy} B_{l,m}(\tau), \tau \in (-\infty, \infty), \sum_{l=1}^{J_n} l^2 (c_l^{xy})^2 \leq v J_n \right\} \tag{3}$$

VFAR & $B-$Splines

# Coefficients relationship

This provides the following relationship of the $B$-spline coefficients:

$$
\begin{aligned}
d^a_{t,h} &= p^a_h + \sum_{i=1}^{J_n} \left\{ \sum_{j=1}^{J_n} \left( \frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c^{aa}_j - c^{aa}_h \right\} \frac{w_{i+m} - w_i}{m} d^a_{t-1,i} \\
&+ \sum_{i=1}^{J_n} \left\{ \sum_{j=1}^{J_n} \left( \frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c^{ab}_j - c^{ab}_h \right\} \frac{w_{i+m} - w_i}{m} d^b_{t-1,i} + d^a_h(\varepsilon^{(a)}_t) \\
d^b_{t,h} &= p^b_h + \sum_{i=1}^{J_n} \left\{ \sum_{j=1}^{J_n} \left( \frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c^{bb}_j - c^{bb}_h \right\} \frac{w_{i+m} - w_i}{m} d^b_{t-1,i} \\
&+ \sum_{i=1}^{J_n} \left\{ \sum_{j=1}^{J_n} \left( \frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c^{ba}_j - c^{ba}_h \right\} \frac{w_{i+m} - w_i}{m} d^a_{t-1,i} + d^b_h(\varepsilon^{(b)}_t)
\end{aligned}
\tag{4}
$$

# Vector AutoRegressive form

Rewriting (4) as a matrix yields a form of Vector AutoRegressive (VAR) of order 1:

$$
\begin{bmatrix} d^a_{t,1} \\ \vdots \\ d^a_{t,J_n} \\ d^b_{t,1} \\ \vdots \\ d^b_{t,J_n} \end{bmatrix}
=
\begin{bmatrix}
p^a_1 & r^{aa}_{1,1} & \cdots & r^{aa}_{1,J_n} & r^{ab}_{1,1} & \cdots & r^{ab}_{1,J_n} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
p^a_{J_n} & r^{aa}_{J_n,1} & \cdots & r^{aa}_{J_n,J_n} & r^{ab}_{J_n,1} & \cdots & r^{ab}_{J_n,J_n} \\
p^b_1 & r^{ba}_{1,1} & \cdots & r^{ba}_{1,J_n} & r^{bb}_{1,1} & \cdots & r^{bb}_{1,J_n} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
p^b_{J_n} & r^{ba}_{J_n,1} & \cdots & r^{ba}_{J_n,J_n} & r^{bb}_{J_n,1} & \cdots & r^{bb}_{J_n,J_n}
\end{bmatrix}
\begin{bmatrix} 1 \\ d^a_{t-1,1} \\ \vdots \\ d^a_{t-1,J_n} \\ d^b_{t-1,1} \\ \vdots \\ d^b_{t-1,J_n} \end{bmatrix}
+
\begin{bmatrix} d^a_1(\varepsilon^{(a)}_t) \\ \vdots \\ d^a_{J_n}(\varepsilon^{(a)}_t) \\ d^b_1(\varepsilon^{(b)}_t) \\ \vdots \\ d^b_{J_n}(\varepsilon^{(b)}_t) \end{bmatrix}
\tag{5}
$$

where $r^{xy}_{h,i}$ denotes $\left\{ \sum_{j=1}^{J_n} \left( \frac{w_{j+m}-w_{j+1}}{w_{j+m}-w_j} - \frac{w_{j+m+1}-w_{j+2}}{w_{j+m+1}-w_{j+1}} \right) c^{xy}_j - c^{xy}_h \right\} \frac{w_{i+m}-w_i}{m}$, for $xy = aa$, $ab$, $ba$, and $bb$.

# Vector AutoRegressive form

Write compactly as the following:

$$Y = BZ + U \tag{6}$$

Assuming

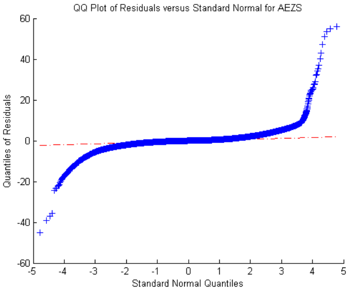
$$\mathbf{u} = \mathrm{vec}(U) = \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix} \sim \mathrm{N}(0, I_T \otimes \Sigma_u),$$

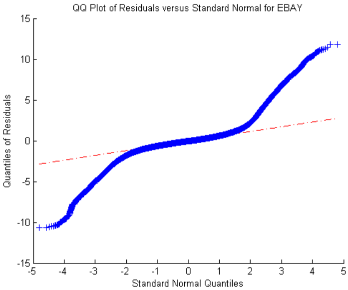
with pdf of $\mathbf{u}$:

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

# Residual analysis

## QQ plots of residuals for AEZS, EBAY




QQ plot

# Maximum likelihood estimation

Using $\mathbf{u} = \mathbf{y} - (Z^\top \otimes I_K)\beta$, the likelihood function is:

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

$$= \frac{1}{(2\pi)^{KT/2}} \left| I_T \otimes \Sigma_u \right|^{-\frac{1}{2}} \exp\left[ -\frac{1}{2}\left(\mathbf{y} - (Z^\top \otimes I_K)\beta\right)^\top (I_T \otimes \Sigma_u^{-1}) \left\{\mathbf{y} - (Z^\top \otimes I_K)\beta\right\}\right].$$

Log-likelihood

# Maximum likelihood estimation

$$
\begin{aligned}
\widehat{B} &= YZ^\top (ZZ^\top)^{-1} \\
\widehat{\Sigma_u} &= \frac{1}{T}(Y - BZ)(Y - BZ)^\top
\end{aligned}
\tag{7}
$$

The first column of $YZ^\top(ZZ^\top)^{-1}$ in (7) is the estimator for $v = \left(p_1^a, \cdots, p_{J_n}^a, p_1^b, \cdots, p_{J_n}^b\right)^\top$.

# Maximum likelihood estimation

To show the estimator for $c_j^{xy}$ for $xy = aa, ab, ba, bb$ as in (2), first define the following notations:

$$
\begin{aligned}
W &= \mathrm{diag}\Big(\frac{m}{w_{1+m} - w_1}, \cdots, \frac{m}{w_{J_n+m} - w_{J_n}}, \\
&\qquad\qquad \frac{m}{w_{1+m} - w_1}, \cdots, \frac{m}{w_{J_n+m} - w_{J_n}}\Big), \\
q_j &= \frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}}, \\
\theta_1 &= (c_1^{aa}, \cdots, c_{J_n}^{aa}, c_1^{ba}, \cdots, c_{J_n}^{ba})^\top, \\
\theta_2 &= (c_1^{ab}, \cdots, c_{J_n}^{ab}, c_1^{bb}, \cdots, c_{J_n}^{bb})^\top, \\
\boldsymbol{\theta} &= (\theta_1, \cdots, \theta_1, \theta_2, \cdots, \theta_2),
\end{aligned}
$$

# Maximum likelihood estimation

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

where $\boldsymbol{\theta}$ contains $J_n$ columns of $\theta_1$ and $J_n$ columns of $\theta_2$.

Therefore the estimator for $c_j^{xy}$ for $xy = aa, ab, ba, bb$ is:

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

# Consistency result

## Theorem (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* $\Theta_{J_n}$, $J_n = 1, 2, \cdots$ *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* $\Theta_{J_n}$, *where*

$$\varphi_{J_n} = \sup_k \inf_{t \geq 0} E_{\boldsymbol{\rho}_{0|\Theta_{J_n}}} \exp\left\{ t \log \frac{g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \Gamma_k)}{g(X_t^{(a)}, X_t^{(b)}, X_{t-1}^{(a)}, X_{t-1}^{(b)}, \boldsymbol{\rho}_{J_n})} \right\}.$$

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

C1 & C2

# Consistency result

### Theorem (2)

*If $J_n = \mathcal{O}(n^{1/3-\eta})$ for $\eta > 0$, then $\|\hat{\boldsymbol{\kappa}}_{J_n} - \boldsymbol{\kappa}_{0|\Theta_{J_n}}\|_2 \to 0$ a.s. when $n \to +\infty$ and $\|\cdot\|_2$ is the $L^2$ norm in $C_{[0,1]}$.*

*$\hat{\boldsymbol{\kappa}}_{J_n} = (\hat{\kappa}_{aa,J_n}, \hat{\kappa}_{ab,J_n}, \hat{\kappa}_{ba,J_n}, \hat{\kappa}_{bb,J_n})$ is the set of sieve estimators on $\Theta_{J_n}$ 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 $\Theta_{J_n}$.*

# Real data analysis

- Liquidity demand and supply curves over 44 trading days from date 2 Jan 2015 to 6 Mar 2015
- Cubic $B$-spline expansions with equally-spaced price percentile as nodes and $J_n = 20$ in the sieve
- 20 coefficients for the bid and another 20 for the ask liquidity curve

$R^2$ with different $J_n$

# Evaluation

The root mean squared estimation error (RMSE), mean absolute percentage error (MAPE), and $R^2$ are computed:

$$
\begin{aligned}
RMSE &= \sqrt{\frac{\sum_{xy=a,b}\sum_{t=1}^{T}\sum_{\tau}\left\{X_t^{(xy)}(\tau)-\hat{X}_t^{(xy)}(\tau)\right\}^2}{N}} \\
MAPE &= \frac{\sum_{xy=a,b}\sum_{t=1}^{T}\sum_{\tau}\frac{\left|X_t^{(xy)}(\tau)-\hat{X}_t^{(xy)}(\tau)\right|}{X_t^{(xy)}(\tau)}}{N} \\
R^2 &= 1-\frac{\sum_{xy=a,b}\sum_{t=1}^{T}\sum_{\tau}\left\{X_t^{(xy)}(\tau)-\hat{X}_t^{(xy)}(\tau)\right\}^2}{\sum_{xy=a,b}\sum_{t=1}^{T}\sum_{\tau}\left\{X_t^{(xy)}(\tau)-\bar{X}\right\}^2}
\end{aligned}
\tag{8}
$$

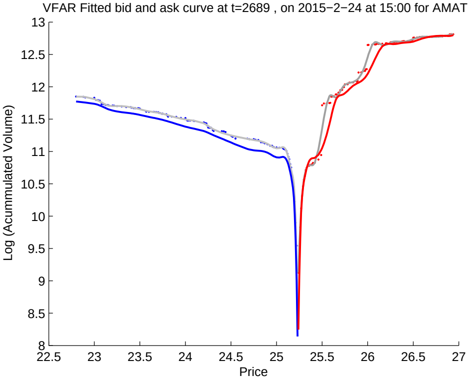
# VFAR fits

## $R^2$, RMSE, and MAPE for in-sample estimation of the nine stocks

| Ticker Symbol | $R^2$ | | RMSE | | MAPE | |
|---|---|---|---|---|---|---|
| | VFAR | Naive | VFAR | Naive | VFAR | Naive |
| AAPL | **92.03%** | 88.87% | **0.3419** | 0.4041 | **3.61%** | 3.80% |
| MSFT | **95.19%** | 93.53% | **0.1831** | 0.2124 | **0.95%** | 1.02% |
| INTC | **94.79%** | 93.13% | **0.1868** | 0.2144 | **0.92%** | 0.98% |
| CSCO | **96.16%** | 95.08% | **0.1926** | 0.2180 | **0.86%** | 0.91% |
| SIRI | **98.29%** | 97.96% | **0.0852** | 0.0931 | **0.29%** | **0.29%** |
| AMAT | **95.83%** | 94.46% | **0.1768** | 0.2040 | **0.89%** | 0.97% |
| CMCSA | **93.39%** | 90.79% | **0.1851** | 0.2185 | **1.20%** | 1.35% |
| AEZS | **98.48%** | 96.82% | **0.4247** | 0.6148 | **2.18%** | 2.28% |
| EBAY | **94.88%** | 93.20% | **0.2250** | 0.2591 | **1.55%** | 1.64% |
| MU | **95.14%** | 93.45% | **0.2157** | 0.2504 | **1.17%** | 1.26% |
| WFM | **95.52%** | 94.00% | **0.2023** | 0.2339 | **1.57%** | 1.59% |
| SBUX | **94.77%** | 92.81% | **0.2241** | 0.2627 | **2.51%** | 2.63% |

# VFAR fits

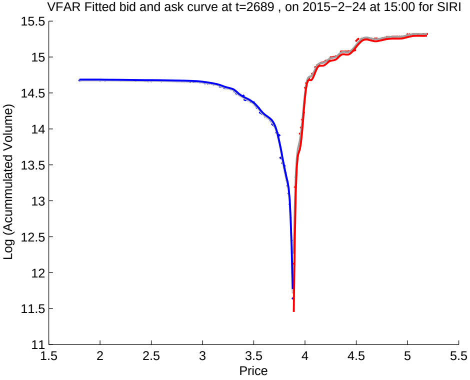
AMAT estimated bid (and ask) supply curves vs. the actually observed ones on 24 February 2015 at 3p.m. ▸ VFAR fits




Estimated VFAR curve

# VFAR fits

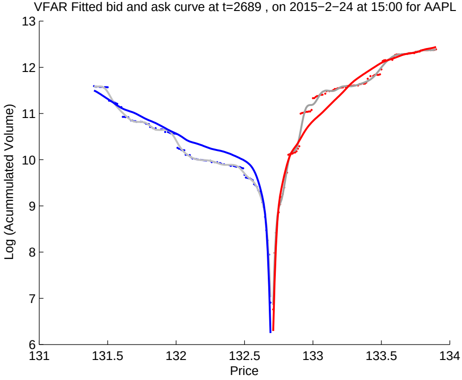
SIRI estimated bid (and ask) supply curves vs. the actually observed ones on 24 February 2015 at 3p.m.




Estimated VFAR curve

# VFAR fits

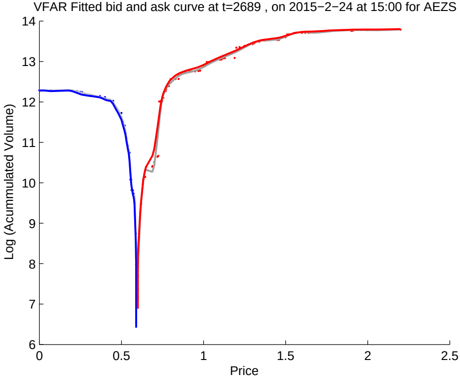
AAPL estimated bid (and ask) supply curves vs. the actually observed ones on 24 February 2015 at 3p.m.




Estimated VFAR curve

# VFAR fits

AEZS estimated bid (and ask) supply curves vs. the actually observed ones on 24 February 2015 at 3p.m.




Estimated VFAR curve

# Forecast

- 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
- Move forward one period, i.e. 5 minutes at a time, and re-estimate and forecast until the last time point at $t = 3300$

# Comparison with Naive Forecast

## RMSE for multistep ahead VFAR and naive forecast of the nine stocks

| Ticker Symbol | RMSE | | | | | |
|---|---|---|---|---|---|---|
| | 1-step | | 5-steps | | 10-steps | |
| | VFAR | Naive | VFAR | Naive | VFAR | Naive |
| AAPL | **0.3742** | 0.4318 | **0.4757** | 0.5678 | **0.5063** | 0.6193 |
| MSFT | **0.1809** | 0.1986 | **0.2444** | 0.2744 | **0.2707** | 0.3082 |
| INTC | **0.1927** | 0.2156 | **0.2559** | 0.2974 | **0.2845** | 0.3377 |
| CSCO | **0.2124** | 0.2317 | **0.3063** | 0.3447 | **0.3614** | 0.4111 |
| SIRI | **0.0894** | 0.0932 | **0.1271** | 0.1404 | **0.1416** | 0.1543 |
| AMAT | **0.1928** | 0.2141 | **0.2533** | 0.2940 | **0.2863** | 0.3374 |
| CMCSA | **0.1959** | 0.2240 | **0.2392** | 0.2952 | **0.2512** | 0.3208 |
| AEZS | **0.4814** | 0.6525 | **0.5483** | 0.7699 | **0.5812** | 0.8075 |
| EBAY | **0.2335** | 0.2620 | **0.2931** | 0.3477 | **0.3228** | 0.3961 |
| MU | **0.2180** | 0.2507 | **0.2797** | 0.3355 | **0.3094** | 0.3787 |
| WFM | **0.2035** | 0.2303 | **0.2575** | 0.3167 | **0.2746** | 0.3529 |
| SBUX | **0.2285** | 0.2623 | **0.2782** | 0.3366 | **0.2975** | 0.3687 |

# Comparison with Naive Forecast

## MAPE for multistep ahead VFAR and naive forecast of the nine stocks

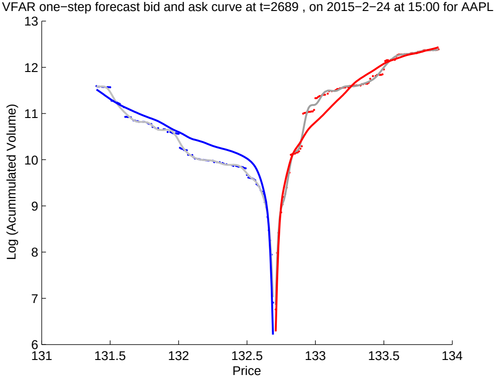
| Ticker Symbol | MAPE | | | | | |
|---|---|---|---|---|---|---|
| | 1-step | | 5-steps | | 10-steps | |
| | VFAR | Naive | VFAR | Naive | VFAR | Naive |
| AAPL | **3.61%** | 3.73% | **4.21%** | 4.68% | **4.49%** | 5.09% |
| MSFT | 0.93% | **0.90%** | 1.42% | **1.39%** | **1.63%** | 1.67% |
| INTC | **0.95%** | 0.97% | **1.46%** | 1.53% | **1.74%** | 1.86% |
| CSCO | **0.88%** | 0.89% | **1.44%** | 1.46% | **1.76%** | 1.85% |
| SIRI | 0.30% | **0.27%** | 0.52% | **0.47%** | 0.62% | **0.56%** |
| AMAT | **1.00%** | 1.01% | **1.47%** | 1.55% | **1.72%** | 1.87% |
| CMCSA | **1.15%** | 1.22% | **1.56%** | 1.76% | **1.69%** | 2.01% |
| AEZS | **2.22%** | 2.32% | **2.85%** | 3.20% | **3.14%** | 3.51% |
| EBAY | **1.31%** | 1.38% | **1.79%** | 1.99% | **2.03%** | 2.36% |
| MU | **1.18%** | 1.27% | **1.70%** | 1.90% | **1.99%** | 2.29% |
| WFM | **1.25%** | 1.31% | **1.76%** | 1.97% | **1.94%** | 2.32% |
| SBUX | **1.81%** | 1.87% | **2.27%** | 2.52% | **2.48%** | 2.86% |

# Findings

- Like the VFAR forecasts, the naive forecasts also have RMSE and MAPE increasing as one forecasts more steps into the future
- All VFAR forecasts outperform the naive forecasts in terms of RMSE
- Only in 5 (out of 36) instances, the naive forecasts perform better than VFAR forecasts in terms of MAPE

# VFAR forecasts

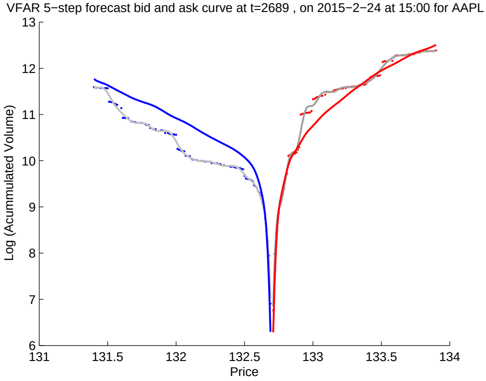
## AAPL 5−minute ahead forecasted bid (and ask) supply curves vs. the actually observed ones for 24 February 2015 at 3p.m. ▸ 1−step




Forecasted VFAR curve

# VFAR forecasts

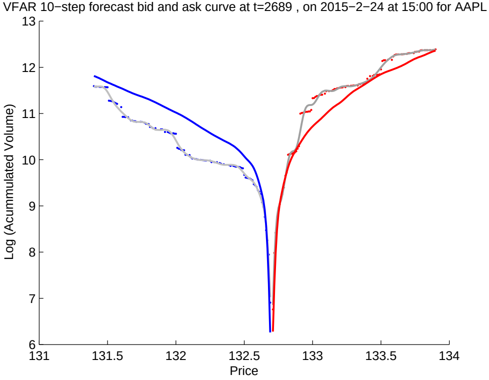
AAPL 25−minute ahead forecasted bid (and ask) supply curves vs. the actually observed ones for 24 February 2015 at 3p.m. ▸ 5−steps




Forecasted VFAR curve

# VFAR forecasts

AAPL 50−minute ahead forecasted bid (and ask) supply curves vs. the actually observed ones for 24 February 2015 at 3p.m. ▸ 10−steps




Forecasted VFAR curve

# Conclusion

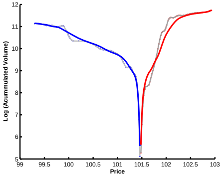
- Proposed the VFAR(1) modeling.
- Developed consistent ML estimators for VFAR(1) model with closed forms.
- In real data analysis, VFAR approach is more successful as compared to the naive model.
- Demand and supply curves are modelled and forecasted successfully.

# Forecasting Limit Order Book Liquidity with Functional AutoRegressive Dynamics

Ying Chen

Wee Song Chua

Wolfgang Karl Härdle




Department of Statistics and Applied Probability
National University of Singapore

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
and Economics
Humboldt-Universität zu Berlin

Sim Kee Boon Institute for Financial Economics
Singapore Management University

# References


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

Antoniadis, A. and Sapatinas, T.
Wavelet methods for continuous-time prediction using hilbert-valued autoregressive processes.
*Journal of Multivariate Analysis*, 87, 135-185, 2003.

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

# References


Besse, P., Cardot, H. and Stephenson, D.
Autoregressive forecasting of some climatic variations.
*Scandinavian Journal of Statistics*, 27, 673-687, 2000.

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

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

# References


Chen, Y. and Li, B.
An Adaptive Functional Autoregressive Forecast Model to Predict Electricity Price Curves.
*Journal of Business and Economic Statistics*, Accepted,2015.

Chordia, T., Sarkar, A. and Subrahmanyam, A.
An Empirical Analysis of Stock and Bond Market Liquidity.
*Federal Reserve Bank of New York Staff Reports*, 164,2003.

Cooper, K., Groth, J.C. and Avers, W.E.
Liquidity, exchange listing, and common stock performance.
*Journal of Economics and Business*, 37, 19-33, 1985.

# References


Fleming, M.J. and Remolona, E.M.
Price Formation and Liquidity in the U.S. Treasury Market: The Response to Public Information.
*Journal of Finance*, 54, 1901-1915, 1999.

Frank, J., Härdle, W.K., and Hafner, C.
Statistics of Financial Markets: an Introduction. 4th ed.
*Springer Verlag, Berlin Heidelberg*, 2015.

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

# References


Gomber, P., Schweickert, U. and Theissen, E.
Liquidity Dynamics in an Electronic Open Limit Order Book: An Event Study Approach.
*European Financial Management*, 21, 52-78, 2015.

Grenander, U.
Abstract Inference.
*Wiley, New York*, 1981.

Groß-Klußmann, A. and Hautsch, N.
Predicting Bid-Ask Spreads Using Long-Memory Autoregressive Conditional Poisson Models.
*Journal of Forecasting*, 32(8), 724-742, 2013.

# References


Guillas, S.
Rates of convergence of autocorrelation estimates for autoregressive Hilbertian processes.
*Statistics & Probability Letters*, 55(3), 281-291, 2001.

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

Härdle, W.K., Mihoci, A. and Ting, C.
Adaptive order flow forecasting with multiplicative error models.
*Quantitative Finance*, Submitted, 2016.

# References


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

Huberman, G., and Halka, D.
Systematic Liquidity.
*Journal of Financial Research*, 24, 161-178, 2001.

Kim, M., Chaudhuri, K., and Shin, Y.
Forecasting Distribution of Inflation Rates: A Functional Autoregressive Approach.
*Journal of the Royal Statistical Society Series A, Statistics in Society*, (Early online publication), 2015.

# References


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

Mourid, T., and Bensmain, N.
Sieves estimator of the operator of a functional autoregressive process.
*Statistics & Probability Letters*, 76(1), 93-108, 2006.

Potters, M. and Bouchaud, J.P.
More statistical properties of order books and price impact.
*Physica A*, 324, 133-140, 2003.

# References


Stoll, H.
The pricing of security dealer services: An empirical study of NASDAQ stocks.
*Journal of Finance*, 33, 1153-1172, 1978.

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

# $B$-spline expansion

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

$$\epsilon_t^{(a)}(\tau) = \sum_{j=1}^{\infty} d_{j}^{a}(\epsilon_t^{(a)}) B_{j,m}(\tau), \qquad \epsilon_t^{(b)}(\tau) = \sum_{j=1}^{\infty} d_{j}^{b}(\epsilon_t^{(b)}) B_{j,m}(\tau),$$

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

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

back

# VFAR & $B-$Splines

$$
\begin{aligned}
X_t^{(a)}(\tau) &= \sum_{j=1}^{J_n} d_{t,j}^a B_{j,m}(\tau) \\
&= \int_0^1 \left\{ \sum_{j=1}^{J_n} \sum_{i=1}^{J_n} c_j^{aa} d_{t-1,i}^a B_{j,m}(\tau - s) B_{i,m}(s) \right\} ds \\
&+ \int_0^1 \left\{ \sum_{j=1}^{J_n} \sum_{i=1}^{J_n} c_j^{ab} d_{t-1,i}^b B_{j,m}(\tau - s) B_{i,m}(s) \right\} ds + \sum_{j=1}^{J_n} d_j^a (\varepsilon_t^{(a)}) B_{j,m}(\tau) \\
&= \sum_{h=1}^{J_n} \sum_{i=1}^{J_n} \left\{ \sum_{j=1}^{J_n} \left( \frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{aa} - c_h^{aa} \right\} \frac{w_{i+m} - w_i}{m} d_{t-1,i}^a B_{h,m}(\tau) \\
&+ \sum_{h=1}^{J_n} \sum_{i=1}^{J_n} \left\{ \sum_{j=1}^{J_n} \left( \frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_j^{ab} - c_h^{ab} \right\} \frac{w_{i+m} - w_i}{m} d_{t-1,i}^b B_{h,m}(\tau) \\
&\quad + \sum_{j=1}^{J_n} d_j^a (\varepsilon_t^{(a)}) B_{j,m}(\tau)
\end{aligned} \tag{9}
$$

# Maximum likelihood estimation

Log-likelihood function:

$$
\begin{aligned}
&\ell\left(X_1^{(a)},\cdots,X_T^{(a)},X_1^{(b)},\cdots,X_T^{(b)};\rho^{aa},\rho^{ab},\rho^{ba},\rho^{bb}\right)\\
&=-\frac{KT}{2}\log 2\pi-\frac{T}{2}\log\left|\Sigma_u\right|-\frac{1}{2}\left(\mathbf{y}-(Z^\top\otimes I_K)\beta\right)^\top(I_T\otimes\Sigma_u^{-1})\left\{\mathbf{y}-(Z^\top\otimes I_K)\beta\right\}\\
&=-\frac{KT}{2}\log 2\pi-\frac{T}{2}\log\left|\Sigma_u\right|-\frac{1}{2}\sum_{t=1}^{T}\left(y_t-v-Cy_{t-1}\right)^\top\Sigma_u^{-1}\left(y_t-v-Cy_{t-1}\right)\\
&=-\frac{KT}{2}\log 2\pi-\frac{T}{2}\log\left|\Sigma_u\right|-\frac{1}{2}\sum_{t=1}^{T}\left(y_t-Cy_{t-1}\right)^\top\Sigma_u^{-1}\left(y_t-Cy_{t-1}\right)\\
&\qquad\qquad+v^\top\Sigma_u^{-1}\sum_{t=1}^{T}\left(y_t-Cy_{t-1}\right)-\frac{T}{2}v^\top\Sigma_u^{-1}v\\
&=-\frac{KT}{2}\log 2\pi-\frac{T}{2}\log\left|\Sigma_u\right|-\frac{1}{2}\operatorname{Tr}\left\{(Y-BZ)^\top\Sigma_u^{-1}(Y-BZ)\right\}
\end{aligned}
$$

back

# Consistency result

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 $\boldsymbol{\psi}$:

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

We consider the following 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 $\left\| \boldsymbol{\rho}_{J_n} - \boldsymbol{\rho}_{0|\Theta_{J_n}} \right\|_{HS} \to 0$. Here $\boldsymbol{\rho}_{0|\Theta_{J_n}}$ denotes the projection of the set of true operators $\boldsymbol{\rho}_0$ on the sieve $\Theta_{J_n}$.

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

back

# $R^2$ with different values of $J_n$

| $J_n$ | Ticker Symbol | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AAPL | MSFT | INTC | CSCO | SIRI | AMAT | CMCSA | AEZS | EBAY | MU | WFM | SBUX |
| 6 | 89.55% | 91.14% | 89.74% | 93.08% | 95.60% | 89.87% | 83.96% | 98.29% | 87.97% | 90.46% | 88.78% | 88.17% |
| 7 | 90.62% | 92.74% | 91.70% | 94.14% | 96.61% | 92.55% | 88.93% | 98.46% | 91.09% | 92.26% | 91.47% | 91.12% |
| 8 | 91.21% | 93.60% | 92.83% | 94.79% | 97.15% | 93.98% | 90.93% | 98.57% | 92.69% | 93.33% | 93.03% | 92.75% |
| 9 | 91.48% | 94.05% | 93.39% | 95.15% | 97.45% | 94.59% | 91.67% | 98.61% | 93.50% | 93.92% | 93.93% | 93.50% |
| 10 | 91.68% | 94.47% | 93.90% | 95.49% | 97.69% | 95.06% | 92.27% | **98.64%** | 94.12% | 94.41% | 94.62% | 94.07% |
| 11 | 91.78% | 94.72% | 94.21% | 95.72% | 97.86% | 95.34% | 92.66% | 98.64% | 94.46% | 94.67% | 94.99% | 94.35% |
| 12 | 91.84% | 94.86% | 94.38% | 95.84% | 97.95% | 95.47% | 92.85% | 98.64% | 94.59% | 94.79% | 95.13% | 94.45% |
| 13 | 91.88% | 94.96% | 94.51% | 95.93% | 98.03% | 95.58% | 92.99% | 98.61% | 94.68% | 94.89% | 95.24% | 94.53% |
| 14 | 91.91% | 95.05% | 94.60% | 96.00% | 98.10% | 95.67% | 93.12% | 98.58% | 94.74% | 94.96% | 95.32% | 94.58% |
| 15 | 91.94% | 95.10% | 94.67% | 96.06% | 98.16% | 95.74% | 93.23% | 98.53% | 94.78% | 95.02% | 95.37% | 94.64% |
| 16 | 91.95% | 95.12% | 94.69% | 96.08% | 98.18% | 95.74% | 93.24% | 98.45% | 94.79% | 95.04% | 95.40% | 94.67% |
| 17 | 91.99% | 95.16% | 94.73% | 96.12% | 98.22% | 95.78% | 93.31% | 98.43% | 94.82% | 95.08% | 95.45% | 94.70% |
| 18 | 92.00% | 95.16% | 94.75% | 96.13% | 98.24% | 95.80% | 93.33% | 98.40% | 94.84% | 95.10% | 95.46% | 94.72% |
| 19 | 92.02% | 95.19% | 94.77% | 96.15% | 98.27% | **95.84%** | 93.38% | 98.40% | 94.86% | 95.13% | 95.50% | 94.75% |
| 20 | **92.03%** | **95.19%** | **94.79%** | **96.16%** | **98.29%** | 95.83% | **93.39%** | 98.48% | **94.88%** | **95.14%** | **95.52%** | **94.77%** |

back