Bayesian Machine Learning in Finance

Patricia(Ning) Ning

Computational Finance & Risk Management
Dept. of Applied Math
Univ. of Washington, Seattle
ningnin@uw.edu

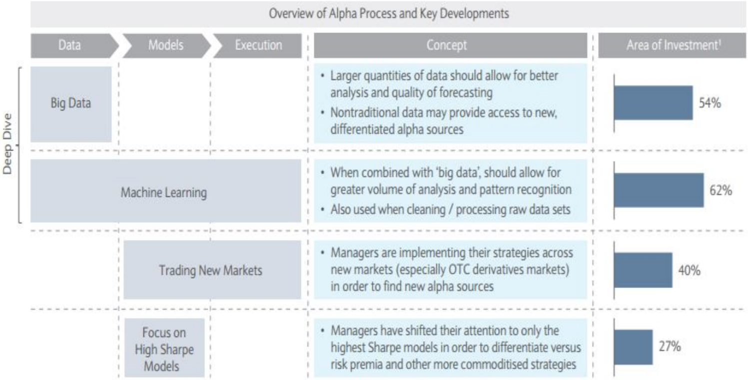
Talk maily based on paper **Multivariate Bayesian Structural Time Series Model**

Forthcoming **Journal of Machine Learning Research**

AI and Big Data Jobs in Hedge Funds

AI and Big Data Jobs in Hedge Funds

FIGURE 15: Current Investment Areas by Systematic Managers



Source: Strategic Consulting survey results and analysis

1. Percentage of managers within sample that are active or investing in these areas

Bayesian Structural Time Series (BSTS) Model

Figure: Steven L. Scott

Figure: Hal R Varian

uc berkeley

school of information management & systems

Biography of Hal R. Varian

Hal R. Varian is the Chief Economist at Google. He started in May 2002 as a consultant and has been involved in many aspects of the company, including auction design, econometric analysis, finance, corporate strategy and public policy.

He is also an emeritus professor at the University of California, Berkeley in three departments: business, economics, and information management.

He received his SB degree from MIT in 1969 and his MA in mathematics and Ph.D. in economics from UC Berkeley in 1973. He has also taught at MIT, Stanford, Oxford, Michigan and other universities around the world.

Dr. Varian is a fellow of the Guggenheim Foundation, the Econometric Society, and the American Academy of Arts and Sciences. He was Co-Editor of the American Economic Review from 1987
1990 and holds honorary doctorates from the University of Oulu, Finland and the University of Karlsruhe, Germany.

Professor Varian has published numerous papers in economic theory, industrial organization, financial economics, econometrics and information economics. He is the author of two major economics textbooks which have been translated into 22 languages. He is the co-author of a bestselling book on business strategy, *Information Rules: A Strategic Guide to the Network Economy* and wrote a monthly column for the *New York Times* from 2000 to 2007.

- Curriculum vitae (PDF)
- Fifty-word biography
- Color photo (100 dpi)
- Color photo (from Google)
- Color photo (high res, from Google)
- Small drawing
- Color photo (300 dpi)
- Color photo (600 dpi)

Bayesian Structural Time Series (BSTS) Model

BSTS model is a machine learning technique used for feature selection, time series forecasting, nowcasting, inferring causal impact and other.
The model consists of three main parts:

1. **Kalman filter:** The technique for time series decomposition. In this step, a researcher can add different state variables: trend, seasonality, regression, and others.
2. **Spike-and-slab method:** In this step, the most important regression predictors are selected.
3. **Bayesian model averaging:** Combining the results and prediction calculation.

Multivariate Bayesian Structural Time Series (MBSTS) Model

Structural Time Series Models belong to state space models

$$\text{Observation Equation: } \tilde{y}_t = Z_t^T \alpha_t + \tilde{\epsilon}_t, \qquad \tilde{\epsilon}_t \sim N_m(0, \Sigma_t),$$

$\tilde{y}_t$: observations, α_t: unobserved latent states

$$\text{Transition Equation: } \alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \qquad \eta_t \sim N_q(0, Q_t),$$

The model matrices Z_t, T_t, and R_t typically contain unknown parameters.

MBSTS Model

In general, the model in state space form can be written as:

$$\tilde{y}_t = \tilde{\mu}_t + \tilde{\tau}_t + \tilde{\omega}_t + \tilde{\xi}_t + \tilde{\epsilon}_t \qquad \tilde{\epsilon}_t \overset{iid}{\sim} N_m(0, \Sigma_\epsilon) \quad t = 1, 2 \ldots, n. \quad (1)$$

Based on state space form, α_t is the collection of these components, namely $\alpha_t = [\tilde{\mu}_t^T, \ \tilde{\tau}_t^T, \ \tilde{\omega}_t^T, \ \tilde{\xi}_t^T]^T$.

MBSTS Model

Trend component

$$\tilde{\mu}_{t+1} = \tilde{\mu}_t + \tilde{\delta}_t + \tilde{u}_t, \qquad \tilde{u}_t \overset{iid}{\sim} N_m(0, \Sigma_\mu), \tag{2}$$

$$\tilde{\delta}_{t+1} = \tilde{D} + \tilde{\rho}(\tilde{\delta}_t - \tilde{D}) + \tilde{v}_t, \qquad \tilde{v}_t \overset{iid}{\sim} N_m(0, \Sigma_\delta). \tag{3}$$

Seasonality component

$$\tau^{(i)}_{t+1} = -\sum_{k=0}^{S_i-2} \tau^{(i)}_{t-k} + w^{(i)}_t, \qquad \tilde{w}_t = [w^{(1)}_t, \cdots, w^{(m)}_t]^T \overset{iid}{\sim} N_m(0, \Sigma_\tau), \tag{4}$$

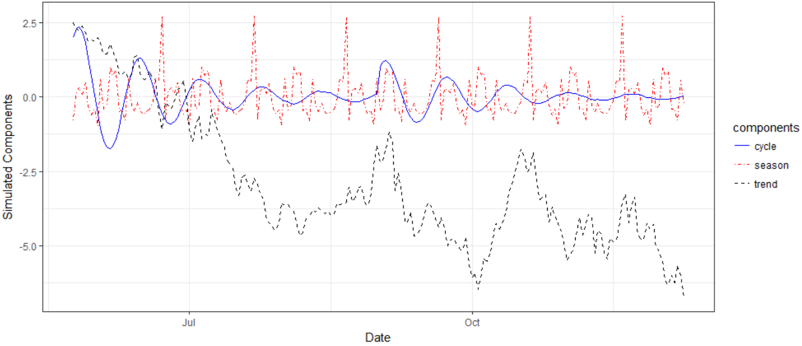
MBSTS Model

Cyclical effect component

$$\begin{aligned} \tilde{\omega}_{t+1} &= \tilde{\varrho}\widehat{cos(\lambda)}\tilde{\omega}_t + \tilde{\varrho}\widehat{sin(\lambda)}\tilde{\omega}^\star_t + \tilde{\kappa}_t, \qquad \tilde{\kappa}_t \overset{iid}{\sim} N_m(0, \Sigma_\omega), \\ \tilde{\omega}^\star_{t+1} &= -\tilde{\varrho}\widehat{sin(\lambda)}\tilde{\omega}_t + \tilde{\varrho}\widehat{cos(\lambda)}\tilde{\omega}^\star_t + \tilde{\kappa}^\star_t, \qquad \tilde{\kappa}^\star_t \overset{iid}{\sim} N_m(0, \Sigma_\omega), \end{aligned} \tag{5}$$

where $\tilde{\varrho}$, $\widehat{sin(\lambda)}$, $\widehat{cos(\lambda)}$ are $m \times m$ diagonal matrices with diagonal entries equal to ϱ_{ii}, $\sin(\lambda_{ii})$ where $\lambda_{ii} = 2\pi/q_i$ is the frequency with q_i being a period such that $0 < \lambda_{ii} < \pi$, and $\cos(\lambda_{ii})$ respectively.

MBSTS Model


2.5
0.0
-2.5
-5.0
Simulated Components
Jul
Oct
Date
components
cycle
season
trend

MBSTS Model

Regression component

$$\xi_t^{(i)} = \beta_i^T x_t^{(i)}. \tag{6}$$

For target series $y^{(i)}$, the $x_t^{(i)} = [x_{t1}^{(i)}, \dots, x_{tk_i}^{(i)}]^T$ is the pool of all available predictors at time t, and $\beta_i = [\beta_{i1}, \dots, \beta_{ij}, \dots, \beta_{ik_i}]^T$ represent corresponding static regression coefficients.

MBSTS Model

Spike and Slab Regression:

- In feature selection, a high degree of sparsity is expected, in the sense that the coefficients for the vast majority of predictors will be zero.
- A natural way to represent sparsity in the Bayesian paradigm is through the spike and slab coefficients.
- One advantage of working in a fully Bayesian setting is that we do not need to commit to a fixed set of predictors.

MBSTS Model

$$\tilde{Y} = \tilde{M} + \tilde{T} + \tilde{W} + X\beta + \tilde{E} \tag{7}$$

where $\tilde{Y} = vec(Y)$, $\tilde{M} = vec(M)$, $\tilde{T} = vec(T)$, $\tilde{W} = vec(W)$, $\tilde{E} = vec(E)$, and X, β are written as:

$$X = \begin{bmatrix} X_1 & 0 & 0 & \dots & 0 \\ 0 & X_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & X_m \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} \tag{8}$$

where X_i being $n \times k_i$ matrix, representing all observations of k_i candidate predictors for $y^{(i)}$.

MBSTS Model

Prior distribution and elicitation

We define $\gamma_{ij} = 1$ if $\beta_{ij} \neq 0$, and $\gamma_{ij} = 0$ if $\beta_{ij} = 0$. Then $\gamma = [\gamma_1, \ldots, \gamma_m]$, where $\gamma_i = [\gamma_{i1}, \ldots, \gamma_{ik_i}]$. The **spike** prior may be written as:

$$\gamma \sim \prod_{i=1}^{m} \prod_{j=1}^{k_i} \pi_{ij}^{\gamma_{ij}} (1 - \pi_{ij})^{1-\gamma_{ij}} \tag{9}$$

where π_{ij} is prior inclusion probability of j^{th} predictor for i^{th} response series.

MBSTS Model

Prior distribution and elicitation

A simple **slab** prior specification is to make β and Σ_ϵ prior independent:

$$\begin{aligned} p(\beta, \Sigma_\epsilon, \gamma) &= p(\beta|\gamma)p(\Sigma_\epsilon|\gamma)p(\gamma) \\ \beta|\gamma &\sim N_K(b_\gamma, A_\gamma^{-1}) \\ \Sigma_\epsilon|\gamma &\sim IW(v_0, V_0) \end{aligned} \tag{10}$$

where b_γ is the vector of prior means with the same dimension as β_γ, and A_γ is the full-model prior information matrix.

MBSTS Model

By the law of total probability, the full likelihood function is given by

$$p(\tilde{Y}^\star, \beta, \Sigma_\epsilon, \gamma) = p(\tilde{Y}^\star|\beta, \Sigma_\epsilon, \gamma) \times p(\beta|\gamma) \times p(\Sigma_\epsilon|\gamma) \times p(\gamma), \quad (11)$$

$$p(\tilde{Y}^\star|\beta, \Sigma_\epsilon, \gamma) \propto |\Sigma_\epsilon|^{-n/2} \exp\left(-\frac{1}{2}(\tilde{Y}^\star - X_\gamma\beta_\gamma)^T(\Sigma_\epsilon^{-1} \otimes I_n)(\tilde{Y}^\star - X_\gamma\beta_\gamma \quad (12)$$

$$p(\beta|\gamma) \propto |A_\gamma|^{1/2} \exp\left(-\frac{1}{2}(\beta_\gamma - b_\gamma)^T A_\gamma(\beta_\gamma - b_\gamma)\right), \quad (13)$$

$$p(\Sigma_\epsilon|\gamma) \propto |\Sigma_\epsilon|^{-(v_0+m+1)/2} \exp\left(tr(-\frac{1}{2}V_0\Sigma_\epsilon^{-1})\right), \quad (14)$$

where $\tilde{Y}^\star = \tilde{Y} - \tilde{M} - \tilde{T} - \tilde{W}$ is the multiple response series $\tilde{Y}$ with time series components subtracted out.

MBSTS Model

Posterior Inference

$$\beta | \hat{Y}^{\star}, \Sigma_{\epsilon}, \gamma \sim N_K(\tilde{\beta}_{\gamma}, (\hat{X}_{\gamma}^T \hat{X}_{\gamma} + A_{\gamma})^{-1}). \tag{15}$$

$$\Sigma_{\epsilon} | \tilde{Y}^{\star}, \beta, \gamma \sim IW(v_0 + n, E_{\gamma}^T E_{\gamma} + V_0). \tag{16}$$

$$p(\gamma | \Sigma_{\epsilon}, \tilde{Y}^{\star}) = C(\Sigma_{\epsilon}, \tilde{Y}^{\star}) \frac{|A_{\gamma}|^{1/2} p(\gamma)}{|\hat{X}_{\gamma}^T \hat{X}_{\gamma} + A_{\gamma}|^{1/2}} \times \exp\left(-\frac{1}{2}\{b_{\gamma}^T A_{\gamma} b_{\gamma} - Z_{\gamma}^T (\hat{X}_{\gamma}^T \hat{X}_{\gamma} + A_{\gamma})^{-1} Z_{\gamma}\}\right). \tag{17}$$

$$\Sigma_{\mu,\delta,\tau,\omega} | \mu, \delta, \tau, \omega \sim IW(w_{\mu,\delta,\tau,\omega} + n, W_{\mu,\delta,\tau,\omega} + AA^T). \tag{18}$$

MBSTS Model

Markov Chain Monte Carlo

MCMC methods are a class of algorithms to sample from a probability distributions ((15), (16), (17) and (18)) based on constructing a Markov chain that has the desired distribution as its equilibrium distribution. The state of the chain after a number of steps is then used as a sample from the desired distribution.

Model Training

Let

$$\theta = (\Sigma_\mu, \Sigma_\delta, \Sigma_\tau, \Sigma_\omega)$$

denote the set of state component parameters. Looping through the five steps yields a sequence of draws

$$\tilde{\psi} = (\alpha, \theta, \gamma, \Sigma_\epsilon, \beta)$$

from a Markov chain with stationary distribution $p(\tilde{\psi}|Y)$, the posterior distribution of $\tilde{\psi}$ given Y.

MBSTS Model Training

1: Draw the latent state $\alpha = (\tilde{\mu}, \tilde{\delta}, \tilde{\tau}, \tilde{\omega})$ from given model parameters and $\tilde{Y}$, namely $p(\alpha|\tilde{Y}, \theta, \gamma, \Sigma_\epsilon, \beta)$, using the posterior simulation algorithm Durbin and Koopman (2002).

2: Draw time series state component parameters θ given α, namely simulating $\theta \sim p(\theta|\tilde{Y}, \alpha)$ based on equation (18).

3: Loop over i in an random order, draw each $\gamma_i|\gamma_{-i}, \tilde{Y}, \alpha, \Sigma_\epsilon$, namely simulating $\gamma \sim p(\gamma|\tilde{Y}^\star, \Sigma_\epsilon)$ one by one based on equation (17), using the stochastic search variable selection (SSVS) algorithm from George and McCulloch (1997).

4: Draw β given Σ_ϵ, γ, α and $\tilde{Y}$, namely simulating $\beta \sim p(\beta|\Sigma_\epsilon, \gamma, \tilde{Y}^\star)$ based on equation (15).

5: Draw Σ_ϵ given γ, α, β and $\tilde{Y}$, namely simulating $\Sigma_\epsilon \sim p(\Sigma_\epsilon|\gamma, \tilde{Y}^\star, \beta)$ based on equation (16).

Target Series Forecasting Let $\hat{Y}$ represents the set of values to be forecast. The posterior predictive distribution of $\hat{Y}$ can be expressed as follows:

$$p(\hat{Y}|Y) = \int p(\hat{Y}|\tilde{\psi})p(\tilde{\psi}|Y)d\tilde{\psi}, \tag{19}$$

where $\tilde{\psi}$ is the set of all model parameters and latent states randomly drawn from $p(\tilde{\psi}|Y)$, then we can draw samples of $\hat{Y}$ from $p(\hat{Y}|\tilde{\psi})$.

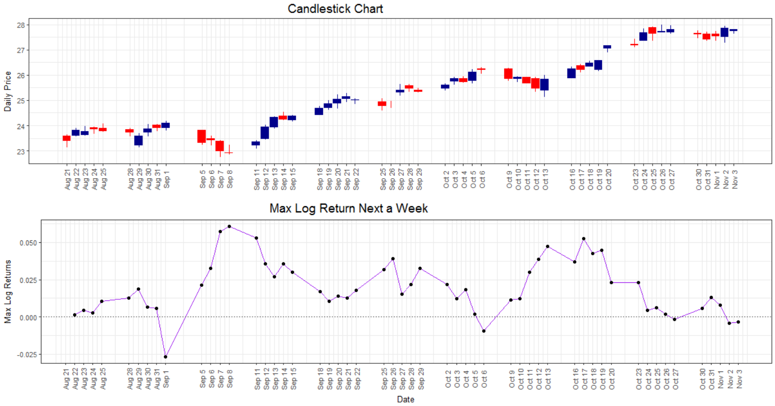
Empirical Analysis

Data: Daily stock price of Bank of America (BOA), Capital One Financial Corporation (COF), J.P. Morgan (JPM) and Wells Fargo (WFC).
Time horizon: 11/27/2006 – 11/03/2017
Source: Google Finance.
Purpose: Trade when its future price is predicted to vary more than $p\%$.
Goal: Forecast the trend of stock movement in the next $k(=5)$ days.

We approximate the daily average price as: $\bar{P}_t = (C_t + H_t + L_t)/3$, where C_t, H_t and L_t are the close, high, and low quotes for day t respectively.

However, instead of using the arithmetic returns, we are interested in the log return V_t defined as $V_t = \{\log(\bar{P}_{t+j}/C_t)\}_{j=1}^{k}$.

We consider the indicator variable $y_t = \max\{v \in V_t\}$, the maximum value of log returns over the next k days.


Candlestick Chart
Daily Price
Max Log Return Next a Week
Max Log Returns
Date

Fundamental analysis claims that markets may incorrectly price a security in the short run but will eventually correct it.

Trend	Abbr.	Trend	Abbr.
Advertising & marketing	advert	Air travel	airtvl
Auto buyers	auto	Auto financing	autoby
Automotive	autofi	Business & industrial	bizind
Bankruptcy	bnkrpt	Commercial Lending	comlnd
Computers & electronics	comput	Construction	constr
Credit cards	crcard	Durable goods	durble
Education	educat	Finance & investing	invest
Financial planning	finpln	Furniture	furntr
Insurance	insur	Jobs	jobs
Luxury goods	luxury	Mobile & wireless	mobile
Mortgage	mtge	Real estate	rlest
Rental	rental	Shopping	shop
Small business	smallbiz	Travel	travel
Unemployment	unempl		

Table: Google domestic trends

Technical analysis claims that useful information is already reflected in the price of a security. We selected a representative set of technical indicators to capture the volatility, momentum and trend, close location value, and potential reversals of each stock.

Variable	Abbr.
Chaikin volatility	ChaVol
Yang and Zhang Volatility historical estimator	Vol
Arms' Ease of Movement Value	EMV
Moving Average Convergence/Divergence	MACD
Money Flow Index	MFI
Aroon Indicator	AROON
Parabolic Stop-and-Reverse	SAR
Close Location Value	CLV

Table: Stock Technical Predictors

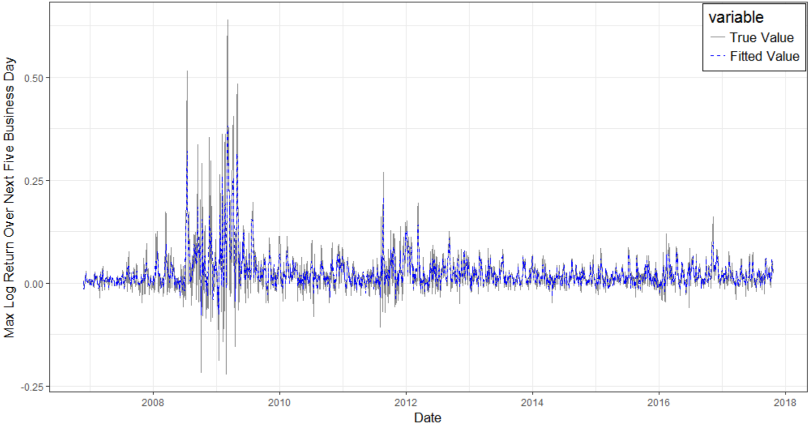
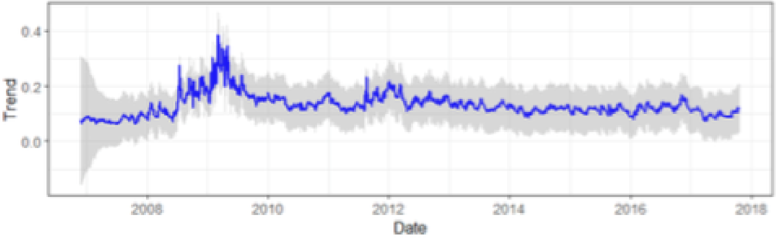
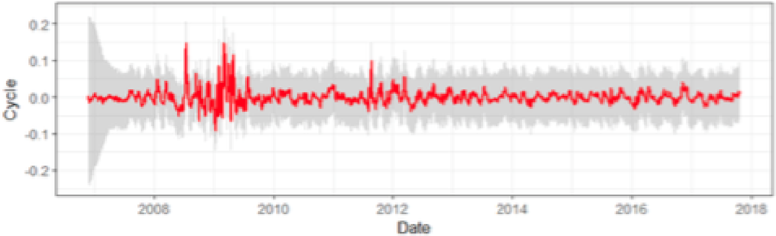


Figure: True and fitted values of max log return from 11/27/2006 to 10/20/2017 (BOA)



(a) Trend Component



(b) Cyclical Component

Empirical posterior inclusion probability for the most likely predictors of max log return.

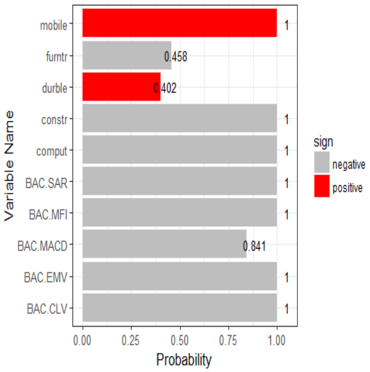


Figure: Bank of America Corp.

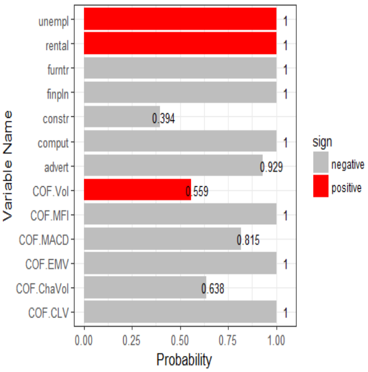


Figure: Capital One Financial Corp.

Empirical posterior inclusion probability for the most likely predictors of max log return.

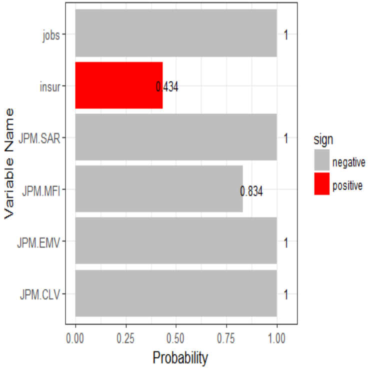


Figure: JPMorgan Chase & Co.

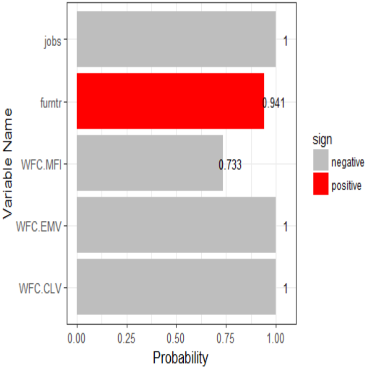


Figure: Wells Fargo & Co.

Cumulative absolute one-step-ahead prediction error.

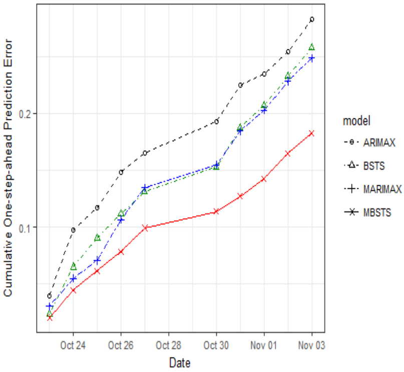


Figure: All Predictors Without Deaseasonal

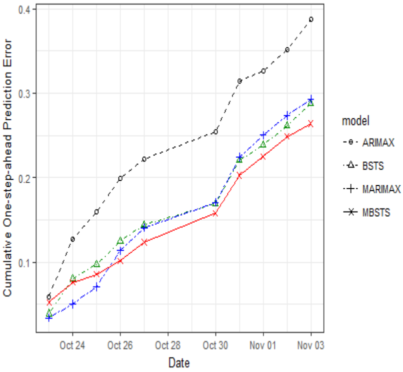


Figure: Partial Predictors With Deaseasonal

One-step-ahead predction of max log return.

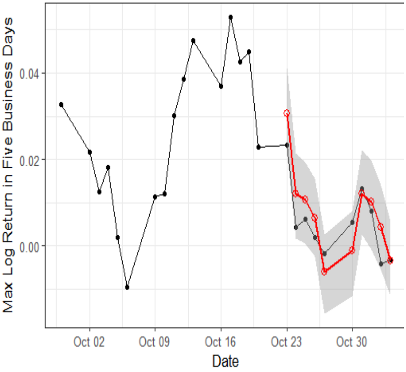


Figure: Bank of America Corp.

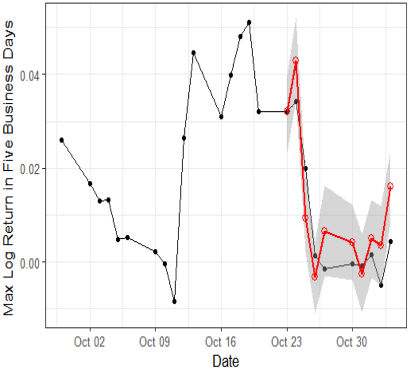


Figure: JPMorgan Chase & Co.

One-step-ahead predction of max log return.

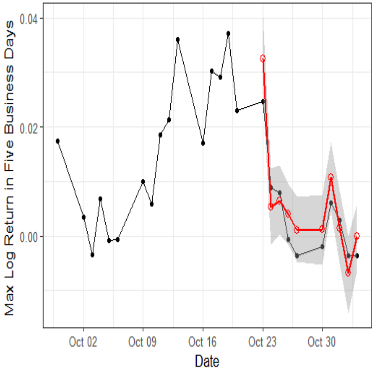


Figure: JPMorgan Chase & Co.

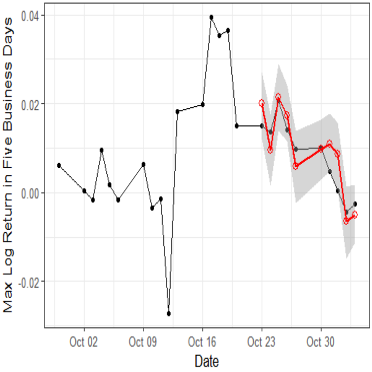


Figure: Wells Fargo & Co.

Thank you!