Journal of
*Risk and Financial Management*

MDPI

*Article*

# Supplementary Material for Is Bitcoin a Relevant Predictor for S&P 500?

**Camilla Muglia, Luca Santabarbara and Stefano Grassi ***

Department of Economics and Finance, University of Rome 'Tor Vergata', Via Columbia 2, 00133 Rome, Italy; camilla.muglia@gmail.com (C.M.); luca-santa@hotmail.it (L.S.)

* Correspondence: stefano.grassi@uniroma2.it

Received: 14 May 2019; Accepted: 17 May 2019; Published: 31 May 2019

---

## 1. Methodology

In the case of $m$ potential regressors there are $2^m$ time-varying parameter (TVP) models to consider. To forecast at each time $t$ it is necessary to contemporary estimate, compare and analyse $K^t = 2^{mt}$ systems. From a technical point of view, handle such a large number of combinations is not just cumbersome but also memory involving. To deal with such a big number of competing models Raftery et al. (2010) and Koop and Korobilis (2012) recently proposed the forgetting factor methodology that allows the fast estimation of time-varying parameters and models weights.

Let's consider $m$ predictors that gives $K = 2^m$ model at each time point $t$. The resulting state space model with all the possible predictors combinations has the following representation:

$$\mathbf{y}_t = \mathbf{z}_t^{(k)} \gamma_t^{(k)} + \boldsymbol{\varepsilon}_t^{(k)} \qquad \boldsymbol{\varepsilon}_t^{(k)} \sim \mathrm{N}\left(\mathbf{0}, \mathbf{H}_t^{(k)}\right) \tag{S1}$$

$$\gamma_{t+1}^{(k)} = \gamma_t^{(k)} + \boldsymbol{\eta}_t^{(k)} \qquad \boldsymbol{\eta}_t^{(k)} \sim \mathrm{N}\left(\mathbf{0}, \mathbf{Q}_t^{(k)}\right) \tag{S2}$$

where $k = \{1, \ldots, K\}$ indicates the model which is characterized by a model-specific "sub"- set of predictors $\mathbf{z}_t^{(k)}$. When $k = K$ the vector of predictors is composed of all the possible regressors: the intercept (first entrance), the lags of the dependent variable, the exogenous variables and their relative lags. $\gamma_t^{(k)}$ are the associated predictor's coefficients that follow a Random Walk dynamic.

The errors $\boldsymbol{\varepsilon}_t^{(k)}$ and $\boldsymbol{\eta}_t^{(k)}$ are mutually independent and uncorrelated. Furthermore, the errors covariance matrices, $\mathbf{H}_t^{(k)}$ and $\mathbf{Q}_t^{(k)}$, are unknown and need to be estimated.

In practice, for each $k$ there is a TVP model expressing a linear and time-evolving relationship among the dependent variable $\mathbf{y}_t$ and the explanatory variables in $\mathbf{z}_t^{(k)}$. For $k = 1, 2, \ldots, K$ the result is a series of TVP models to be inspected at every $t$. The contemporaneous estimation of these models can be computationally involving, or even infeasible, with the maximum likelihood or MCMC methods.

To overcome this issue Raftery et al. (2010) introduced an approximation based on three hyperparameters. The first one, $\lambda$, avoids the calculation of variance-covariance matrix $\mathrm{Q}_t$. Koop and Korobilis (2012) applied this methodology to economics, estimating the time-varying volatility ($\mathbf{H}_t$) via an EWMA with decay factor $\kappa$. Lastly, $\alpha$ is the hyperparameter which measures the model's weight based on its forecast performance (model switching).

Note that $\lambda$ and $\alpha$ are involved in the state equation for parameters and models respectively. The KF starts with:

$$\gamma_{t-1} | \mathrm{Y}_{t-1} \sim \mathrm{N}\left(\hat{\gamma}_{t-1|t-1}, \Sigma_{t-1|t-1}\right) \tag{S3}$$

where, $Y_{t-1} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{t-1})$, $\hat{\gamma}_{t-1|t-1} = \mathrm{E}\left(\gamma_{t-1}|Y_{t-1}\right)$ and $\Sigma_{t-1|t-1} = \mathrm{Var}\left(\gamma_{t-1}|Y_{t-1}\right)$. At each time point $t$, the algorithm iterates between: prediction Equation (S4), updating Equation (S5) and the predictive density (S6):

$$\gamma_t|Y_{t-1} \sim \mathrm{N}\left(\hat{\gamma}_{t|t-1}, \Sigma_{t|t-1},\right) \tag{S4}$$

$$\gamma_t|Y_t \sim \mathrm{N}\left(\hat{\gamma}_{t|t}, \Sigma_{t|t}\right), \tag{S5}$$

$$\mathbf{y}_t|Y_{t-1} \sim \mathrm{N}\left(\mathbf{z}_t\hat{\gamma}_{t|t-1}, \mathbf{H}_t + \mathbf{z}_t\Sigma_{t|t-1}\mathbf{z}_t'\right). \tag{S6}$$

The quantity $\Sigma_{t|t-1}$ depends on the error variances: $\Sigma_{t|t-1} = \Sigma_{t-1|t-1} + \mathbf{Q}_t$. Raftery et al. (2010) proposed an approximation given by:

$$\Sigma_{t|t-1} = \frac{1}{\lambda}\Sigma_{t-1|t-1}. \tag{S7}$$

Correspondingly, $\mathbf{Q}_t = \left(\frac{1}{\lambda} - 1\right)\Sigma_{t-1|t-1}$ with $\lambda \in (0, 1]$. The tuning parameter $\lambda$ plays a crucial role in adjusting the effective memory of the algorithm, leading to a weighted estimation where data $i$ times ago has weight $\lambda^i$. For example, in the case of daily data, setting $\lambda = 0.99$ implies that observations ten days ago will receive 90% as much weight as last periods observation. Whereas for $\lambda = 0.92$, observations ten days ago will receive 43% as much weight as last periods observation. The first case, $\lambda = 0.99$, is consistent with models where changes in $\gamma_t$ are gradual. The second, $\lambda = 0.92$, is consistent with models where changes in $\gamma_t$ are quite rapid and abrupt.

It is well known that both macroeconomic and financial time series are characterized by heteroskedastic effects. Therefore, following Koop and Korobilis (2012) $H_t$ is assumed to follow an EWMA:

$$H_t = \kappa H_{t-1} + (1-\kappa)\,\nu_t^2. \tag{S8}$$

The EWMA estimator requires to select a value for $\kappa$. As suggested in Koop and Korobilis (2012), the value of $\kappa$ is set to 0.94 for daily data.

In the multi-model framework, the state vector $\mathbf{\Gamma}_t = \left(\gamma_t^{(1)}, \gamma_t^{(2)}, \gamma_t^{(3)}, \ldots, \gamma_t^{(K)}\right)$ can be split into independent blocks. Predictions, outputs and other results are conditioned on model $k$ ($M_t = k$), with $k = 1, 2, \ldots, K$. Raftery et al. (2010) introduced the parameter $\alpha$ to easily move among models without using more complicated methodology like reversible jump type of algorithm. The switching is based on posterior probabilities:

$$\begin{aligned} p\left(\mathbf{\Gamma}_{t-1}|Y_{t-1}\right) &= \sum_{k=1}^{K} p\left(\gamma_{t-1}^{(k)}|M_{t-1} = k, Y_{t-1}\right) p\left(M_{t-1} = k|Y_{t-1}\right) = \\ &= \sum_{k=1}^{K} p\left(\gamma_{t-1}^{(k)}|M_{t-1} = k, Y_{t-1}\right) \pi_{t-1|t-1,k}. \end{aligned} \tag{S9}$$

The estimation of the $k^{th}$- model weight at time $t$ using $Y_{t-1}$, $\pi_{t|t-1,k} = p\left(M_t = k|Y_{t-1}\right)$, is obtained with the model prediction Equation (S10). The calculation of the $k^{th}$- model weight at time $t$ using $Y_t$, $\pi_{t|t,k} = p\left(M_t = k|Y_t\right)$, comes from the model updating Equation (S11):

$$\pi_{t|t-1,k} = \sum_{l=1}^{K} \pi_{t-1|t-1,l}\,p_{kl} = \frac{\pi_{t-1|t-1,k}^{\alpha}}{\sum_{l=1}^{K} \pi_{t-1|t-1,l}^{\alpha}}. \tag{S10}$$

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k}\,p_k\left(\mathbf{y}_t|Y_{t-1}\right)}{\sum_{l=1}^{K} \pi_{t|t-1,l}\,p_l\left(\mathbf{y}_t|Y_{t-1}\right)}. \tag{S11}$$

The probability attached to model $k$ by DMA is equivalent to:

$$\pi_{t|t-1,k} \propto \left[\pi_{t-1|t-2,k} p_k \left(\mathbf{y}_{t-1}|\Upsilon_{t-2}\right)\right]^{\alpha} = \prod_{i=1}^{t-1} \left[p_k \left(\mathbf{y}_{t-i}|\Upsilon_{t-i-1}\right)\right]^{\alpha^i}. \tag{S12}$$

where $p_k \left(\mathbf{y}_{t-i}|\Upsilon_{t-i-1}\right)$ is the predictive density for the model $k$ evaluated at $\mathbf{y}_{t-i}$ with $i = 1, \ldots, t-1$. The forgetting factor $\alpha \in (0, 1]$ gives a measure of the model performance rate of decay. The forecast performance recorded $i$ periods ago has a significance equal to $\alpha^i$. Note that when $\alpha = 0$ all models are equally probable for every $t$: the models weight remain unchanged from the prior, $\pi_{0|0,k} = \frac{1}{K}$. From the recursive iteration a prediction for every model $k$ is obtained:

$$\mathbf{y}_t|M_t = k, \Upsilon_{t-1} \sim \mathrm{N}\left(\mathbf{z}_t^{(k)} \hat{\gamma}_{t|t-1}^{(k)}, \mathbf{H}_t^{(k)} + \mathbf{z}_t^{(k)} \Sigma_{t|t-1}^{(k)} \mathbf{z}_t^{(k)'}\right) \tag{S13}$$

DMA point estimates result averaging at every $t$ across the obtained $K$ models predictions. In other words, the DMA dependent variable forecast, $\mathbf{y}_{DMA_t}$, comes from a weighted average of all the models forecasts where the weights are the conditional probabilities $P\left(M_t = k|\Upsilon_{t-1}\right) = \pi_{t|t-1,k}$ computed using the information up to time $t-1$ for $k = 1, 2, \ldots, K$:

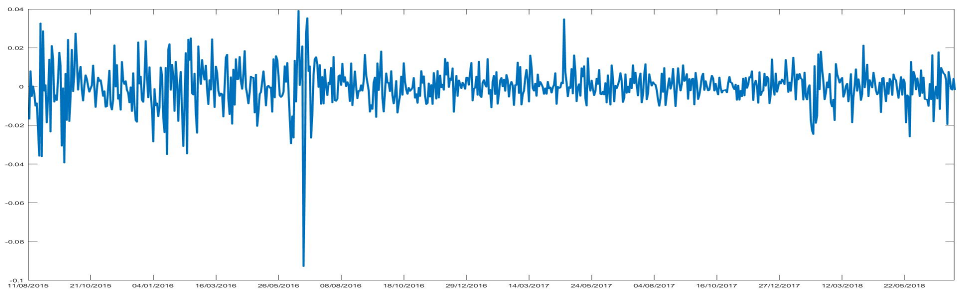
$$\mathbf{y}_{DMA_t} = E\left(\mathbf{y}_t|\Upsilon_{t-1}\right) = \sum_{k=1}^{K} \pi_{t|t-1,k} \mathbf{z}_t^{(k)} \hat{\gamma}_{t|t-1}^{(k)} \tag{S14}$$

DMS selects and uses a single model to make predictions of the dependent variable. The selection process is based on the evaluation of the conditional probability $P\left(M_t = k|\Upsilon_{t-1}\right)$: the methodology picks, among all the analysed models, the one with the highest predictive power capacity. Specifically, DMS extracts the model with the highest conditional probability value $\pi_{t|t-1,k}$. This happens at each time, $t = 1, 2, \ldots, T$.

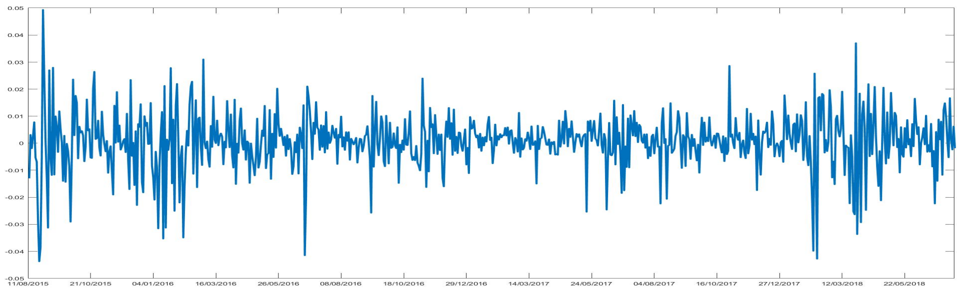
Koop and Korobilis (2012) refer to the special case $\lambda = \alpha = 1$ as Bayesian Model Averaging (BMA) which is very popular in macroeconomics and finance, see Koop and Potter (2004). Using the simultaneous combination $\lambda = \alpha = 1$, means performing forecast with conventional linear fixed coefficients models in a Bayesian framework.

The definition of a prior for $\pi_{0|0,k}$ and $\gamma_0^{(k)}$ is essential to implement DMA, DMS and BMA. A non-informative prior is chosen for both the states and the weights. In particular, $\pi_{0|0,k} = \frac{1}{K}$ and $\gamma_0^{(k)} \sim \mathrm{N}\left(\mathbf{0}, \mathrm{I}\right)$ for $k = 1, 2, 3, \ldots, K$. This means that, at the beginning, all models are equally likely. The peculiarity to be part of the "adaptive algorithms" class can be translated in the need to wait some time to have coherent, accurate and reliable results.

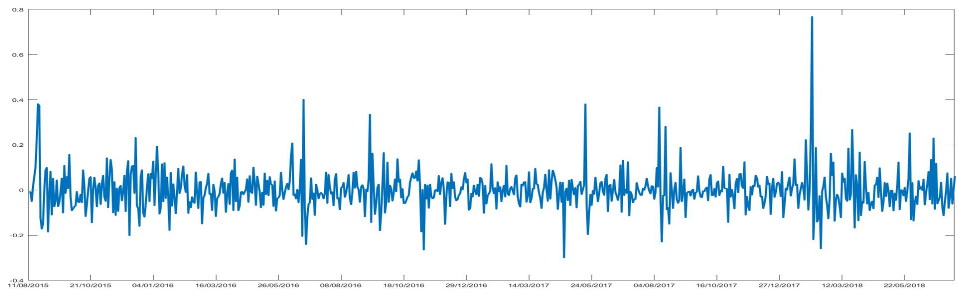
## 2. Further Results




(a) EF300

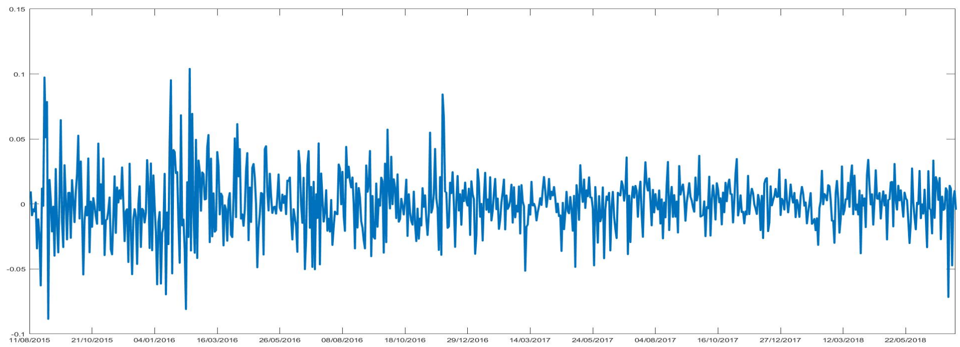



(b) NASDAQ

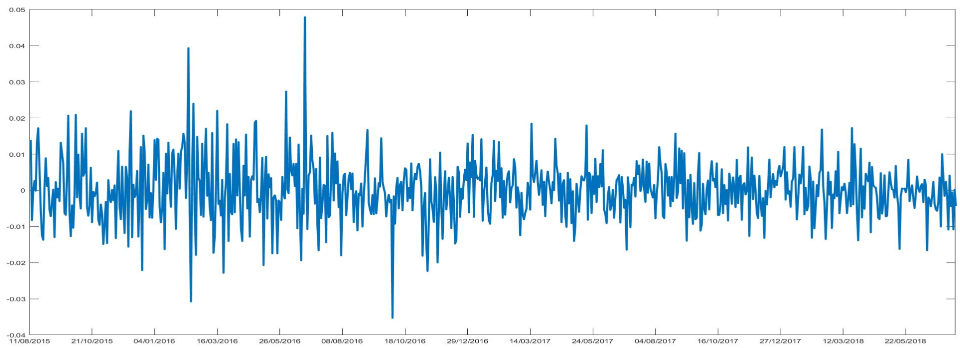



(c) VIX

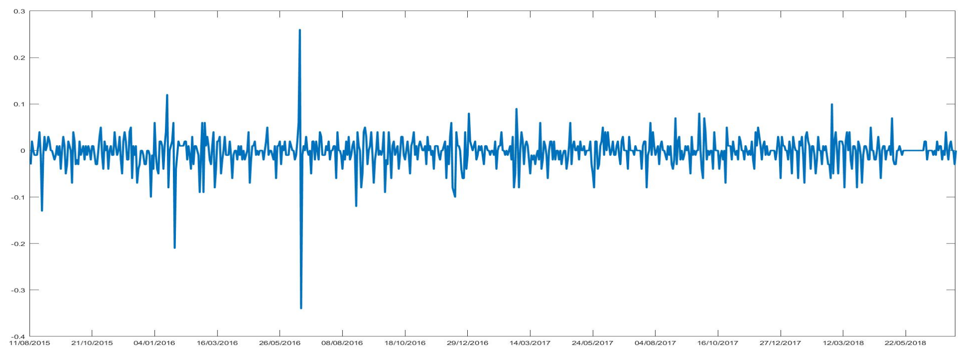
**Figure S1.** Other predictors: financial indexes. FTSEuroFirst300 (Panel (a)) shows substantially low volatility. The huge drop in June 2016 is due to the Brexit effect. Nasdaq index (Panel (b)) appears to be more volatile than the European index. Recall that Nasdaq is strongly dependent on the IT sector. Thus its fluctuations are determined by the movements of the biggest IT companies, such as Apple and Microsoft. VIX (Panel (c)) is the most volatile. Note that the scale is different. The highest peak is on March 2018, the period which followed the drop in price of Bitcoin.



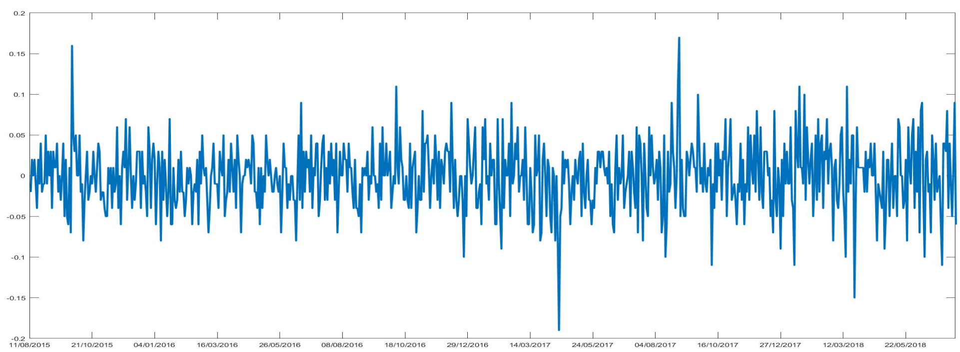

(a) OIL




(b) GOLD

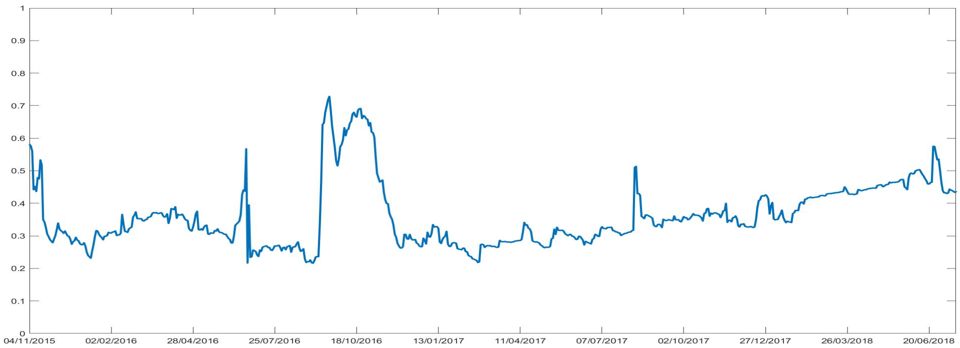



(c) 1mUS

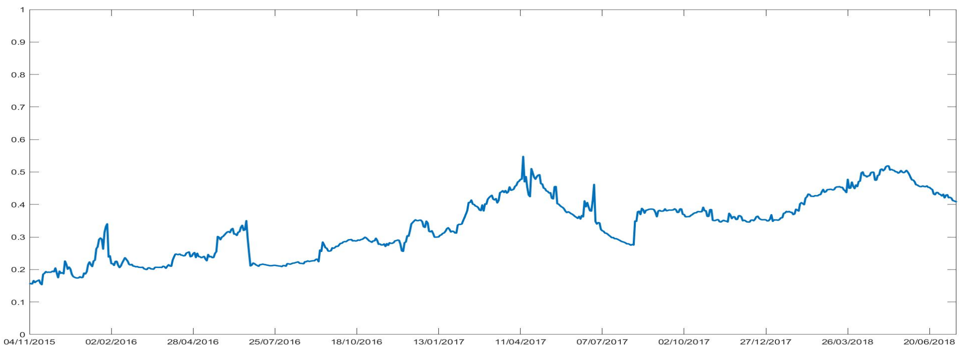



(d) 10yUS

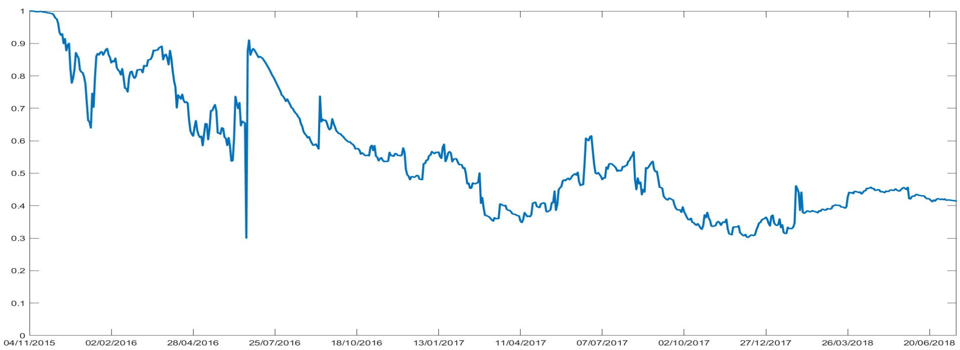
**Figure S2.** Other predictors: commodities and constant maturities. Both ICE Brent (Panel (a)) and SPDR Gold Shares (panel (b)) show low volatility. They are usually considered as safe investments. In the period under analysis they display lower volatility than the Treasury Constant Maturity Rate. Constant maturities panel (c) and (d) reflect the economic recovery which characterized the U.S. economy in this last years, after the great recession of 2007-2009.




(a) Time-Varying Probability of Inclusion EF300

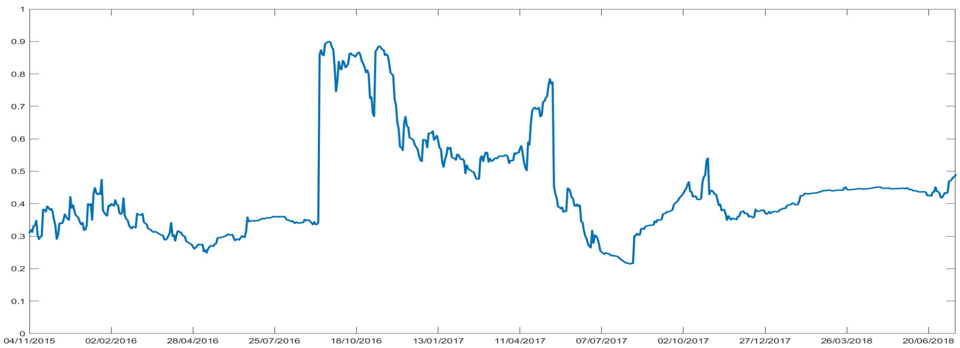



(b) Time-Varying Probability of Inclusion NASDAQ

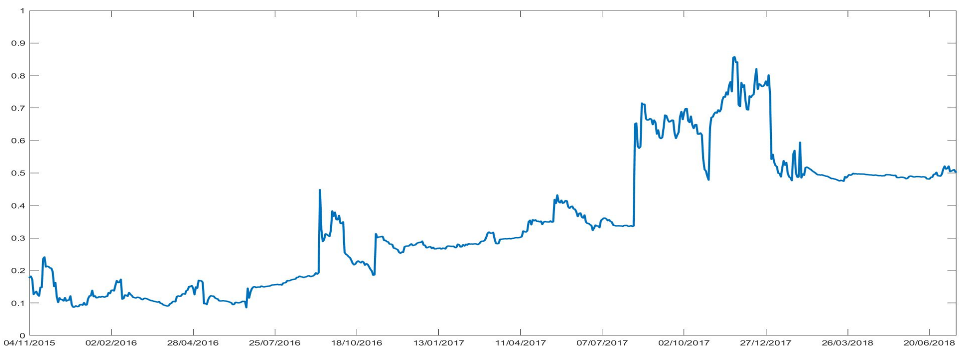



(c) Time-Varying Probability of Inclusion VIX

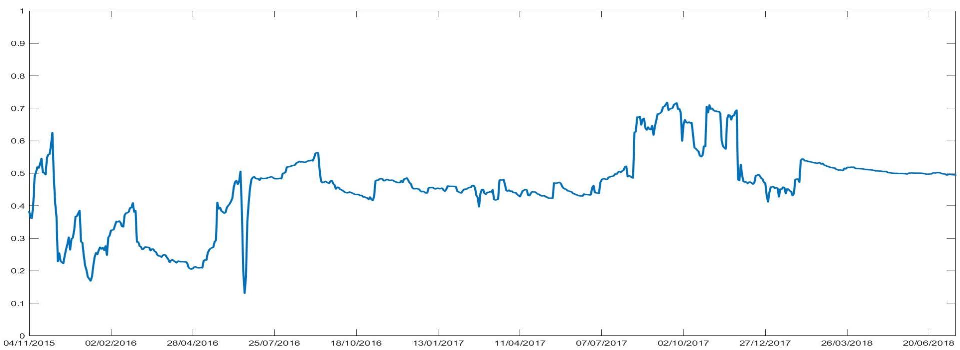
**Figure S3.** Posterior inclusion probabilities: financial indeces. VIX, panel (c) seems to be the one which affects the most the S&P 500. This is in line with the way in which VIX is defined: it is a measure of market's volatility implied by S&P 500. Its importance seems to be highly persistent especially in 2016. This reflects once again the turmoil of the financial markets during that year. After that, it shows a flatter period. The predictive power of Nasdaq, panel (b), is almost irrelevant. Similarly for the European index, FTSEuroFirst300, panel (a). The only period in which it seems to have importance is represented, once again, by the months which followed the Brexit.



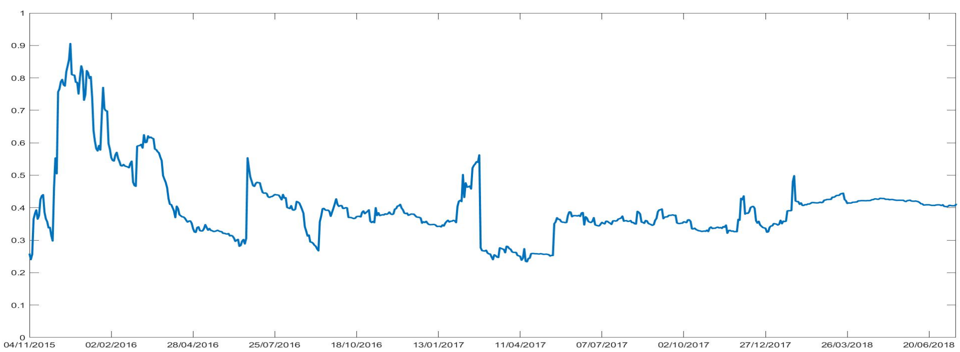

(a) Time-Varying Probability of Inclusion OIL




(b) Time-Varying Probability of Inclusion GOLD




(c) Time-Varying Probability of Inclusion 1mUS




(d) Time-Varying Probability of Inclusion 10yUS

**Figure S4.** Posterior inclusion probabilities: commodities and constant maturities. Both ICE Brent (Panel (a)) and SPDR Gold Shares (panel (b)) show low volatility. They are usually considered as safe investments. In the period under analysis they display lower volatility than the Treasury Constant Maturity Rate. Constant maturities panel (c) and (d) reflect the economic recovery which characterized the U.S. economy in this last years, after the great recession of 2007-2009.

**Table S1.** Correlation matrix of the predictors: the BTC appears to be highly positively correlated especially with all the financial indexes S&P500, EF300 and NASDAQ.

| | S&P500 | BTC | BHL | OIL | GOLD | VIX | EF300 | NASDAQ | 1mUS | 10yUS |
|---|---|---|---|---|---|---|---|---|---|---|
| S&P500 | 1.000 | 0.840 | 0.586 | 0.882 | 0.539 | -0.427 | 0.872 | 0.986 | -0.878 | -0.642 |
| BTC | | 1.000 | 0.802 | 0.761 | 0.440 | -0.192 | 0.809 | 0.834 | -0.629 | -0.472 |
| BHL | | | 1.000 | 0.489 | 0.300 | -0.141 | 0.597 | 0.558 | -0.400 | -0.365 |
| OIL | | | | 1.000 | 0.464 | -0.258 | 0.749 | 0.879 | -0.716 | -0.372 |
| GOLD | | | | | 1.000 | -0.299 | 0.354 | 0.494 | -0.642 | -0.650 |
| VIX | | | | | | 1.000 | -0.293 | -0.335 | 0.511 | 0.557 |
| EF300 | | | | | | | 1.000 | 0.880 | -0.632 | -0.551 |
| NASDAQ | | | | | | | | 1.000 | -0.848 | -0.600 |
| 1mUS | | | | | | | | | 1.000 | 0.791 |
| 10yUS | | | | | | | | | | 1.000 |

**Table S2.** DM statistics for the point forecast: $M_1$ vs $M_2$. Again, most of them are included in the interval $[-1.96, 1.96]$. Thus they are almost never falling outside the boundaries. This means that the null hypothesis of equal forecasting ability is never rejected.

| | $\lambda = \alpha = 0.99$ | | $\lambda = \alpha = 0.95$ | | $\lambda = 0.99$ | $\alpha = 1$ | $\lambda = 1$ | $\alpha = 0.99$ | $\lambda = \alpha = 1$ |
|---|---|---|---|---|---|---|---|---|---|
| | DMA | DMS | DMA | DMS | DMA | DMS | DMA | DMS | BMA |
| $h = 1$ | -1.573 | -0.061 | 0.543 | -0.443 | -1.893 | -0.768 | -2.627 | -1.085 | -2.581 |
| $h = 2$ | -1.670 | 0.340 | -1.492 | 0.299 | -1.579 | 0.302 | -1.722 | -0.837 | -1.739 |
| $h = 3$ | -1.283 | -1.197 | -1.226 | -0.336 | -1.288 | -1.271 | -1.302 | -1.326 | -1.312 |
| $h = 4$ | -1.263 | -0.149 | -1.204 | 0.058 | -1.246 | 0.008 | -1.314 | -0.480 | -1.287 |
| $h = 5$ | -1.527 | -0.959 | -1.374 | -0.834 | -1.534 | -0.954 | -1.611 | -0.996 | -1.618 |
| $h = 6$ | -1.197 | -0.364 | -1.181 | 1.501 | -1.191 | -0.334 | -1.252 | -0.349 | -1.230 |
| $h = 7$ | -0.291 | -0.562 | 0.101 | 0.058 | -0.427 | -1.000 | -0.076 | 0.250 | -0.316 |

**Table S3.** DM statistics for the point forecast: ARMA(1,1)-GARCH(1,1) vs $M_1$. Results are in line with Table 2 in the paper. Regarding point forecast, the benchmark model performs much better than both DMA and DMS. Therefore, the null hypothesis of equal forecasting ability is always rejected.

| | $\lambda = \alpha = 0.99$ | | $\lambda = \alpha = 0.95$ | | $\lambda = 0.99$ | $\alpha = 1$ | $\lambda = 1$ | $\alpha = 0.99$ | $\lambda = \alpha = 1$ |
|---|---|---|---|---|---|---|---|---|---|
| | DMA | DMS | DMA | DMS | DMA | DMS | DMA | DMS | BMA |
| $h = 1$ | 11.516 | 10.988 | 12.080 | 11.947 | 11.417 | 10.925 | 11.724 | 11.373 | 11.233 |
| $h = 2$ | 8.217 | 6.887 | 8.314 | 7.217 | 8.277 | 8.943 | 8.271 | 6.882 | 8.334 |
| $h = 3$ | 8.128 | 8.668 | 8.345 | 5.024 | 8.140 | 8.609 | 8.153 | 8.703 | 8.174 |
| $h = 4$ | 8.249 | 5.228 | 8.136 | 4.945 | 8.207 | 5.280 | 8.452 | 5.648 | 8.431 |
| $h = 5$ | 5.905 | 5.158 | 5.873 | 5.058 | 5.878 | 5.152 | 6.125 | 5.313 | 6.094 |
| $h = 6$ | 5.316 | 4.026 | 4.881 | 3.330 | 5.303 | 4.039 | 5.683 | 4.274 | 5.669 |
| $h = 7$ | 8.818 | 8.193 | 8.193 | 7.923 | 8.774 | 8.175 | 9.448 | 8.829 | 9.371 |

| | $\lambda = \alpha = 0.99$ | | $\lambda = \alpha = 0.95$ | | $\lambda = 0.99$ | $\alpha = 1$ | $\lambda = 1$ | $\alpha = 0.99$ | $\lambda = \alpha = 1$ |
|---|---|---|---|---|---|---|---|---|---|
| | DMA | DMS | DMA | DMS | DMA | DMS | DMA | DMS | BMA |
| $h = 1$ | -11.482 | -10.994 | -12.004 | -11.780 | -11.434 | -11.009 | -11.780 | -11.380 | -11.341 |
| $h = 2$ | -7.602 | -8.902 | -7.657 | -8.354 | -7.628 | -8.929 | -7.694 | -8.979 | -7.748 |
| $h = 3$ | -3.901 | -4.878 | -3.694 | -4.823 | -3.922 | -4.882 | -4.000 | -5.038 | -4.030 |
| $h = 4$ | -6.016 | -5.366 | -5.711 | -5.258 | -5.994 | -5.337 | -6.243 | -5.494 | -6.233 |
| $h = 5$ | -3.918 | -3.033 | -3.653 | -2.797 | -3.898 | -3.026 | -4.126 | -3.157 | -4.104 |
| $h = 6$ | -4.354 | -3.832 | -3.983 | -3.271 | -4.338 | -3.839 | -4.671 | -4.064 | -4.651 |
| $h = 7$ | -8.579 | -8.201 | -8.297 | -7.924 | -8.546 | -8.181 | -9.201 | -8.827 | -9.143 |

**Table S4.** DM statistics for the point forecast: ARMA-GARCH(1,1) vs $M_2$. Again, results are in line with Table 2 in the paper and with Table S3. Once more, the benchmark model performs better than both DMA and DMS. Therefore, the null hypothesis of equal forecasting ability is always rejected.

| | $h = 1$ | $h = 2$ | $h = 3$ | $h = 4$ | $h = 5$ | $h = 6$ | $h = 7$ |
|---|---|---|---|---|---|---|---|
| DM Statistic | -2.236 | 0.812 | 0.245 | -0.556 | -1.604 | 0.183 | 0.858 |
| p-value | 0.987 | 0.207 | 0.402 | 0.710 | 0.944 | 0.427 | 0.194 |

**Table S5.** DM statistics for the density forecast: $M_1$ vs $M_2$. Again, the modified DM test is carried out. The alternative hypothesis is that $M_2$ is more accurate than $M_1$. When $h = 1, 5$, $H_1$ is accepted. Regarding the other cases $H_0$ is rejected, but, looking at the p-values, it is not possible to accept the alternative. However, this result does not influence the conclusion of the thesis. Bitcoin still does not reveal any predictive effect over the S&P 500 index.

**Table S6.** Point forecast for $\kappa = 0.97$: $M_1$ vs $M_0$ at the top and $M_2$ vs $M_0$ at the bottom.

| $M_1$ **vs** $M_0$ $\kappa = 0.97$ | | DMA $\lambda = 0.99$ $\alpha = 0.99$ | DMS $\lambda = 0.99$ $\alpha = 0.99$ | DMA $\lambda = 0.95$ $\alpha = 0.95$ | DMS $\lambda = 0.95$ $\alpha = 0.95$ | DMA $\lambda = 0.99$ $\alpha = 1$ | DMS $\lambda = 0.99$ $\alpha = 1$ | DMA $\lambda = 1$ $\alpha = 0.99$ | DMS $\lambda = 1$ $\alpha = 0.99$ | BMA $\lambda = 1$ $\alpha = 1$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $h = 1$ | MAFE | 1.008 | 1.014 | 1.065 | 1.100 | 1.006 | 1.003 | 1.001 | 1.008 | 1.002 |
| $h = 2$ | MAFE | 1.388 | 1.389 | 1.475 | 1.521 | 1.391 | 1.398 | 1.392 | 1.387 | 1.395 |
| $h = 3$ | MAFE | 1.661 | 1.674 | 1.774 | 1.826 | 1.668 | 1.681 | 1.662 | 1.675 | 1.673 |
| $h = 4$ | MAFE | 1.908 | 1.907 | 2.083 | 2.112 | 1.921 | 1.917 | 1.914 | 1.915 | 1.930 |
| $h = 5$ | MAFE | 2.134 | 2.141 | 2.342 | 2.396 | 2.121 | 2.123 | 2.131 | 2.122 | 2.129 |
| $h = 6$ | MAFE | 2.304 | 2.297 | 2.492 | 2.501 | 2.297 | 2.296 | 2.337 | 2.332 | 2.336 |
| $h = 7$ | MAFE | 2.514 | 2.504 | 2.714 | 2.738 | 2.503 | 2.503 | 2.569 | 2.570 | 2.561 |

| $M_2$ **vs** $M_0$ $\kappa = 0.97$ | | DMA $\lambda = 0.99$ $\alpha = 0.99$ | DMS $\lambda = 0.99$ $\alpha = 0.99$ | DMA $\lambda = 0.95$ $\alpha = 0.95$ | DMS $\lambda = 0.95$ $\alpha = 0.95$ | DMA $\lambda = 0.99$ $\alpha = 1$ | DMS $\lambda = 0.99$ $\alpha = 1$ | DMA $\lambda = 1$ $\alpha = 0.99$ | DMS $\lambda = 1$ $\alpha = 0.99$ | BMA $\lambda = 1$ $\alpha = 1$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $h = 1$ | MAFE | 1.003 | 1.005 | 1.053 | 1.075 | 1.005 | 1.008 | 1.004 | 1.004 | 1.016 |
| $h = 2$ | MAFE | 1.386 | 1.387 | 1.460 | 1.481 | 1.382 | 1.382 | 1.394 | 1.393 | 1.399 |
| $h = 3$ | MAFE | 1.659 | 1.669 | 1.769 | 1.786 | 1.661 | 1.666 | 1.660 | 1.666 | 1.666 |
| $h = 4$ | MAFE | 1.905 | 1.907 | 2.089 | 2.114 | 1.900 | 1.901 | 1.911 | 1.915 | 1.912 |
| $h = 5$ | MAFE | 2.129 | 2.137 | 2.338 | 2.407 | 2.109 | 2.109 | 2.139 | 2.131 | 2.133 |
| $h = 6$ | MAFE | 2.301 | 2.301 | 2.496 | 2.523 | 2.304 | 2.307 | 2.346 | 2.339 | 2.342 |
| $h = 7$ | MAFE | 2.510 | 2.505 | 2.726 | 2.724 | 2.503 | 2.503 | 2.569 | 2.563 | 2.618 |

**Table S7.** Point forecast for $\kappa = 0.99$: $M_1$ vs $M_0$ at the top and $M_2$ vs $M_0$ at the bottom.

| $M_1$ **vs** $M_0$ $\kappa = 0.99$ | | DMA $\lambda = 0.99$ $\alpha = 0.99$ | DMS $\lambda = 0.99$ $\alpha = 0.99$ | DMA $\lambda = 0.95$ $\alpha = 0.95$ | DMS $\lambda = 0.95$ $\alpha = 0.95$ | DMA $\lambda = 0.99$ $\alpha = 1$ | DMS $\lambda = 0.99$ $\alpha = 1$ | DMA $\lambda = 1$ $\alpha = 0.99$ | DMS $\lambda = 1$ $\alpha = 0.99$ | BMA $\lambda = 1$ $\alpha = 1$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $h = 1$ | MAFE | 1.006 | 1.000 | 1.063 | 1.089 | 1.013 | 1.011 | 1.014 | 1.013 | 1.026 |
| $h = 2$ | MAFE | 1.388 | 1.392 | 1.474 | 1.467 | 1.393 | 1.402 | 1.403 | 1.403 | 1.408 |
| $h = 3$ | MAFE | 1.651 | 1.668 | 1.760 | 1.798 | 1.651 | 1.659 | 1.667 | 1.655 | 1.658 |
| $h = 4$ | MAFE | 1.903 | 1.904 | 2.104 | 2.133 | 1.913 | 1.916 | 1.920 | 1.917 | 1.934 |
| $h = 5$ | MAFE | 2.121 | 2.125 | 2.368 | 2.401 | 2.118 | 2.118 | 2.151 | 2.150 | 2.188 |
| $h = 6$ | MAFE | 2.285 | 2.280 | 2.499 | 2.528 | 2.287 | 2.287 | 2.369 | 2.358 | 2.366 |
| $h = 7$ | MAFE | 2.481 | 2.466 | 2.699 | 2.702 | 2.483 | 2.482 | 2.619 | 2.621 | 2.611 |

| $M_2$ **vs** $M_0$ $\kappa = 0.99$ | | DMA $\lambda = 0.99$ $\alpha = 0.99$ | DMS $\lambda = 0.99$ $\alpha = 0.99$ | DMA $\lambda = 0.95$ $\alpha = 0.95$ | DMS $\lambda = 0.95$ $\alpha = 0.95$ | DMA $\lambda = 0.99$ $\alpha = 1$ | DMS $\lambda = 0.99$ $\alpha = 1$ | DMA $\lambda = 1$ $\alpha = 0.99$ | DMS $\lambda = 1$ $\alpha = 0.99$ | BMA $\lambda = 1$ $\alpha = 1$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $h = 1$ | MAFE | 1.002 | 1.000 | 1.059 | 1.065 | 1.013 | 1.011 | 1.017 | 1.015 | 1.039 |
| $h = 2$ | MAFE | 1.386 | 1.387 | 1.462 | 1.460 | 1.391 | 1.390 | 1.403 | 1.402 | 1.403 |
| $h = 3$ | MAFE | 1.652 | 1.663 | 1.752 | 1.782 | 1.655 | 1.661 | 1.669 | 1.657 | 1.670 |
| $h = 4$ | MAFE | 1.904 | 1.910 | 2.112 | 2.128 | 1.908 | 1.910 | 1.917 | 1.914 | 1.923 |
| $h = 5$ | MAFE | 2.123 | 2.131 | 2.367 | 2.399 | 2.110 | 2.109 | 2.159 | 2.159 | 2.154 |
| $h = 6$ | MAFE | 2.288 | 2.289 | 2.506 | 2.527 | 2.291 | 2.295 | 2.385 | 2.374 | 2.386 |
| $h = 7$ | MAFE | 2.482 | 2.473 | 2.692 | 2.687 | 2.483 | 2.483 | 2.618 | 2.613 | 2.614 |

**Table S8.** Results for $h = 10$. In this case $\lambda = \alpha = 0.99$ and $\kappa = 0.94$ are fixed. When the forecast horizon increases to 10 days ahead, the outcome gets much worse. The forecast accuracy coming from $M_1$ and $M_2$ is very poor in this case. Whereas the benchmark model is converging to its unconditional mean.

| | DMA | DMS | log PL |
|---|---|---|---|
| $M_0$ | 2.096 | | -1617.275 |
| $M_1$ | 6.320 | 6.339 | -3376.700 |
| $M_2$ | 9.676 | 9.786 | -3644.715 |

**Table S9.** This table refers to the case in which Dow Jones (DJ henceforth) index is substituted to S&P500. The only case analysed is the one with $\alpha = \lambda = 0.99$ and $\kappa = 0.94$. The aim is to understand whether the Bitcoin can have an impact to an index different from the case studied in the paper. The predictors do not differ from previous case. Two models are considered: the first one, former column, which assumes BTC among its predictors, and the second one, latter column, which excludes it. Results have to been compared with Table 2 of the paper. It emerges that using DJ instead of S&P500 improves the result of both point and density forecast for shorter horizon (one or two days ahead).

| | $\lambda = \alpha = 0.99$ | With BTC | | Without BTC | |
|---|---|---|---|---|---|
| | $\kappa = 0.94$ | DMA | DMS | DMA | DMS |
| $h = 1$ | MAFE | 1.576 | 1.589 | 1.571 | 1.577 |
| | MSFE | 0.108 | 0.109 | 0.108 | 0.109 |
| | log PL | -2322.688 | | -2319.416 | |
| $h = 2$ | MAFE | 2.784 | 2.833 | 2.769 | 2.803 |
| | MSFE | 0.192 | 0.194 | 0.192 | 0.193 |
| | log PL | -2772.328 | | -2769.355 | |
| $h = 3$ | MAFE | 3.459 | 3.494 | 3.444 | 3.463 |
| | MSFE | 0.234 | 0.236 | 0.234 | 0.235 |
| | log PL | -2979.653 | | -2975.552 | |
| $h = 4$ | MAFE | 4.148 | 4.140 | 4.136 | 4.123 |
| | MSFE | 0.277 | 0.276 | 0.277 | 0.276 |
| | log PL | -3082.744 | | -3087.355 | |
| $h = 5$ | MAFE | 4.742 | 4.723 | 4.723 | 4.705 |
| | MSFE | 0.321 | 0.319 | 0.321 | 0.318 |
| | log PL | -3183.221 | | -3188.199 | |
| $h = 6$ | MAFE | 5.269 | 5.218 | 5.258 | 5.218 |
| | MSFE | 0.356 | 0.353 | 0.356 | 0.353 |
| | log PL | -3259.553 | | -3263.172 | |
| $h = 7$ | MAFE | 5.786 | 5.738 | 5.780 | 5.737 |
| | MSFE | 0.387 | 0.385 | 0.387 | 0.385 |
| | log PL | -3320.115 | | -3319.068 | |

**Table S10.** This table refers to the case in which Russell 2000 (RUT henceforth) index is substituted to S&P500. The only case analysed is the one with $\alpha = \lambda = 0.99$ and $\kappa = 0.94$. The aim is to understand whether the Bitcoin can have an impact to an index different from the case studied in the paper. The predictors do not differ from previous case. Two models are considered: the first one, former column, which assumes BTC among its predictors, and the second one, latter column, which excludes it. Results have to been compared with Table 2 of the paper. It emerges that results are much worse in this case. Both point and density forecasts shows poor outcomes when compared with the ones in the paper. This means that RUT index, which is a good proxy of the small-cap companies in the US market, does not suffers from changes in the BTC market. It is reasonable to assume that smaller companies in terms of capitalization are not disturbed by fluctuations of cryptocurrencies' price.

| | $\lambda = \alpha = 0.99$ | With BTC | | Without BTC | |
|---|---|---|---|---|---|
| | $\kappa = 0.94$ | DMA | DMS | DMA | DMS |
| $h = 1$ | MAFE | 3.186 | 3.200 | 3.182 | 3.192 |
| | MSFE | 0.193 | 0.194 | 0.194 | 0.194 |
| | log PL | -2905.296 | | -2904.874 | |
| $h = 2$ | MAFE | 4.463 | 4.451 | 4.465 | 4.457 |
| | MSFE | 0.272 | 0.271 | 0.272 | 0.271 |
| | log PL | -3153.631 | | -3150.552 | |
| $h = 3$ | MAFE | 5.400 | 5.459 | 5.395 | 5.449 |
| | MSFE | 0.333 | 0.335 | 0.332 | 0.334 |
| | log PL | -3290.512 | | -3285.752 | |
| $h = 4$ | MAFE | 6.315 | 6.324 | 6.322 | 6.355 |
| | MSFE | 0.392 | 0.393 | 0.391 | 0.393 |
| | log PL | -3392.327 | | -3399.891 | |
| $h = 5$ | MAFE | 7.054 | 7.138 | 7.627 | 7.724 |
| | MSFE | 0.442 | 0.448 | 0.442 | 0.447 |
| | log PL | -3470.732 | | -3474.810 | |
| $h = 6$ | MAFE | 7.634 | 7.701 | 7.627 | 7.724 |
| | MSFE | 0.499 | 0.496 | 0.484 | 0.490 |
| | log PL | -3525.020 | | -3521.943 | |
| $h = 7$ | MAFE | 8.213 | 8.300 | 8.222 | 8.271 |
| | MSFE | 0.511 | 0.503 | 0.527 | 0.528 |
| | log PL | -3577.549 | | -3573.526 | |

## References

Koop, Gary, and Dimitris Korobilis. 2012. Forecasting Inflation Using Dynamic Model Averaging. *International Economic Review* 53: 867–86.

Koop, Gary and Simon Potter. 2004. Forecasting in Dynamic Factor Models using Bayesian Model Averaging. *Econometrics Journal* 7: 550–565.

Raftery, Adrian. E., Miroslav Kárný, and Pavel Ettler. 2010. Online Prediction under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill. *Technometrics* 52: 52–66.