Variable selection in crude oil forecasting

Master thesis - P10

10th semester
Mathematics-Economics
Group: JQ

Supervisor: Charisios Grivas
Submission date: 27th of May

Aalborg University
Department of Mathematical Sciences
Thomas Manns Vej 23
DK-9220 Aalborg Øst

AALBORG UNIVERSITY
STUDENT REPORT

Department of Mathematical Sciences
Mathematics-Economics
Thomas Manns Vej 23
9220 Aalborg Øst
http://math.aau.dk

Title:

Variable selection in crude oil Forecasting: a comparison of penalised and greedy methods

Project:

Master thesis - P10

Project period:

02/02-2026 - 27/05-2026

Project group

Group JQ

Participants:

Jacob Qazizada

Supervisor:

Charisios Grivas

Pages: 93 + title page
Project completed 27/05-2026

Abstract:

This project compares penalised regression methods (LASSO, Elastic Net) with greedy selection algorithms (OCMT, BMT) for forecasting crude oil returns and realised volatility out-of-sample. OCMT (One Covariate at a Time Multiple Testing) selects variables by sequentially testing their statistical significance, while BMT (Boosting Multiple Testing) adopts a more conservative stepwise approach that adds at most one variable per stage, reducing the risk of including spurious predictors when variables are highly correlated.

The results show that sparse models generally improve forecast accuracy relative to a historical benchmark, particularly for volatility. Momentum variables drive return forecasts, while uncertainty indicators are important for volatility. No single method dominates across all evaluation criteria, but penalised regression methods achieve a favourable balance between precision and stability. Despite improved forecast performance, economic gains are limited by transaction costs and model uncertainty, demonstrating that practical implementation of statistical forecasts in oil markets remains challenging.

Preface

This project was written by Group Jacob Qazizada from Aalborg University. The writing period is from the 02/02-2026 to 27/05-2026. The project mainly revolves around variable selection methods.

Referencing in this project is done in the manner *[Number of source]*. This is done in the beginning of a chapter and in between sentences if referenced directly. All sources are listed in chronological order in the bibliography with their corresponding number of source.

Generative AI has been used in the coding and debugging part, as well as grammar in the project.

The group gives their gratitude towards their supervisor Charisios Grivas, for their supervision and advice through the project period.

Signature

Jacob Qazizada

Contents

1 Introduction

Crude oil is a critical input to the global economy, with price fluctuations affecting inflation, industrial production and financial markets [1]. Accurate forecasts of oil price movements and volatility are therefore essential for central banks, governments and investors. However, forecasting remains challenging due to the complex interplay of supply and demand, geopolitical events, financial speculation and technical trading patterns. The recent proliferation of high-frequency data and uncertainty indicators has enriched the information environment but also increased the risk of overfitting.

Previous literature has predicted oil returns using macroeconomic variables and technical indicators [2, 3] and more recently uncertainty indices to forecast oil volatility [4]. Yet one gap persists, while penalised methods like LASSO and Elastic Net have been applied to oil returns, newer greedy selection algorithms, OCMT and BMT [5, 6] have not been systematically compared in crude oil markets. These methods offer alternative variable selection based on sequential hypothesis testing and may perform differently.

This thesis addresses this gap with two complementary exercises. The first replicates and extends [3] by comparing LASSO, Elastic Net, OCMT and BMT on their predictor set, with two additional predictors included, to forecast monthly oil returns. The second extends [4] by applying the same methods to forecast crude oil realised volatility using a comprehensive set of uncertainty indices. The paper uses recursive expanding windows, tune hyperparameters via cross-validation and information criteria and evaluate forecasts using out-of-sample R^2 and directional accuracy tests. In the latter part of the analysis, an economic intuition has been explored.

1.1 Problem statement

How do penalised regression methods compare with greedy selection algorithms in forecasting crude oil returns and volatility out-of-sample?

2 Theory

2.1 The variable selection problem

Consider the linear forecasting model (DGP)

$$y_{t+1} = x_t'\beta^* + \varepsilon_{t+1}, \qquad \mathbb{E}[\varepsilon_{t+1} \mid x_t] = 0,$$

where $x_t \in \mathbb{R}^p$ is a vector of candidate predictors observed at time t and $\beta^* \in \mathbb{R}^p$ is the unknown coefficient vector. Write

$$x_t = (x_{1t}, \ldots, x_{pt})',$$

where x_{it} denotes predictor i observed at time t.

Noise variables inflate the variance of the forecasts. The expected mean squared prediction error can be decomposed as

$$\mathbb{E}\big[(y_{t+1} - x_t'\hat{\beta})^2\big] = \sigma^2 + \mathbb{E}\big[(\hat{\beta} - \beta^*)'\mathbb{E}[x_t x_t'](\hat{\beta} - \beta^*)\big],$$

where $\sigma^2 = \mathbb{E}[\varepsilon_{t+1}^2]$ denotes the irreducible error. The second term captures the additional prediction error arising from estimation uncertainty and depends on how accurately $\hat{\beta}$ approximates β^* as well as on the covariance structure of the predictors.

The decomposition shows that prediction error consists of two parts: the irreducible noise σ^2 and an additional term that depends on how far $\hat{\beta}$ is from β^*. Including many noise variables increases the dimension of $\hat{\beta}$, which tends to raise estimation variance without reducing bias. Even when an inactive variable has $\beta_i^* = 0$, its sample estimate is rarely exactly zero, adding extraneous noise. Variable selection addresses this by restricting the model to a smaller set of predictors, thereby lowering variance and improving out-of-sample forecasts.

To formalise this idea, define the *active set*

$$S = \{i \in \{1, \ldots, p\} : \beta_i^* \neq 0\},$$

and let $s = |S|$ denote its cardinality. Thus S contains the true signals.

Let $\mu_t = \sum_{j \in S} \beta_j^* x_{jt}$ be the linear combination of the true signals. The remaining inactive variables with $\beta_i^* = 0$ are partitioned as follows:

$$\mathcal{P} = \big\{i \in S^c : \mathrm{Cov}(x_{it},\, \mu_t) \neq 0\big\} \qquad \text{(proxy variables)},$$

$$\mathcal{N} = \big\{i \in S^c : \mathrm{Cov}(x_{it},\, \mu_t) = 0\big\} \qquad \text{(noise variables)}.$$

Hence $S^c = \mathcal{P} \cup \mathcal{N}$. Proxies have zero population coefficients in the DGP but appear relevant because they co-move with true signals; noise variables are inactive variables that are uncorrelated with the signal component μ_t.

A central assumption in sparse modelling is *sparsity*. Throughout, the active set S and its cardinality $s = |S|$ are treated as fixed population quantities. Hence, as the total number of candidate predictors increases,

$$\frac{s}{p} \to 0 \qquad \text{as } p \to \infty.$$

In words, only a negligible fraction of the candidate predictors are true signals [7]. In the context of this paper, the focus is on crude oil markets. Many economic and financial variables are correlated and the fundamental drivers of oil price movements are relatively few [1]. Hence, the sparsity assumption is a natural starting point.

Under sparsity, the true model can be viewed as a low-dimensional structure embedded in a higher-dimensional space. The key challenge is that the active set S is unknown and must be recovered from the data, making variable selection a fundamentally ill-posed problem without additional structure.

2.2 LASSO

A natural approach to variable selection is to impose sparsity directly in the estimation problem. The LASSO [8] achieves this by adding an ℓ_1 penalty to the least squares objective, shrinking some coefficients exactly to zero.

Formally, the LASSO estimator is defined as

$$\hat{\beta}^{\text{Lasso}} = \arg\min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{n} \sum_{t=1}^{n} (y_{t+1} - x_t'\beta)^2 + \lambda \|\beta\|_1 \right\},$$

where $\lambda \geq 0$ is a tuning parameter and $\|\beta\|_1 = \sum_{i=1}^{p} |\beta_i|$. The ℓ_1 penalty induces sparsity by shrinking small coefficients to exactly zero, thereby performing variable selection and estimation simultaneously. The LASSO requires the following conditions:

Assumption L1 - Sparsity
The true parameter vector satisfies the sparsity condition, $\lim_{p\to\infty} \frac{s}{p} = 0$.

Assumption L2 - Restricted eigenvalue
There exists $\psi > 0$ such that for all $\delta \in \mathbb{R}^p$ satisfying $\|\delta_{S^c}\|_1 \leq 3\|\delta_S\|_1$,

$$\frac{1}{n}\|X\delta\|_2^2 \geq \psi \|\delta_S\|_2^2.$$

Assumption L3 - Irrepresentable condition
Let $\Sigma = \mathbb{E}[x_t x_t']$ be the population covariance matrix and partition it according to (S, S^c). Then there exists $\eta > 0$ such that

$$\|\Sigma_{S^c S}\Sigma_{SS}^{-1}\operatorname{sign}(\beta_S^*)\|_\infty \leq 1 - \eta.$$

Assumption L4 - Beta-min
The non-zero coefficients are sufficiently large:

$$\min_{i \in S} |\beta_i^*| \gg \sqrt{\frac{\log p}{n}}.$$

Assumption L5 - Regularisation parameter rate
The tuning parameter λ satisfies $\lambda \to 0$ and $\lambda\sqrt{n} \to \infty$ as $n \to \infty$, which is satisfied by $\lambda \asymp \sqrt{\frac{\log p}{n}}$.

Assumption L6 - Weak dependence
The process $\{(x_t, \varepsilon_{t+1})\}$ is stationary and strongly mixing with exponentially decaying mixing coefficients. Moreover,

$$\frac{1}{n}X'X \xrightarrow{p} \Sigma, \qquad \Sigma = \mathbb{E}[x_t x_t'],$$

where Σ is positive definite.

These assumptions are strong and may fail in many practical settings, particularly when inactive variables are highly correlated with the true signals or signals are weak.

To address some of these limitations, the Elastic Net [9] adds an ℓ_2 penalty to the LASSO objective.

2.3 Elastic Net

The Elastic Net estimator [9] is defined with a mixing parameter $\alpha \in [0, 1]$ as

$$\hat{\beta}^{\mathrm{EN}} = \arg\min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{n}\sum_{t=1}^{n}(y_{t+1} - x_t'\beta)^2 + \lambda\big(\alpha\|\beta\|_1 + (1-\alpha)\|\beta\|_2^2\big) \right\}.$$

When $\alpha = 1$ the Elastic Net reduces to the LASSO; when $\alpha = 0$ it becomes ridge regression. The ℓ_2 penalty stabilises the solution when predictors are highly correlated and encourages a grouping effect: coefficients of correlated variables tend to be similar in magnitude, so

such variables are either selected or omitted together – unlike the LASSO, which often picks only one from a correlated group.

The Elastic Net relies on assumptions analogous to the LASSO, including sparsity, restricted eigenvalue and beta-min conditions, which are Assumptions L1, L2, L4, L5 and L6. The main difference is a modified irrepresentable condition that accounts for the ridge term.

Assumption EN - Irrepresentable condition
Let $\Sigma = \mathbb{E}[x_t x_t']$ and partition it according to (S, S^c). There exists $\eta > 0$ such that

$$\left\|\Sigma_{S^cS}(\Sigma_{SS} + \gamma I)^{-1} \operatorname{sign}(\beta_S^*)\right\|_\infty \leq 1 - \eta,$$

where $\gamma = \lambda(1-\alpha)/n$. This condition is less restrictive than the LASSO version because the ridge term γI stabilises the inversion.

Even with these assumptions, the Elastic Net does not guarantee exact recovery of the true active set, especially when proxies are present. Its main advantage over the LASSO is improved stability and prediction accuracy in the presence of multicollinearity.

The next method, OCMT, takes a fundamentally different approach that avoids many of these design-based assumptions altogether.

2.4 OCMT

The OCMT procedure [5] differs from penalised regression methods by selecting variables through sequential hypothesis testing with a multiple testing correction. Instead of solving a global optimisation problem, OCMT evaluates each predictor individually using its associated t-statistic.

Let $q \in (0,1)$ denote the nominal significance level and let p be the number of candidate predictors. The critical value is stage-dependent: a more liberal threshold is used in the first stage to avoid omitting weak signals and a stricter threshold is applied in subsequent stages. Define

$$c_p^{(1)}(q, \delta_1) = \Phi^{-1}\left(1 - \frac{q}{2c_q\, p^{\delta_1}}\right), \qquad c_p^{(k)}(q, \delta_2) = \Phi^{-1}\left(1 - \frac{q}{2c_q\, p^{\delta_2}}\right), \quad k \geq 2,$$

where Φ^{-1} is the inverse standard normal cumulative distribution function, $c_q > 0$, $\delta_1 \geq 1$ and $\delta_2 \geq 2$. In this paper we set $c_q = 1$. A predictor is selected at stage k if its absolute t-statistic, conditional on previously selected variables, exceeds $c_p^{(k)}(q, \delta_k)$.

The procedure is implemented sequentially. In the first stage, all candidate predictors are tested individually. In subsequent stages, tests are conducted conditional on the variables

already selected, so that each new variable is evaluated based on its incremental contribution. Larger values of the tuning exponents make the procedure more conservative and reduce false positives.

The OCMT procedure relies on the following assumptions.

Assumption C1 - Signal strength
For all $i \in S$, $|\beta_i^*| \geq c_\beta > 0$.

Assumption C2 - Design regularity
The eigenvalues of $\Sigma = \mathbb{E}[x_t x_t']$ are bounded away from zero and infinity.

Assumption C3 - Weak dependence
The process $\{(x_t, \varepsilon_{t+1})\}$ is stationary and strongly mixing with exponentially decaying mixing coefficients.

Assumption C4 - Exponential tails
There exist constants $C_1, C_2, C_3, C_4 > 0$ and $s_x, s_u > 0$ such that for all $\alpha > 0$,

$$\sup_{i,t} \Pr(|x_{it}| > \alpha) \leq C_1 \exp(-C_2 \alpha^{s_x}), \qquad \sup_{t} \Pr(|\varepsilon_{t+1}| > \alpha) \leq C_3 \exp(-C_4 \alpha^{s_u}).$$

Assumption C5 - Dimensionality
The number of predictors satisfies $p = O(n^\kappa)$ for some $\kappa > 0$.

Assumption C6 - Martingale difference
The error process $\{\varepsilon_{t+1}\}$ satisfies $\mathbb{E}[\varepsilon_{t+1} \mid \mathcal{F}_t] = 0$ and $\mathbb{E}[\varepsilon_{t+1}^2 \mid \mathcal{F}_t] = \sigma^2$, where $\mathcal{F}_t = \sigma(\{x_s, \varepsilon_s\}_{s \leq t})$ is the information set up to time t.

Assumption C7 - Eigenvalue condition for selected predictors
Let $\hat{S}$ denote the set of predictors selected by OCMT and let $\ell = |\hat{S}|$ be its cardinality. Define the selected predictor vector

$$x_{t,\hat{S}} = (x_{it})_{i \in \hat{S}} \in \mathbb{R}^\ell,$$

and let

$$\Sigma_{\hat{S}\hat{S}} = \mathbb{E}[x_{t,\hat{S}} x_{t,\hat{S}}']$$

denote its population covariance matrix. Let

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_\ell$$

be the eigenvalues of $\Sigma_{\hat{S}\hat{S}}$. There exist constants $c_1 > 0$, $C_0 > 0$, $c_2 \geq 0$ and a finite integer $M \geq 0$ such that:

1. $\lambda_i \geq c_1$ for all $i = 1, \ldots, \ell$;

2. $\lambda_i \leq C_0$ for all $i = 1, \ldots, \ell - M$;
3. $\lambda_i \leq c_2 \ell$ for all $i = \ell - M + 1, \ldots, \ell$.

Conditions (i) and (ii) imply that all eigenvalues are bounded away from zero and that all but the largest M eigenvalues are uniformly bounded above. Condition (iii) allows the largest M eigenvalues to grow at most linearly with ℓ.

Recall that S denotes the active set of true signals, $\mathcal{N} \subset S^c$ the set of pure noise variables and $\hat{S}$ the set selected by OCMT. To evaluate selection performance, define

$$\text{TPR} = \frac{|\hat{S} \cap S|}{|S|}, \qquad \text{FPR} = \frac{|\hat{S} \cap \mathcal{N}|}{|\mathcal{N}|}, \qquad \text{FDR} = \mathbb{E}\left[\frac{|\hat{S} \cap S^c|}{|\hat{S}|}\right],$$

with the convention $0/0 = 0$. The TPR measures the fraction of true signals correctly selected, the FPR measures the fraction of pure noise variables incorrectly selected and the FDR measures the expected proportion of inactive variables among the selected predictors.

These quantities summarise the variable selection performance of OCMT. An effective selection procedure should asymptotically recover all true signals while avoiding the inclusion of irrelevant predictors. The following theorem shows that, under Assumptions C1-C6, OCMT achieves these properties for true signals and pure noise variables.

Theorem 2.4.1. Suppose Assumptions C1-C6 hold. Let $\delta_1 \geq 1$ and $\delta_2 \geq 2$. Then

$$\text{TPR} = \frac{|\hat{S} \cap S|}{|S|} \xrightarrow{p} 1, \qquad \text{FPR} = \frac{|\hat{S} \cap \mathcal{N}|}{|\mathcal{N}|} \xrightarrow{p} 0.$$

Proof. This proof follows the main ideas of Theorem 1 in [5] and its Online Supplement (Lemmas A2, A10 and the proof on pages 1504–1508, short versions can be found in Appendix A).

Step 0 - Preliminaries

Define the signal component

$$\mu_t = x_{t,S}' \beta_S^* = \sum_{j \in S} \beta_j^* x_{jt}.$$

For each predictor $i = 1, \ldots, p$, let

$$x_i = (x_{i1}, \ldots, x_{in})'$$

denote the $n \times 1$ vector of observations on predictor i and let

$$\mu = (\mu_1, \ldots, \mu_n)'.$$

At each stage of OCMT, inference is conducted conditional on the predictors selected in previous stages. Let Z denote the $n \times r$ matrix of previously selected predictors and define the residual-maker matrix

$$M_Z = I - Z(Z'Z)^{-1}Z'.$$

For any predictor x_{it}, define its net effect as

$$\theta_i := \mathbb{E}\left[n^{-1}x_i'M_Z\mu\right].$$

In the first stage, where Z is empty and $M_Z = I$, this reduces to

$$\theta_i = \mathrm{Cov}(x_{it}, \mu_t).$$

Step 1 – True signals are selected with probability approaching 1.

Take $i \in S$. Under Assumptions C1 and C2, $\theta_i \neq 0$ and is bounded away from zero. Let $\hat{\sigma}_i$ denote the estimated standard deviation of $n^{-1/2}x_i'M_Z\varepsilon$. Decompose the OCMT t-statistic as

$$t_i = \underbrace{\frac{n^{-1/2}x_i'M_Z\mu}{\hat{\sigma}_i\sqrt{n^{-1}x_i'M_Zx_i}}}_{A_i} + \underbrace{\frac{n^{-1/2}x_i'M_Z\varepsilon}{\hat{\sigma}_i\sqrt{n^{-1}x_i'M_Zx_i}}}_{B_i}.$$

By Assumptions C2 and C6,

$$n^{-1}x_i'M_Z\mu \xrightarrow{p} \theta_i \neq 0, \qquad n^{-1}x_i'M_Zx_i \xrightarrow{p} c_i > 0,$$

for some finite constant c_i. Hence,

$$A_i = \sqrt{n}\,\frac{\theta_i + o_p(1)}{\hat{\sigma}_i\sqrt{c_i + o_p(1)}} = \sqrt{n}\,C_i + o_p(\sqrt{n}),$$

for some finite constant $C_i \neq 0$. Therefore A_i diverges at rate $\sqrt{n}$.

For B_i, note that x_{it} is $\mathcal{F}_t$-measurable and ε_{t+1} is a martingale difference by Assumption C6. Hence $\{x_{it}\varepsilon_{t+1}\}$ is a martingale difference. Under the exponential tail condition (Assumption C4), the martingale CLT in Appendix A.0.1 implies that the scaled sum $n^{-1/2}x_i'M_Z\varepsilon$ is $O_p(1)$. Since the denominator converges to a positive constant, $B_i = O_p(1)$. Consequently $|t_i| \xrightarrow{p} \infty$ at rate $\sqrt{n}$.

By Appendix Lemma A.0.7, the critical value satisfies $c_p^{(k)}(q, \delta_k) = O(\sqrt{\log p})$. By Assumption C5, $p = O(n^\kappa)$, so $c_p^{(k)}(q, \delta_k) = O(\sqrt{\log n})$. Because $\sqrt{n}$ dominates $\sqrt{\log n}$, $\Pr(i \in \hat{S}) \to 1$. Since $|S| = s$ is finite, $\Pr(S \subseteq \hat{S}) \to 1$.

Step 2 – Noise variables are not selected.

Let $j \in \mathcal{N}$. Then

$$\beta_j^* = 0, \qquad \theta_j = 0.$$

Hence x_{jt} has no net predictive effect on μ_t.

Under Assumptions C3-C6, Appendix Lemma A.0.8 establishes exponential tail bounds for the null t-statistics. The OCMT critical values are given by

$$c_p^{(k)}(q, \delta_k) = \Phi^{-1}\left(1 - \frac{q}{2c_q p^{\delta_k}}\right),$$

whose asymptotic order is characterised in Appendix Lemma A.0.7. Using these results, the proof of Theorem 1 in [5, pp. 1504–1508] shows that the probability of selecting pure noise variables vanishes asymptotically. Therefore,

$$\Pr(\exists j \in \mathcal{N} : \, j \in \hat{S}) \to 0,$$

or equivalently,

$$\Pr(\hat{S} \cap \mathcal{N} = \emptyset) \to 1.$$

Step 3 – Hidden signals

If a true signal has $\theta_i = 0$ in the first stage, it is called hidden. Appendix Lemma A.0.9 guarantees that in a later stage $j \geq 2$ its net effect becomes nonzero. The same argument as in Step 1 (now with a stricter critical value, since $\delta_2 \geq 2$) shows that the hidden signal is selected with probability tending to 1. By Appendix Proposition A.0.10 and Appendix Lemma A.0.11, the total number of stages is bounded, so $\Pr(S \subseteq \hat{S}) \to 1$ remains true.

From Steps 1 and 3, $\Pr(S \subseteq \hat{S}) \to 1$; from Step 2, $\Pr(\hat{S} \cap \mathcal{N} = \emptyset) \to 1$. Therefore TPR $\xrightarrow{p} 1$ and FPR $\xrightarrow{p} 0$. □

The theorem implies that OCMT asymptotically selects all true signals while excluding pure noise variables. However, proxies may also be selected. The resulting model is therefore an *approximating model*. Consequently, the FDR need not vanish and the selected model need not coincide with the true active set, even asymptotically. This feature motivates the refinement introduced by the BMT procedure. A key advantage of OCMT is that it avoids the strong design assumptions required by the previously discussed penalised regression methods.

2.5 BMT

The BMT procedure [6] refines OCMT by selecting at most one variable per stage. At each step, it considers the candidate predictor with the largest absolute t-statistic conditional on the variables already selected and includes it only if it exceeds the critical value.

Once included, its effect is partialled out from the remaining predictors and the procedure continues until no variable satisfies the threshold.

BMT uses the same stage-dependent critical values as OCMT:

$$c_p^{(1)}(q,\delta_1) = \Phi^{-1}\left(1 - \frac{q}{2c_q\, p^{\delta_1}}\right), \qquad c_p^{(k)}(q,\delta_2) = \Phi^{-1}\left(1 - \frac{q}{2c_q\, p^{\delta_2}}\right), \quad k \geq 2,$$

with $c_q = 1$ in this paper. A predictor is selected at stage k if its absolute conditional t-statistic exceeds $c_p^{(k)}(q,\delta_k)$.

Partialling out selected variables ensures that subsequent t-statistics reflect only incremental contributions, so that proxies and noise variables lose significance once the true signals are included.

BMT builds on the same probabilistic framework as OCMT. In particular, Assumptions C2–C6 are maintained. In addition, the following conditions are required:

Assumption B1 - Signal strength
There exists a fixed constant $c_0 > 0$ independent of n and p such that for every true signal $i \in S$,

$$|\beta_i^*| \geq c_0.$$

Assumption B2 - Signal-proxy dominance
There exists a constant $c_\nu > 0$ such that for any true signal $i \in S$, any inactive variable $j \in S^c$ and any subset $\mathcal{A} \subseteq \{1,\ldots,p\}$ with $i \notin \mathcal{A}$,

$$\nu_{i|\mathcal{A}} - \nu_{j|\mathcal{A}} \geq c_\nu \sqrt{n},$$

where $\nu_{k|\mathcal{A}}$ denotes the non-centrality parameter of the t-statistic for variable k conditional on the set $\mathcal{A}$.

Under these conditions, BMT achieves the *oracle property* [10], which requires both selection consistency and asymptotic efficiency. The following theorem formalises this result.

Theorem 2.5.1 (BMT oracle property)**.** Under Assumptions C2–C6, B1 and B2, with $\delta_1 \geq 1$ in the first stage and $\delta_2 > \delta_1$ in later stages, the BMT procedure yields a post-selection OLS estimator $\tilde{\beta}$ such that

1. $\Pr(\hat{S} = S) \to 1$,
2. $\sqrt{n}(\tilde{\beta}_S - \beta_S^*) \xrightarrow{d} \mathcal{N}\big(0, \sigma^2 \Sigma_{SS}^{-1}\big)$.

Here, $\hat{S}$ is the set selected by BMT, $\sigma^2 = \mathbb{E}[\varepsilon_{t+1}^2]$ is the error variance and $\Sigma_{SS} = \mathbb{E}[x_{t,S} x_{t,S}']$ is the covariance matrix of the signals.

Thus, unlike OCMT, achieving the oracle property implies exact recovery of the active set together with asymptotically efficient estimation.

Proof. A complete rigorous proof is given in [6]. Here we outline the main idea.

Step 1. Selection consistency
Define the events

$$\mathcal{E}_k = \big\{A_k \subseteq S \text{ and } |A_k| = k\big\}, \qquad k = 0, 1, \ldots, s,$$

where A_k is the set selected after k stages ($A_0 = \emptyset$). We show $\Pr(\mathcal{E}_k) \to 1$ for all $k \leq s$ by induction.

Base. $\mathcal{E}_0$ holds with probability 1.

Induction step. Assume $\Pr(\mathcal{E}_{k-1}) \to 1$ and $k - 1 < s$. Condition on $\mathcal{E}_{k-1}$. Then $A_{k-1} \subsetneq S$ and $|A_{k-1}| = k - 1$.

For any true signal $i \in S \setminus A_{k-1}$ and any inactive variable $j \notin S$, Assumption B2 yields

$$\nu_{i|A_{k-1}} - \nu_{j|A_{k-1}} \geq c_\nu \sqrt{n}, \qquad c_\nu > 0.$$

By the exponential inequalities established in Appendix Lemma A.0.8, there exist constants $C, \omega > 0$ such that for any threshold b,

$$\Pr\big(|t_{r|A_{k-1}} - \nu_{r|A_{k-1}}| > b\big) \leq Ce^{-\omega b^2}.$$

Taking $b = \frac{c_\nu}{4}\sqrt{n}$, we obtain

$$\Pr\left(\max_{j \notin S} |t_{j|A_{k-1}}| \geq \max_{i' \in S \setminus A_{k-1}} |t_{i'|A_{k-1}}|\right) \leq p \cdot Ce^{-\omega(c_\nu/4)^2 n}.$$

Because $p = O(n^\kappa)$ by Assumption C5 and $\omega(c_\nu/4)^2 > 0$, the right-hand side tends to 0 as $n \to \infty$. Consequently,

$$\lim_{n \to \infty} \Pr\left(\arg\max_{r \notin A_{k-1}} |t_{r|A_{k-1}}| \in S \;\middle|\; \mathcal{E}_{k-1}\right) = 1.$$

Now, let i denote that true signal. By Assumptions B1 and C2, the partial coefficient of x_i conditional on A_{k-1} is bounded away from zero and its standard error is of order $n^{-1/2}$. Hence

$$\nu_{i|A_{k-1}} \asymp \sqrt{n}.$$

and therefore

$$|t_{i|A_{k-1}}| \geq \nu_{i|A_{k-1}} - O_p(1) \asymp \sqrt{n}.$$

The critical value satisfies $c_p^{(k)} \sim \sqrt{2\delta_k \log p} = o(\sqrt{n})$ because $p = O(n^\kappa)$ again by Assumption C5. Hence,

$$\Pr\Big(|t_{i|A_{k-1}}| > c_p^{(k)} \mid \mathcal{E}_{k-1}\Big) \to 1.$$

Thus BMT selects i, giving $A_k = A_{k-1} \cup \{i\}$ and consequently $\Pr(\mathcal{E}_k) \to 1$.

By induction, $\Pr(\mathcal{E}_s) \to 1$. Since $|S| = s$, the conditions $A_s \subseteq S$ and $|A_s| = s$ imply $A_s = S$.

We now argue that no inactive variable is selected after S is exhausted. Condition on $A_s = S$. For any inactive variable $j \notin S$, the corresponding conditional non-centrality parameter satisfies

$$\nu_{j|S} = 0.$$

By Appendix Lemma A.0.8, the conditional t-statistics of noise variables satisfy an exponential tail inequality. Combined with the stage-dependent critical values

$$c_p^{(k)}(q, \delta_k) = \Phi^{-1}\left(1 - \frac{q}{2c_q p^{\delta_k}}\right),$$

the proof of Theorem 1 in [5, pp. 1504–1508] establishes that

$$\Pr(\exists j \notin S : \ j \text{ selected} \mid A_s = S) \to 0.$$

Moreover, Assumption B2 implies that for every stage $k < s$,

$$\max_{i \in S \setminus A_k} \nu_{i|A_k} - \max_{j \in S^c} \nu_{j|A_k} \geq c_\nu \sqrt{n},$$

with $c_\nu > 0$. Hence, with probability approaching one,

$$\arg\max_{r \notin A_k} |t_{r|A_k}| \in S \setminus A_k,$$

so a true signal is selected before any proxy variable.

After all true signals have been selected, i.e. conditional on $A_s = S$, every inactive variable $j \in S^c$ satisfies

$$\nu_{j|S} = 0.$$

Thus proxy variables become asymptotically indistinguishable from pure noise variables once conditioning on the true active set. Hence,

$$\Pr(\hat{S} = S) \to 1.$$

Step 2 - Asymptotic efficiency
In this part of the proof, we invoke other theorems that are presented in Appendix A.

On the event $\hat{S} = S$, we have $\tilde{\beta}_S = \hat{\beta}_S^{\text{OLS}}$, where $\hat{\beta}_S^{\text{OLS}} = (X_S' X_S)^{-1} X_S' y$ is the OLS estimator using only the signals. Hence, it is sufficient to determine the limit distribution of $\hat{\beta}_S^{\text{OLS}}$.

Write the standard OLS expansion:

$$\sqrt{n}(\hat{\beta}_S^{\text{OLS}} - \beta_S^*) = \Big(\frac{1}{n}X_S'X_S\Big)^{-1}\frac{1}{\sqrt{n}}X_S'\varepsilon.$$

Recall that Σ_{SS} denotes the population covariance matrix of the true signals. By Assumptions C3-C4, the weak law of large numbers for mixing processes (Theorem A.0.2) yields

$$\frac{1}{n}X_S'X_S \xrightarrow{p} \mathbb{E}[x_{t,S}x_{t,S}'] =: \Sigma_{SS}.$$

Define $z_t := x_{t,S}\varepsilon_{t+1}$, where $z_t \in \mathbb{R}^s$. By Assumption C6, $\{\varepsilon_{t+1}\}$ is a martingale difference sequence with $\mathbb{E}[\varepsilon_{t+1} \mid \mathcal{F}_t] = 0$ and constant conditional variance $\mathbb{E}[\varepsilon_{t+1}^2 \mid \mathcal{F}_t] = \sigma^2$. Hence $\{z_t\}$ is a martingale difference sequence. Assumptions C3-C4 imply that $\{z_t\}$ is stationary, strongly mixing and has finite second moments. Consequently, the mixing central limit theorem (Theorem A.0.3) gives

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t \xrightarrow{d} \mathcal{N}\big(0,\ \mathbb{E}[z_t z_t']\big).$$

Using the constant conditional variance from Assumption C6,

$$\mathbb{E}[z_t z_t'] = \mathbb{E}[x_{t,S}x_{t,S}'\varepsilon_{t+1}^2] = \sigma^2\mathbb{E}[x_{t,S}x_{t,S}'] = \sigma^2\Sigma_{SS}.$$

Since $\frac{1}{n}X_S'X_S \xrightarrow{p} \Sigma_{SS}$ and Σ_{SS} is nonsingular by Assumption C2, the continuous mapping theorem (Theorem A.0.4) implies $(\frac{1}{n}X_S'X_S)^{-1} \xrightarrow{p} \Sigma_{SS}^{-1}$.

Slutsky's theorem (Theorem A.0.5) then yields

$$\sqrt{n}(\hat{\beta}_S^{\text{OLS}} - \beta_S^*) \xrightarrow{d} \mathcal{N}\big(0,\ \sigma^2\Sigma_{SS}^{-1}\big).$$

Finally, because $\Pr(\tilde{S} = S) \to 1$ and $\tilde{\beta}_{\tilde{S}} = \hat{\beta}_S^{\text{OLS}}$ on this event, Theorem A.0.6 implies that the unconditional distribution of $\sqrt{n}(\tilde{\beta}_S - \beta_S^*)$ converges to the same limit. □

Thus, BMT combines the screening power of multiple testing with a stagewise selection rule that excludes proxies, achieving the oracle property. The next chapter evaluates its empirical performance.

2.6 Weak assumptions in practice

Although the previous sections establish asymptotic properties for each variable selection method, several assumptions may be difficult to justify empirically in macroeconomic and

financial forecasting applications. In crude oil markets, many candidate predictors are driven by common global factors and exhibit substantial persistence, collinearity and time variation [1, 2, 3, 4]. Consequently, the conditions required for consistent variable selection and asymptotic efficiency may fail in practice.

For the LASSO and Elastic Net, the most restrictive assumptions are the irrepresentable conditions (Assumption L3 and Assumption EN). These conditions restrict the dependence between inactive predictors and the active set. In particular, inactive predictors cannot be too strongly represented by linear combinations of the signals. In macroeconomic applications, however, predictors often share latent common components, generating substantial cross-sectional dependence between signals and inactive predictors. When the irrepresentable condition fails, inactive predictors may remain highly correlated with the residual variation generated by omitted signals, preventing exact recovery of the active set. Consequently, the probability of false selection need not vanish asymptotically. This issue is particularly severe for the LASSO, since Assumption L3 is essentially necessary for variable selection consistency. Although the ridge component in the Elastic Net weakens the condition through the regularised inverse $(\Sigma_{SS} + \gamma I)^{-1}$, strong collinearity may still prevent consistent separation of signals from inactive predictors.

Assumption L2 is also difficult to verify empirically. The restricted eigenvalue condition requires that there exists a constant $\psi > 0$ such that

$$\frac{1}{n}\|X\delta\|_2^2 \geq \psi\|\delta_S\|_2^2$$

for sparse deviations δ. Hence, sparse linear combinations of predictors must retain sufficient variation after accounting for dependence among predictors. In empirical macroeconomic datasets, predictors are often highly persistent and strongly correlated due to common economic factors. Such dependence can make the lower bound ψ arbitrarily small, weakening identification of the active set. When this occurs, distinct sparse coefficient vectors may generate very similar fitted values, reducing the stability of variable selection [7].

The beta-min condition (Assumption L4) may also be restrictive in forecasting environments. This assumption requires the smallest signal coefficient to dominate estimation noise at rate $\sqrt{\log p/n}$, ensuring that true signals are asymptotically distinguishable from zero. As mentioned several times, in financial applications, predictive relationships are often weak, unstable and local to particular subsamples. As a result, economically meaningful signals may violate the beta-min condition even when they improve forecast accuracy. In such settings, shrinkage estimators may asymptotically set non-zero coefficients equal to zero, implying failure of selection consistency despite satisfactory predictive performance.

The assumptions underlying OCMT are weaker with respect to the predictor dependence structure because the method avoids sparse inversion of the covariance matrix and does not require an irrepresentable condition. This is an important advantage when predictors are strongly correlated. However, the consistency arguments for OCMT rely heavily on

the probabilistic behaviour of the sequential t-statistics. In the proof of 2.4.1, selection consistency follows from an asymptotic separation argument: signal t-statistics diverge at rate $\sqrt{n}$, whereas the critical values grow only at rate $\sqrt{\log p}$.

This separation depends critically on Assumptions C3-C6. Assumption C6 imposes a martingale difference structure together with conditional homoskedasticity, while Assumption C4 requires exponential tail decay. Again, crude oil prices and futures frequently exhibit volatility clustering, conditional heteroskedasticity and heavy tails. Such features may invalidate the martingale limit approximations and exponential probability bounds used to control the null t-statistics. In finite samples, inactive predictors may therefore generate spuriously large t-statistics, weakening the false positive control established by OCMT.

The dimensionality condition in Assumption C5 is comparatively mild since it allows p to grow polynomially with n. Nevertheless, when signals are weak, the finite-sample separation between signal and inactive predictor t-statistics may be insufficient even if the asymptotic rate conditions hold.

For BMT, the strongest requirement is Assumption B2. The oracle property relies on a stronger separation condition requiring the non-centrality parameter of every remaining signal to dominate that of every inactive predictor by order $\sqrt{n}$ at each stage of the procedure. This ensures that signals asymptotically enter the model before proxy predictors. In our application, however, many predictors respond to common economic shocks and may therefore generate t-statistics similar to those of the true signals. When this separation condition fails, proxy predictors may enter before signals, weakening exact recovery of the active set.

Assumption B1 may also be restrictive because it imposes a fixed lower bound on signal strength. In practice, predictive relationships may vary across subsamples due to structural breaks, changing market conditions or parameter instability. Under such circumstances, signals may become asymptotically weak, preventing exact recovery even when the resulting forecasts remain economically informative.

Taken together, the assumptions reveal a trade-off across methods. Penalised regression methods rely primarily on structural assumptions concerning the geometry of the predictor space, whereas OCMT and BMT rely more heavily on probabilistic assumptions governing the behaviour of sequential test statistics. The LASSO and Elastic Net may therefore become unreliable under strong predictor dependence, while OCMT and BMT may become sensitive to weak signals, heavy tails and deviations from the assumed error structure. Since crude oil markets are characterised by persistent comovement, structural instability and periods of elevated volatility [1, 2], no method is guaranteed to satisfy all assumptions simultaneously. Consequently, empirical forecast performance provides an important complement to asymptotic theory.

2.7 Forecast evaluation

Let $\mathcal{F}_t$ denote the information set available at time t. For a given forecasting model, the one-step-ahead forecast is defined as

$$\hat{y}_{t+1} = \mathbb{E}(y_{t+1} \mid \mathcal{F}_t),$$

where model parameters are estimated recursively using an expanding estimation window.

Following [3], forecast performance is evaluated relative to the historical mean benchmark forecast,

$$\bar{y}_{t+1} = \frac{1}{t}\sum_{s=1}^{t} y_s,$$

which is the recursive average of all observations available up to time t.

Point forecast accuracy. The mean squared prediction error (MSPE) is defined as

$$\text{MSPE} = \frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}}(y_{t+1} - \hat{y}_{t+1})^2,$$

where $\mathcal{T}$ denotes the set of out-of-sample periods.

Forecast performance is also evaluated using the out-of-sample coefficient of determination,

$$R^2_{\text{OOS}} = 1 - \frac{\sum_{t\in\mathcal{T}}(y_{t+1} - \hat{y}_{t+1})^2}{\sum_{t\in\mathcal{T}}(y_{t+1} - \bar{y}_{t+1})^2}.$$

A positive value of R^2_{OOS} indicates that the forecasting model outperforms the historical mean benchmark.

Directional accuracy. Let $\Delta y_{t+1} = y_{t+1} - y_t$ denote the change. The success ratio is

$$\text{Success ratio} = \frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}} \mathbf{1}\big(\operatorname{sign}(\Delta y_{t+1}) = \operatorname{sign}(\Delta \hat{y}_{t+1})\big).$$

The Pesaran–Timmermann test [11] tests whether the success ratio is significantly greater than 0.5. The test statistic is asymptotically standard normal under the null of independence.

Clark-West test. Because the historical mean is nested in any model that adds predictors, the Clark-West test [12] is appropriate. Define

$$d_{t+1} = (y_{t+1} - \bar{y}_{t+1})^2 - \big((y_{t+1} - \hat{y}_{t+1})^2 - (\bar{y}_{t+1} - \hat{y}_{t+1})^2\big).$$

The Clark–West statistic is the t-statistic from regressing d_{t+1} on a constant. Under the null that the benchmark MSPE is less than or equal to the model's MSPE, the statistic asymptotically follows a standard normal distribution. A positive significant value indicates that the model improves forecast accuracy.

3 Application

3.1 Presenting the data

Extending the sample period of [3, 2019], this study constructs a monthly dataset from February 1986 to July 2025 to forecast crude oil returns. The dataset combines oil spot prices, oil futures prices and volumes, macroeconomic variables from the Federal Reserve Economic Data (FRED) and technical indicators derived from crude oil futures. All series are aligned at monthly frequency.

For all models considered, training in-sample period starts in February 1986 and ends in December 2007, while the out-of-sample evaluation period runs from January 2008 to July 2025.

3.1.1 Oil price and return

The nominal West Texas Intermediate (WTI) crude oil spot price is obtained from the U.S. Energy Information Administration (EIA, series *RWTCm*). Daily observations are converted to monthly frequency by taking the last trading day of each month. The real oil price is obtained by deflating the nominal price with the Consumer Price Index (CPI, FRED series CPIAUCSL):

$$P_t^{\text{real}} = \frac{P_t^{\text{nominal}}}{CPI_t} \times 100.$$

Monthly real oil returns are computed as simple returns:

$$r_t = \frac{P_t^{\text{real}} - P_{t-1}^{\text{real}}}{P_{t-1}^{\text{real}}}.$$

The dependent variable in the following forecasting exercises is the one-month-ahead return r_{t+1}.

3.1.2 Macroeconomic predictors

A set of 16 macroeconomic variables is collected from FRED. The series include:

- Interest rates: 3-month Treasury bill rate (TB3MS), 10-year Treasury yield (GS10) and Moody's seasoned Baa and Aaa corporate bond yields (BAA, AAA).

- Inflation: monthly CPI inflation (infl_m) constructed from CPIAUCSL.
- Financial market variables: S&P 500 monthly variance (SVOL, computed from daily returns), the economic policy uncertainty index (epu) and the Kilian global economic activity index (kilian).
- Real activity: industrial production growth (ip_growth), first difference of the unemployment rate (unemp_diff), total capacity utilization (TCU) and the Chicago Fed National Activity Index (cfnai).
- Monetary aggregate: M2 money supply growth (m2_growth).
- Oil-specific supply variables (from the EIA): U.S. crude oil production, ending stocks and imports.

All macroeconomic series are lagged by one month to avoid look-ahead bias. The only exception is the CPI, which is used contemporaneously for deflation.

3.1.3 Oil supply variables

As mentioned above, monthly data on U.S. crude oil production, imports and ending stocks are taken from the EIA. These are transformed into monthly growth rates:

$$\text{growth}_t = 100\left(\log X_t - \log X_{t-1}\right),$$

for production, imports and stocks respectively. The growth rates are then lagged by one month to serve as predictors.

3.1.4 Technical indicators

18 technical indicators are constructed from the monthly WTI futures price (Price_Futures) and trading volume (Vol) obtained from the New York Mercantile Exchange (NYMEX). All indicators are binary signals:

Moving average (MA) signals compare a short-term moving average (windows $s = 1, 2, 3$ months) with a long-term moving average (windows $\ell = 9, 12$ months):

$$MA_{s,\ell,t} = \mathbf{1}\big(MA_{s,t} \geq MA_{\ell,t}\big), \quad \text{where} \quad MA_{j,t} = \frac{1}{j}\sum_{i=0}^{j-1} P_{t-i}^{\text{futures}}.$$

Momentum (MOM) signals indicate whether the current futures price is above its level m months earlier ($m \in \{1, 2, 3, 6, 9, 12\}$):

$$MOM_{m,t} = \mathbf{1}\Big(P_t^{\text{futures}} \geq P_{t-m}^{\text{futures}}\Big).$$

On-balance volume (OBV) signals combine price direction and volume. First define

$$D_t = \begin{cases} 1 & \text{if } P_t^{\text{futures}} > P_{t-1}^{\text{futures}}, \\ -1 & \text{if } P_t^{\text{futures}} < P_{t-1}^{\text{futures}}, \\ 0 & \text{otherwise.} \end{cases}$$

Cumulative OBV is $OBV_t = \sum_{k=1}^{t} D_k \cdot \text{Vol}_k$. Moving averages of OBV (short windows $1, 2, 3$; long windows $9, 12$) are then used to generate signals:

$$VOL_{s,\ell,t} = \mathbf{1}\Big(MA_{s,t}^{\text{OBV}} \geq MA_{\ell,t}^{\text{OBV}}\Big).$$

3.1.5 Portfolio exercise

To evaluate the economic value of the return forecasts, we consider a mean-variance investor who allocates her portfolio between crude oil futures and risk-free Treasury bills at the end of each month t. The optimal portfolio weight in oil futures for month $t+1$ is given by

$$w_t = \frac{1}{\xi} \frac{\hat{r}_{t+1}}{\hat{\sigma}_{t+1}^2},$$

where $\xi = 3$ is the coefficient of relative risk aversion, $\hat{r}_{t+1}$ denotes the forecast of the one-month-ahead spot oil return, and $\hat{\sigma}_{t+1}^2$ is the corresponding variance forecast estimated using an expanding window of the past 60 months. The portfolio weight is constrained to lie in the interval $[-1.5, 1.5]$ to mitigate extreme leverage.

Let r_{t+1}^F denote the simple return on crude oil futures and let r_{t+1}^f denote the monthly risk-free rate, which is obtained from the annualised 3-month Treasury bill rate (TB3MS) according to

$$r_{t+1}^f = \left(1 + \frac{\text{TB3MS}_{t+1}}{100}\right)^{1/12} - 1.$$

The excess return on crude oil futures is then defined as

$$r_{t+1}^{\text{excess}} = r_{t+1}^F - r_{t+1}^f.$$

Using the forecasted spot return as the trading signal, the resulting portfolio return is

$$R_{p,t+1} = w_t r_{t+1}^{\text{excess}} + r_{t+1}^f,$$

and the corresponding portfolio excess return is

$$R_{p,t+1}^{\text{excess}} = R_{p,t+1} - r_{t+1}^{f} = w_t r_{t+1}^{\text{excess}}.$$

We evaluate the performance of the forecasting models using both the annualised certainty equivalent return (CER) gain and the annualised Sharpe ratio of the portfolio excess returns. The CER gain is defined as

$$\text{CER gain} = 1200 \left(\overline{R}_p^{\text{model}} - \frac{\xi}{2}\sigma_{p,\text{model}}^2 - \left(\overline{R}_p^{\text{bench}} - \frac{\xi}{2}\sigma_{p,\text{bench}}^2 \right) \right),$$

while the annualised Sharpe ratio is given by

$$SR = \sqrt{12}\, \frac{\overline{R_p^{\text{excess}}}}{\sigma\left(R_p^{\text{excess}}\right)},$$

where $\overline{R_p^{\text{excess}}}$ and $\sigma(R_p^{\text{excess}})$ denote the sample mean and standard deviation of the portfolio excess returns, respectively. A positive CER gain indicates that the investor would prefer the forecasts from the model of interest over the benchmark.

3.2 LASSO

3.2.1 Setup

We apply the LASSO estimator defined in Section 2.2 recursively with an expanding window. At each forecast origin t:

1. The available training sample is split into an inner training set (first 60%) and a validation set (last 40%).
2. A LASSO path is estimated on the inner training set.
3. The λ that minimises the validation MSPE is selected.
4. The LASSO is re-estimated on the full training sample using the chosen λ and a one-month-ahead forecast $\hat{y}_{t+1}$ is produced.

3.2.2 Alternative LASSO λ selection: CV and BIC

The LASSO estimator is again defined as in Section 2.2. Instead of using a fixed 60/40 validation split, the shrinkage parameter λ can be selected by cross-validation, the BIC. All methods are applied recursively within an expanding window.

CV:

1. The available training sample (up to month $t-1$) is randomly partitioned into $K = 10$ folds.
2. For a grid of λ values, a LASSO model is trained on $K - 1$ folds and evaluated on the remaining fold.
3. The cross-validated MSPE is computed for each λ, and the value minimizing this criterion is selected.
4. The LASSO is re-estimated on the full training sample using the chosen λ, and a one-month-ahead forecast $\hat{y}_t$ is produced.

BIC:

1. A LASSO path is estimated on the full training sample for a grid of λ values.
2. For each value of λ, the residual sum of squares RSS and the number of non-zero coefficients df are computed.
3. The BIC is computed as $BIC = n \log(RSS/n) + df \log(n)$, where n is the training sample size.
4. The λ that minimizes the BIC is selected, and the corresponding LASSO model is used to produce the forecast $\hat{y}_t$.

Findings In Table 3.1, it is observed that LASSO (BIC) delivers the best out-of-sample forecasting performance among the four LASSO variants. Both Clark-West and Pesaran-Timmermann tests are strongly significant, which means that the LASSO (BIC) model not only produces a statistically significant reduction in MSPE relative to the benchmark, but also correctly predicts the direction of oil returns more often than a random walk.

Table 3.1: LASSO out-of-sample forecasting and portfolio performance

Model	MSPE	OOS R^2 (%)	SR (%)	CER	Sharpe
Benchmark	11.61	0.00	–	0.00	-0.1757
LASSO (60/40)	10.05	13.42	68.72	-9.19	0.3151
LASSO (CV)	10.02	13.73	68.72	-5.05	0.3777
LASSO (BIC)	9.93	14.47	68.72	-5.65	0.3190

Selection method	CW p	PT p
LASSO (60/40)	0.010	< 0.0001
LASSO (CV)	< 0.0001	< 0.0001
LASSO (BIC)	0.0003	< 0.0001

In regards to the portfolio exercise, LASSO (BIC) yields a negative annualised CER gain (–5.65%) and a Sharpe ratio of 0.319, compared to the benchmark's Sharpe ratio of -0.176. Although the CER gain is negative, the positive Sharpe ratio indicates that the model achieves superior risk-adjusted returns relative to the benchmark.

Table 3.2 shows the predictor selection frequencies of the four LASSO variants. MOM_1 is selected 100% of the time, confirming that the ℓ_1 penalty consistently retains the strongest predictor irrespective of the tuning method. The sparsity level varies considerably across methods. BIC selects only six predictors, corresponding to a relatively large effective λ, while the 60/40 split and CV select intermediate model sizes.

Table 3.2: Variable selection frequencies (%) – LASSO variants

Predictor	**60/40**	**CV**	**BIC**
MOM_1	100.0	100.0	100.0
MOM_2	95.3	99.5	92.9
MOM_3	25.1	16.6	—
imports_growth	61.1	76.8	12.3
stocks_growth	52.6	79.1	13.7
m2_growth	29.9	29.9	29.9
cfnai	30.3	30.3	29.9
MA_1_12	21.3	10.4	—
MA_1_9	2.8	—	—
MA_3_9	15.6	—	—
MA_3_12	—	0.5	—
VOL_1_9	7.1	0.9	—
VOL_1_12	3.8	0.5	—
VOL_2_12	0.9	—	—
VOL_3_9	6.6	—	—
VOL_3_12	9.0	—	—
TB3MS	10.0	—	—
GS10	3.3	—	—
epu	1.9	1.4	—
AAA	—	1.4	—
ip_growth	7.6	0.5	—
unemp_diff	6.6	—	—
BAA	0.9	—	—
MOM_9	1.4	—	—

The cumulative squared prediction error (CSPE) difference between LASSO (BIC) and the benchmark is plotted in Figure 3.1. The steady upward drift confirms that the model consistently outperforms the benchmark over the entire out-of-sample period. The cumulative

wealth dynamics are illustrated in Figure 3.2.

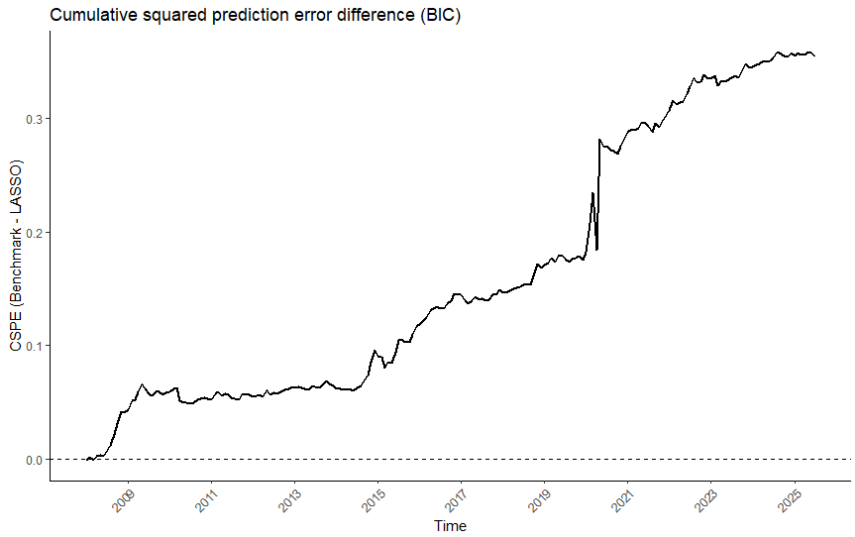


Figure 3.1: CSPE difference: LASSO (BIC) vs benchmark

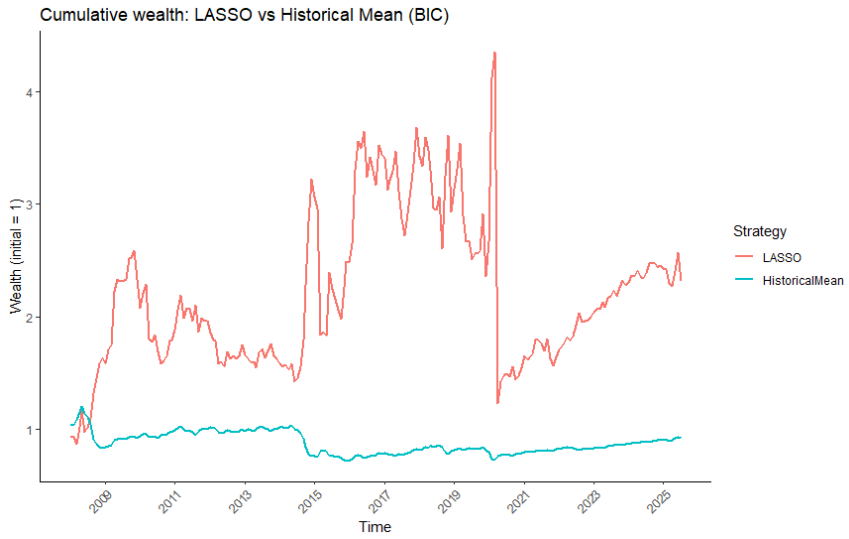


Figure 3.2: Cumulative wealth: LASSO (BIC) vs benchmark

The selection dynamics of LASSO (BIC) are visualised in the heatmap (Figure 3.3). The few predictors that enter most often are short-horizon momentum indicators (MOM_1, MOM_2) and occasionally supply variables (imports growth, stocks growth). This sparse pattern is consistent with the theoretical expectation that an effective LASSO selection should retain only the most informative signals.

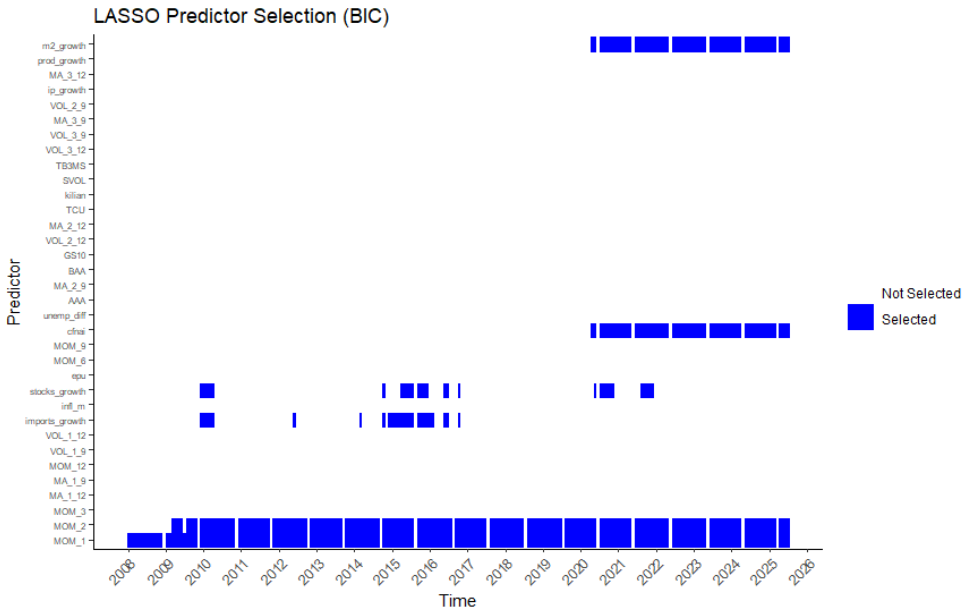


Figure 3.3: Variable selection heatmap – LASSO (BIC)

Every plot including the evolution of λ can be found in Appendix B.1.

3.3 Elastic Net

3.3.1 Setup

We apply the elastic net estimator defined in Section 2.3 recursively with an expanding window. At each forecast origin t:

1. The available training sample is split into an inner training set (first 60%) and a validation set (last 40%).
2. For each candidate mixing parameter α in a grid $\{0.1, 0.3, 0.5, 0.7, 0.9\}$, an Elastic Net path is estimated on the inner training set.
3. For each combination (α, λ), the MSPE is computed on the validation set.
4. The pair (α, λ) that minimises the validation MSPE is selected.
5. The Elastic Net is re-estimated on the full training sample using the chosen (α, λ) and a one-month-ahead forecast $\hat{y}_{t+1}$ is produced.

3.3.2 Alternative Elastic Net λ selection: CV and BIC

The Elastic Net can also be estimated using CV or the BIC to select λ, while still tuning α via a grid search. The procedures for CV and BIC are identical to those described for the

LASSO in the section earlier, except that for each α in the grid, the optimal λ is chosen by CV or BIC and the final (α, λ) pair is the one with the lowest cross-validated MSPE or BIC, respectively.

Findings The Elastic Net results in Table 3.3 lead to two main conclusions. First, all three λ selection methods significantly outperform the benchmark in terms of both point forecast accuracy and directional accuracy. Second, CV delivers the best overall performance among the three methods, achieving the highest out-of-sample R^2. Overall, the Elastic Net performs similarly to the LASSO, and the choice of λ selection matters more than the addition of the ℓ_2 penalty.

Table 3.3: Elastic Net out-of-sample forecasting and portfolio performance

Model	MSPE	OOS R^2 (%)	SR (%)	CER	Sharpe
Benchmark	11.61	0.00	–	0.00	-0.1757
EN (60/40)	10.08	13.18	68.72	-10.01	0.3037
EN (CV)	9.94	14.44	68.25	-4.77	0.3777
EN (BIC)	9.95	14.30	68.72	-5.65	0.3257

Selection method	CW p	PT p
EN (60/40)	0.011	< 0.0001
EN (CV)	< 0.0001	< 0.0001
EN (BIC)	0.0003	< 0.0001

The Elastic Net selection frequencies in Table 3.4 are almost identical to those of the LASSO (Table 3.2). Short-horizon momentum variables are selected in nearly all months across all λ selection methods. The BIC criterion produces the sparsest model, retaining only a few core predictors, while CV and the 60/40 split keep a wider set that includes technical indicators and macroeconomic variables. The close correspondence between the two methods confirms that the Elastic Net's ℓ_2 penalty does not alter the set of relevant predictors; it only affects the shrinkage of coefficients.

Table 3.4: Elastic Net variable selection frequencies (%)

Predictor	**60/40 split**	**CV**	**BIC**
MOM_1	100.0	100.0	100.0
MOM_2	95.7	100.0	95.3
imports_growth	61.6	88.6	43.6
stocks_growth	52.6	92.9	50.2
cfnai	31.3	31.8	31.3
m2_growth	29.9	29.9	29.9
MOM_3	25.1	23.2	17.1
MA_1_12	22.3	22.3	17.1
MA_3_9	15.6	1.4	–
TB3MS	11.8	1.4	–
ip_growth	9.0	14.2	17.1
VOL_3_12	9.0	1.9	–
VOL_3_9	7.6	1.9	–
unemp_diff	7.1	1.9	–
VOL_1_9	7.1	0.9	–
VOL_1_12	3.8	–	–
GS10	3.3	–	–
MA_1_9	2.8	1.9	–
BAA	1.9	–	–
epu	1.9	5.7	0.5
MOM_9	1.4	–	–
VOL_2_12	0.9	–	–
AAA	0.9	1.9	–
MA_3_12	–	0.5	–
prod_growth	–	0.5	–

The CV specification of the Elastic Net delivers the best out-of-sample performance. Its CSPE plot, heatmap, and cumulative wealth plot are presented in Figures 3.4–3.6.

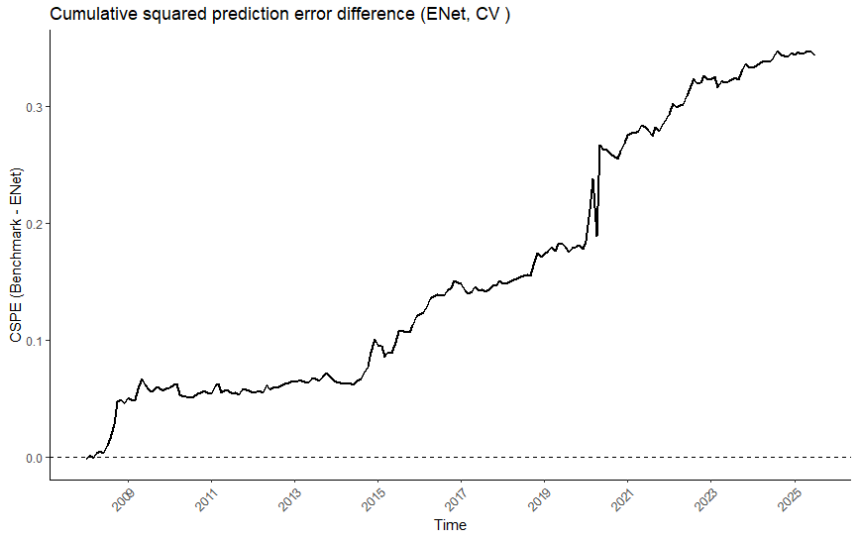


Figure 3.4: CSPE difference: Elastic Net (CV) vs benchmark

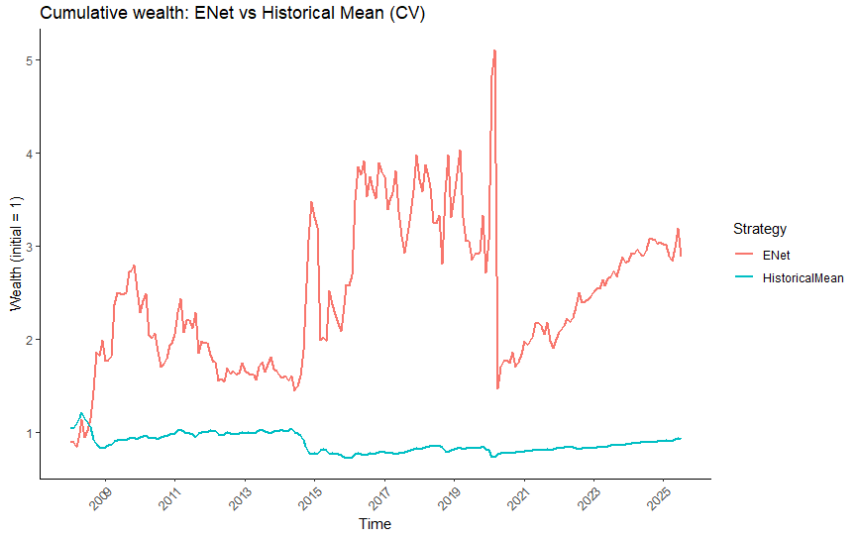


Figure 3.5: Cumulative wealth: Elastic Net (CV) vs benchmark

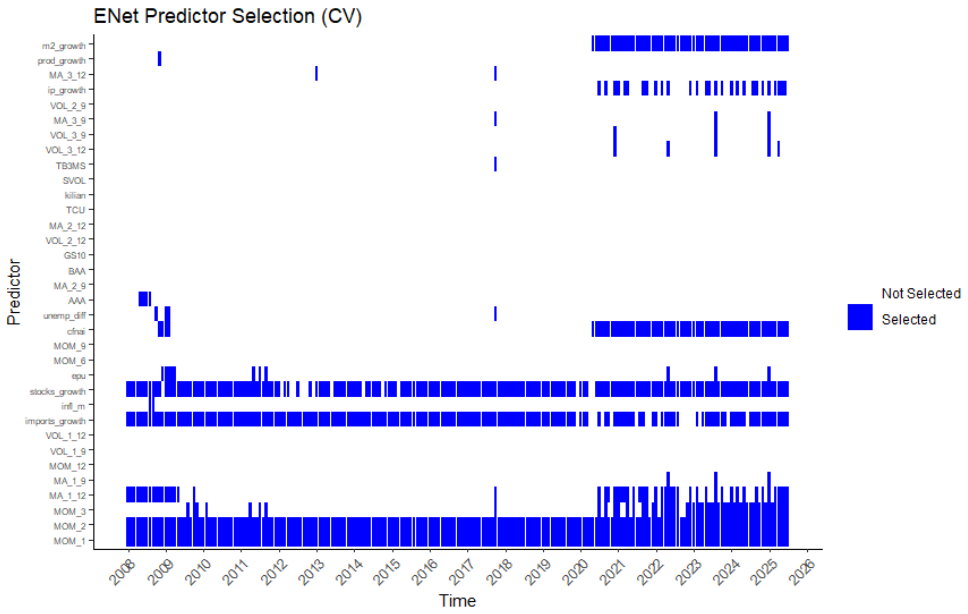


Figure 3.6: Variable selection heatmap – Elastic Net (CV)

Every plot including the evolution of λ and α for each variant can be found in Appendix B.2.

3.4 OCMT and BMT

3.4.1 Setup

We apply the OCMT and BMT procedures described in Sections 2.4 and 2.5 to forecast the one-month-ahead oil return r_{t+1}, using the predictor vector

$$x_t = (x_{1t}, \ldots, x_{pt})'.$$

OCMT OCMT selects all predictors whose t-statistics exceed the stage-dependent critical value.

1. **Stage 1:** For each predictor x_{it}, compute the marginal t-statistic t_i from the regression of r_{t+1} on an intercept and x_{it}. Define

$$S_1 = \left\{ i \in \{1, \ldots, p\} : |t_i| > c_p^{(1)}(q, \delta_1) \right\}.$$

2. **Stage $k \geq 2$:** Conditional on the previously selected set S_{k-1}, compute the conditional t-statistics

$$t_{j|S_{k-1}}, \qquad j \notin S_{k-1},$$

from regressions of r_{t+1} on an intercept, the predictors in S_{k-1}, and predictor x_{jt}. Update

$$S_k = S_{k-1} \cup \left\{ j \notin S_{k-1} : |t_{j|S_{k-1}}| > c_p^{(k)}(q, \delta_2) \right\}.$$

3. Stop when

$$S_k = S_{k-1}.$$

The final selected set is denoted by

$$\hat{S} = S_k,$$

and the post-selection forecasting model is estimated by OLS using the predictors in $\hat{S}$.

BMT BMT follows the same sequential testing procedure but selects at most one predictor at each stage.

1. **Stage 1:** Compute the marginal t-statistics t_i for all predictors and define

$$i^* = \arg\max_{1 \leq i \leq p} |t_i|.$$

If

$$|t_{i^*}| > c_p^{(1)}(q, \delta_1),$$

set

$$S_1 = \{i^*\};$$

otherwise let

$$S_1 = \emptyset.$$

2. **Stage** $k \geq 2$**:** For each remaining predictor $j \notin S_{k-1}$, compute the conditional t-statistics $t_{j|S_{k-1}}$ and define

$$j^* = \arg\max_{j \notin S_{k-1}} |t_{j|S_{k-1}}|.$$

If

$$|t_{j^*|S_{k-1}}| > c_p^{(k)}(q, \delta_2),$$

update

$$S_k = S_{k-1} \cup \{j^*\};$$

otherwise terminate the procedure.

3. The final selected set is denoted by

$$\hat{S} = S_k,$$

and the post-selection forecasting model is estimated by OLS using the predictors in $\hat{S}$.

The significance level parameter is set to $q = 0.05$ for OCMT and $q = 0.01$ for BMT.

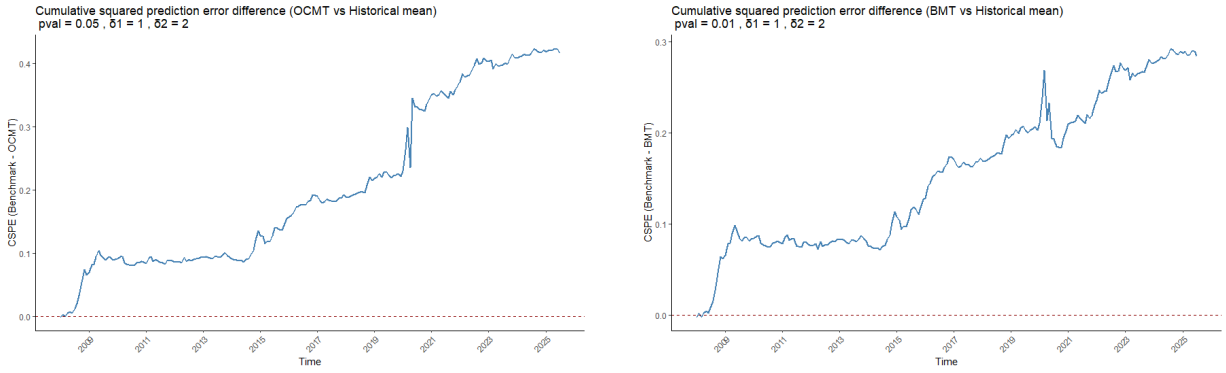
Table 3.5: OCMT and BMT out-of-sample forecasting, economic performance, and model size

Model	MSPE	OOS R^2 (%)	SR (%)	CER	Sharpe
Benchmark	11.61	0.00	–	0.00	-0.176
OCMT	9.64	16.95	48.57	-10.31	0.417
BMT	10.26	11.62	68.72	-8.90	0.378

Model	CW p-value	PT p-value
OCMT	0.000231	0.669
BMT	2.10×10^{-9}	< 0.0001

Model size	Mean	Median (Min/Max)
OCMT	4.01	3 (2 / 7)
BMT	1.29	1 (1 / 2)

Table 3.5 compares the two greedy methods. Both models show improvement in point forecast accuracy compared to the benchmark, achieving positive out-of-sample R^2's. Both methods exhibit directional accuracy, whereas BMT shows a substantially higher success ratio. BMT is also far sparser, selecting on average only one or two predictors per month, while OCMT selects on average four predictors per month.



(a) OCMT (b) BMT

Figure 3.7: CSPE difference

Figure 3.7 shows the CSPE differences. Both boosting methods curves exhibit a steady upward drift throughout the out-of-sample period, closely resembling the patterns observed for LASSO and Elastic Net.

Since BMT achieves the best overall performance, its cumulative wealth dynamics are illustrated in Figure 3.8. Despite a negative CER gain, the Sharpe ratio is positive and substantially higher than the benchmark.

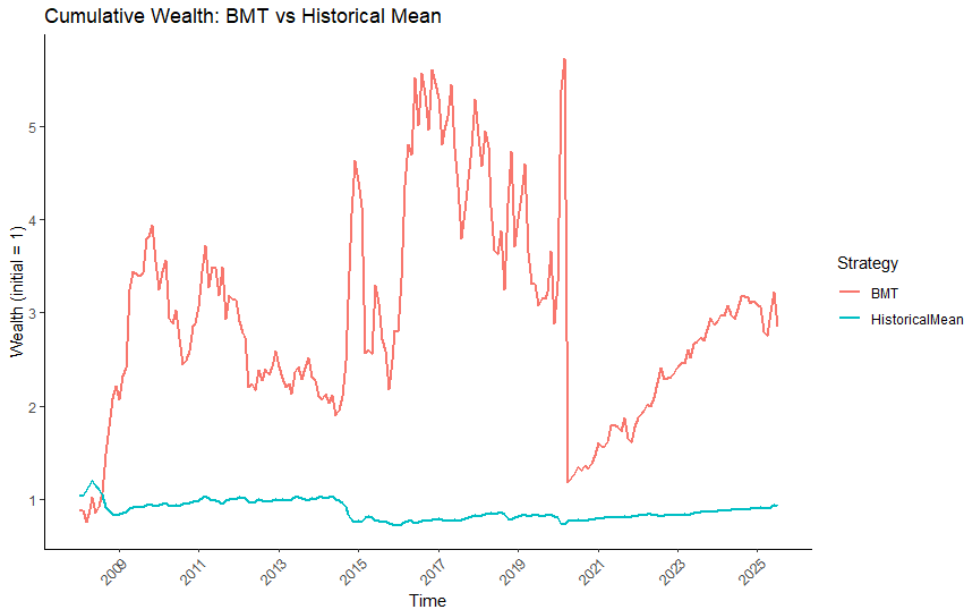


Figure 3.8: Cumulative wealth: BMT

The heatmaps of selected predictors at time t is illustrated below.

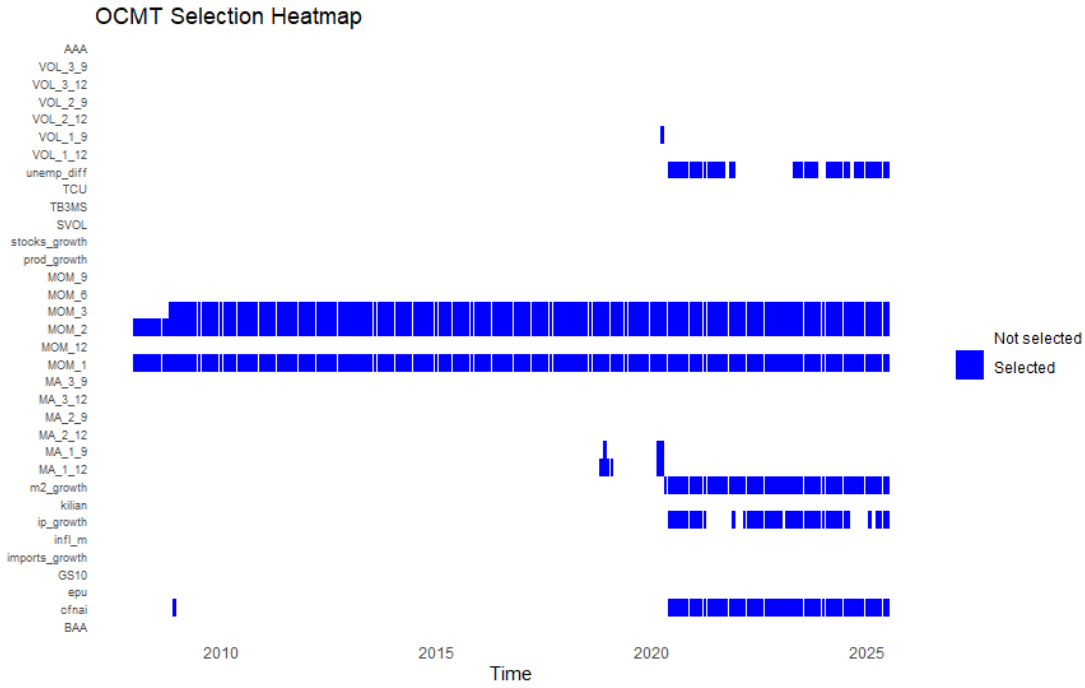


Figure 3.9: OCMT variable selection heatmap

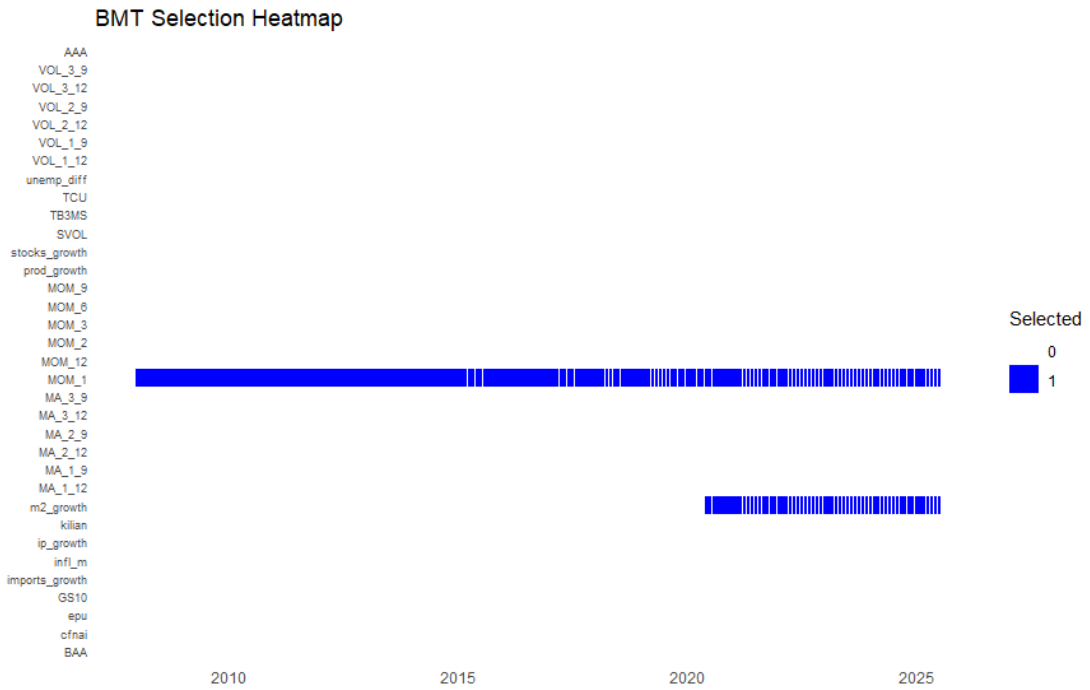


Figure 3.10: BMT variable selection heatmap

Table 3.6: Variable selection frequencies (%)

Predictor	**OCMT**	**BMT**
MOM_1	100.0	100.0
MOM_2	100.0	–
MOM_3	95.3	–
m2_growth	29.9	29.4
cfnai	29.9	–
ip_growth	21.8	–
unemp_diff	19.9	–
MA_1_12	2.8	–
MA_1_9	1.4	–
VOL_1_9	0.5	–

OCMT selects more macroeconomic predictors, while BMT is extremely sparse. Compared to LASSO and Elastic Net, OCMT favours macroeconomic uncertainty over technical indicators, and BMT is even sparser than LASSO (BIC).

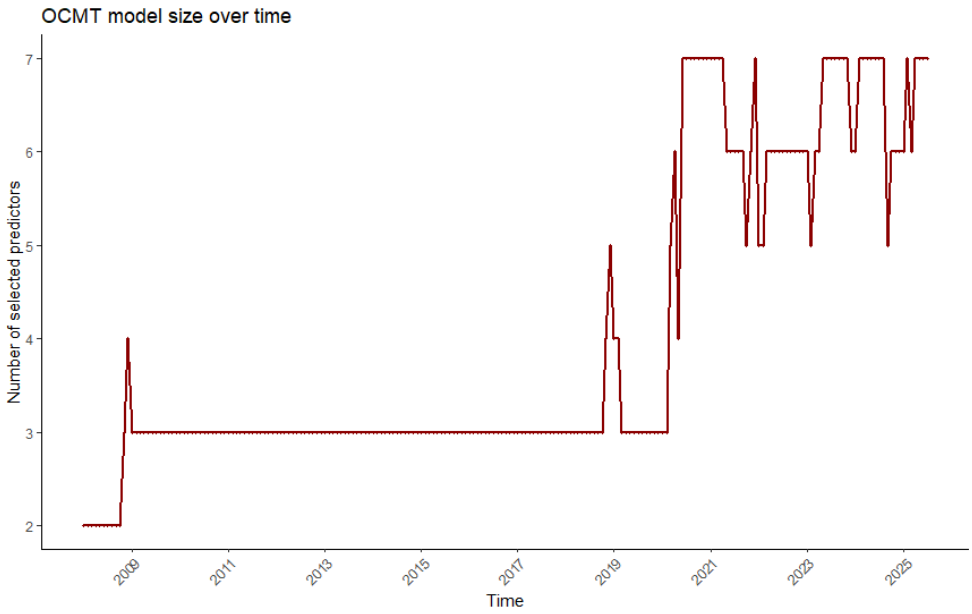


Figure 3.11: OCMT model size over time

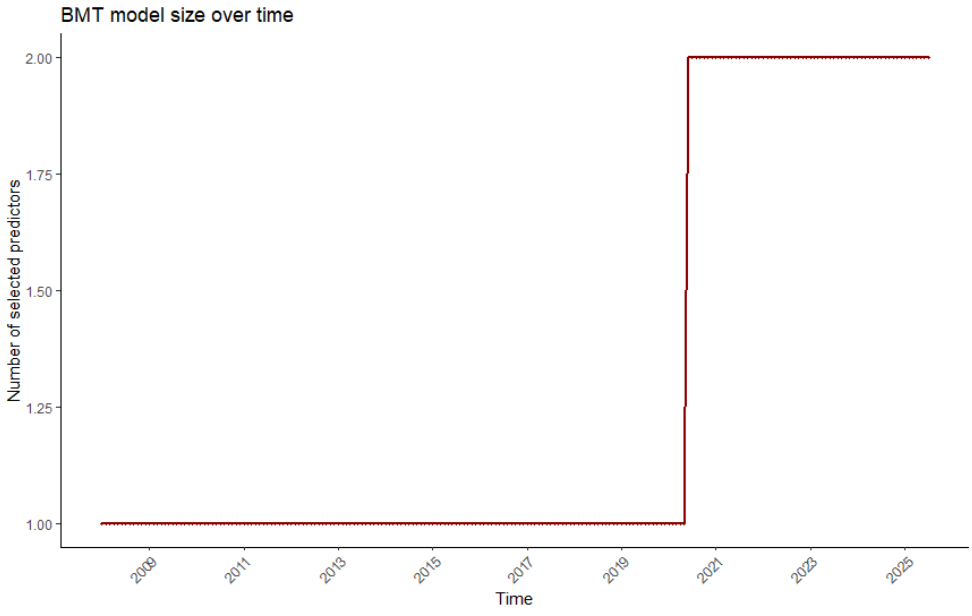


Figure 3.12: BMT model size over time

Figures 3.11 and 3.12 show the evolution of the number of selected predictors. OCMT's model size fluctuates a little, while BMT remains stable.

In summary, OCMT and BMT differ fundamentally in their selection rules. BMT outperforms OCMT in point forecast accuracy and obtains almost same CSPE as LASSO and

Elastic Net. OCMT's simultaneous selection rule appears too aggressive, while BMT's stepwise approach successfully isolates the dominant momentum signal and rarely adds other predictors.

Every plot for OCMT and BMT can be found in Appendix B.3 and B.4, respectively.

3.5 Realised volatility of crude oil

3.5.1 Presenting the predictors

Following [4], we compute monthly realised volatility from daily WTI spot prices as $RV_t = \sum_{i=1}^{N_t} r_{t,i}^2$, where $r_{t,i}$ is the daily log return and N_t is the number of daily observations in month t. The dependent variable is the log realised volatility: $LV_t = \ln(RV_t)$.

We collect the following 14 uncertainty variables:

1. EPU (economic policy uncertainty)
2. GPR (geopolitical risk index)
3. EMV (equity market volatility tracker)
4. MPU (monetary policy uncertainty)
5. SVAR (S&P 500 realised volatility)
6. RSK_SP (S&P 500 realised skewness)
7. RSK_Oil (crude oil realised skewness)
8. RA (risk aversion index)
9. EU (economic uncertainty)
10. RU (real uncertainty)
11. MU (macro uncertainty)
12. FU (financial uncertainty)
13. VIX (CBOE volatility index)
14. OVX (CBOE crude oil volatility index)

For exploratory analysis, we also gather several other uncertainty indices from different articles, that have explored uncertainty indeces before:

15. TPU (trade policy uncertainty; [13])
16. CPU (climate policy uncertainty; [14])
17. WUI (world uncertainty index; [15])
18. IDEMV (infectious disease equity market volatility; [16])
19. GEPU (global economic policy uncertainty; [17])

All series are taken at monthly frequency.

To ensure stationarity, the paper applies the ADF test to each uncertainty predictor. Table 3.7 reports the test results and the chosen transformation. Variables that do not reject the unit root hypothesis in the ADF test are treated as non-stationary and are therefore taken in first differences, since the first difference is found to be stationary for all variables. Variables that reject the unit root hypothesis are used in levels. VIX and OVX are log-transformed, following [18, 19]. All predictors are lagged one month. Because OVX starts in June 2007, the sample runs from January 2008 to February 2024.

Table 3.7: Unit root tests and final transformations

Variable	**ADF decision**	**Transformation**
EPU, MPU, GEPU, TPU, CPU, RU, MU, FU, WUI, IDEMV	Non-stationary	First difference
GPR, RSK_Oil, RA, EU, EMV, SVAR, RSK_SP	Stationary	Level
VIX, OVX	Non-stationary	Log level

Based on [20], technical indicators derived from futures prices significantly improve crude oil volatility forecasts out-of-sample. We therefore augment the uncertainty set with the 18 technical indicators described in Section 3.1.4.

3.5.2 Value-at-Risk and straddle strategy evaluation

To assess the economic value of the volatility forecasts, we implement two complementary exercises: a Value-at-Risk (VaR) calculation and a simple straddle strategy based on predicted volatility.

Value-at-Risk savings

From the log realised volatility forecast $\widehat{LV}_{t+1}$ produced by each model, we obtain the predicted monthly standard deviation of oil futures returns as $\hat{\sigma}_{t+1} = \exp(\widehat{LV}_{t+1}/2)$. We compute VaR using Filtered Historical Simulation (FHS) following [21]. For each forecast month, we take the past 60 months of actual futures returns, standardise them to zero mean and unit variance using the rolling sample mean and standard deviation: $\tilde{r}_i = (r_i - \mu_{60})/\sigma_{60}$ for $i = t-59, \ldots, t$. The empirical 5th percentile of these standardised returns is denoted $q^{emp}_{0.05}$. The 95% VaR is then

$$\mathrm{VaR}^{0.05}_{t+1} = -\hat{\sigma}_{t+1} \cdot q^{emp}_{0.05},$$

This approach directly incorporates the observed fat tails and asymmetry of oil returns without imposing a parametric distribution. The VaR saving of a model relative to the

historical mean benchmark is defined as VaR saving = $\text{VaR}_{\text{bench}} - \text{VaR}_{\text{model}}$. A positive saving indicates that the model predicts a smaller VaR, implying lower required capital reserves. We also report the actual violation rate (the fraction of months in which the realised return falls below the predicted VaR) to check calibration. A well-calibrated VaR model should have a violation rate close to the nominal 5% level. Rates significantly above 5% indicate underestimation of tail risk, while rates below 5% suggest overestimation.

Straddle strategy

A straddle consists of buying both a call and a put option with the same strike and maturity, profiting from large absolute price movements. Since the models only forecast volatility, we implement a stylised straddle strategy based on the predicted monthly standard deviation. At time t, a trading signal is formed using the one-step-ahead volatility forecast $\hat{\sigma}_{t+1}$:

$$\text{Signal}_t = \begin{cases} +1 & \text{if } \hat{\sigma}_{t+1} > \text{median}(\hat{\sigma}_{\text{history}}), \\ -1 & \text{otherwise,} \end{cases}$$

where $\text{median}(\hat{\sigma}_{\text{history}})$ denotes the median realised monthly volatility over the in-sample period. This threshold is fixed before the out-of-sample evaluation begins, thereby avoiding look-ahead bias. A value of $+1$ corresponds to a long straddle position, while -1 corresponds to a short straddle position.

The payoff in month $t+1$ is defined as

$$\text{Payoff}_{t+1} = \text{Signal}_t \cdot \big(|r_{t+1}| - \bar{p}\big),$$

where r_{t+1} is the monthly oil futures return and $\bar{p}$ is the average absolute return over the in-sample period, which serves as a proxy for the option premium and therefore the capital at risk. The corresponding monthly return of the strategy is

$$R_{t+1} = \frac{\text{Payoff}_{t+1}}{\bar{p}}.$$

The excess return relative to the monthly risk-free rate is computed as

$$R^e_{t+1} = R_{t+1} - r_{f,t+1},$$

where $r_{f,t+1}$ denotes the monthly risk-free rate derived from the annualised 3-month Treasury bill rate (TB3MS), defined analogously to the portfolio exercise in Section 3.1.5:

$$r_{f,t+1} = \left(1 + \frac{\text{TB3MS}_{t+1}}{100}\right)^{1/12} - 1.$$

The annualised Sharpe ratio is then given by

$$\text{Sharpe} = \frac{\overline{R^e}}{\sigma(R^e)}\sqrt{12},$$

where $\overline{R^e}$ and $\sigma(R^e)$ denote the sample mean and standard deviation of the excess returns over the out-of-sample period. Transaction costs are assumed to be zero. This trading framework is inspired by [22], who evaluate volatility forecasts using a straddle strategy based on predicted volatility movements.

3.6 LASSO and Elastic Net

Following the same recursive forecasting procedure as described in Section 3.2 and Section 3.3 for oil returns, the paper applies the LASSO and Elastic Net variants to the realised volatility dataset.

Table 3.8: LASSO and Elastic Net out-of-sample forecasting performance

Model	MSPE	OOS R^2 (%)	SR (%)	CW p	PT p
Benchmark	0.9737	0.00	–	–	–
LASSO (60/40)	0.5535	43.16	56.30	0.0011	0.0878
LASSO (CV)	0.5253	46.05	55.46	0.0002	0.1398
LASSO (BIC)	0.5107	47.55	60.50	0.0001	0.0138
EN (60/40)	0.5034	48.30	56.30	0.0012	0.0973
EN (CV)	0.5198	46.62	54.62	0.0008	0.1632
EN (BIC)	0.5143	47.18	57.98	0.0002	0.0532

All LASSO and Elastic Net variants substantially reduce MSPE relative to the benchmark, with OOS R^2 values ranging from 43% to 48%. The Clark–West test is highly significant ($p < 0.01$) for every variant, confirming that the improvements in point forecast accuracy are not due to chance. Among the LASSO specifications, LASSO (BIC) delivers the best performance: it achieves the lowest MSPE, the highest OOS R^2 and the highest success ratio. Moreover, it is the only LASSO variant with a significant PT test at the 5% level ($p = 0.0138$), indicating reliable directional predictions of volatility changes.

The Elastic Net variants perform similarly or slightly better. Elastic Net (60/40 split) attains the highest OOS R^2 overall and the lowest MSPE, marginally outperforming LASSO (BIC) in point forecast accuracy. Its success ratio and PT test are not statistically significant, however. Elastic Net (BIC) achieves a competitive OOS R^2 and a success ratio of 57.98%, with a PT test close to significance. Overall, the paper believes Elastic Net (60/40

split) is the best model for point forecasting, while LASSO (BIC) excels in directional accuracy out of the shrinkage models.

Table 3.9: Variable selection frequencies (%)

Predictor	LASSO (BIC)	Elastic Net (60/40)
log_OVX	100.0	100.0
MOM_2	92.5	94.2
d_FU	82.5	92.5
RA	50.0	65.0
d_MU	44.2	79.2
EU	35.0	82.5
d_IDEMV	31.7	45.0
MOM_12	24.2	65.8
RSK_SP	24.2	50.0
GPR	20.0	68.3
MOM_3	10.0	60.8
d_RU	5.8	48.3
SVAR	3.3	40.8
MOM_6	0.0	45.8
VOL_3_9	0.0	45.8
MA_2_12	0.0	45.8
d_CPU	0.0	45.0
d_TPU	0.0	43.3
log_VIX	0.0	42.5
d_EPU	0.0	35.8
RSK_Oil	0.0	33.3
EMV	0.0	28.3
d_MPU	0.0	22.5
VOL_2_12	0.0	22.5
d_GEPU	0.0	15.8
MA_1_9	0.0	15.8
MA_3_9	0.0	15.0
MA_1_12	0.0	11.7
MOM_1	0.0	9.2
MA_3_12	0.0	7.5
d_WUI	0.0	6.7
MOM_9	0.0	5.0
VOL_1_9	0.0	2.5
VOL_1_12	0.0	1.7

The variable selection frequencies for LASSO (BIC) and Elastic Net (60/40 split) are

reported in Table 3.9. log_OVX is selected in every month by both models, confirming its dominant role in volatility forecasting. LASSO (BIC) selects a parsimonious set of predictors, primarily log_OVX, MOM_2, d_FU, RA, and d_MU, consistent with the sparse selection pattern observed in the first dataset. Elastic Net (60/40 split) selects a much larger set, including many technical indicators and differences of uncertainty indices, reflecting the effect of its ℓ_2 penalty that allows correlated variables to enter together. The heatmaps of both models are seen in Figures 3.13 and 3.14.

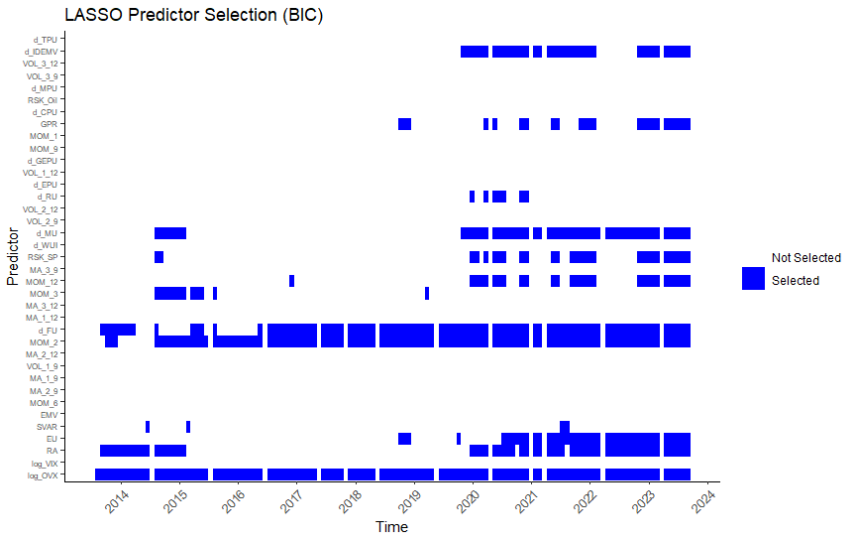
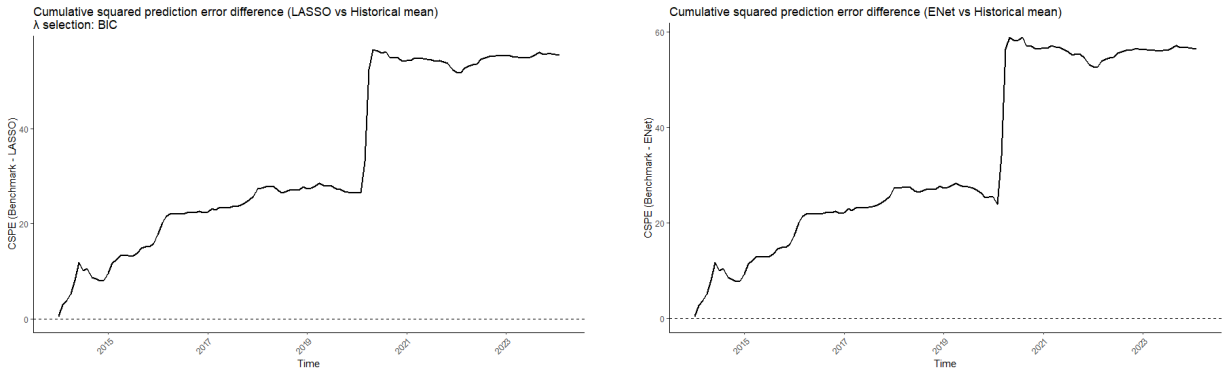


Figure 3.13: LASSO (BIC) heatmap

Figure 3.14: Elastic Net (60/40 split) heatmap

Figure 3.15 shows the CSPE plots of both models, even though LASSO (BIC) chooses less predictors, their CSPE plots look the same.

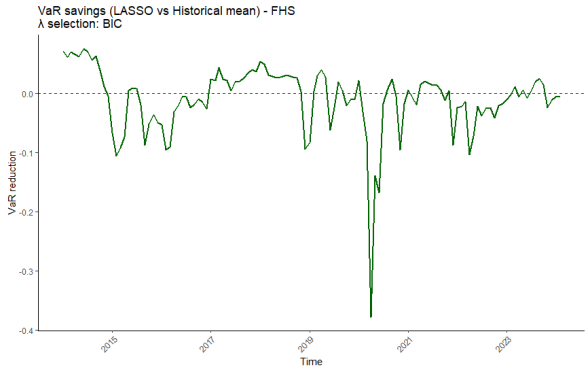


(a) LASSO (BIC).

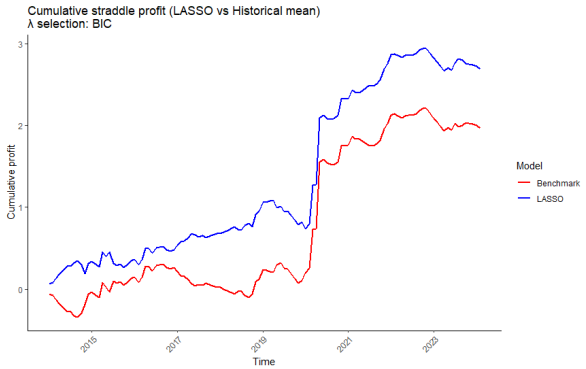
(b) Elastic Net (60/40 split).

Figure 3.15: CSPE plots

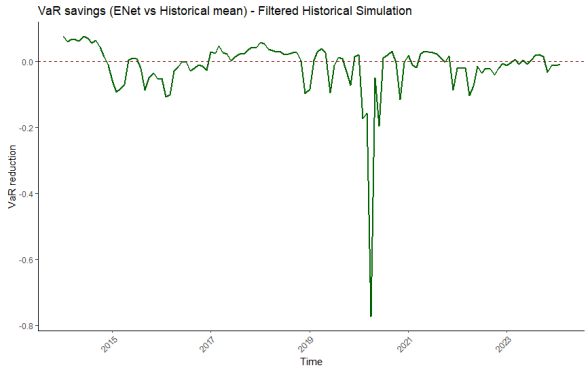
Figure 3.16 presents the economic value of the two models. The left column shows VaR savings computed using FHS. For LASSO (BIC), the average saving is -7.4%, meaning it predicts a higher VaR than the historical mean; its violation rate is 10.8%, slightly above the benchmark's 10.0% and the nominal 5%. Elastic Net (60/40) has a larger average VaR saving of -10.9% and the same violation rate of 10.8%.



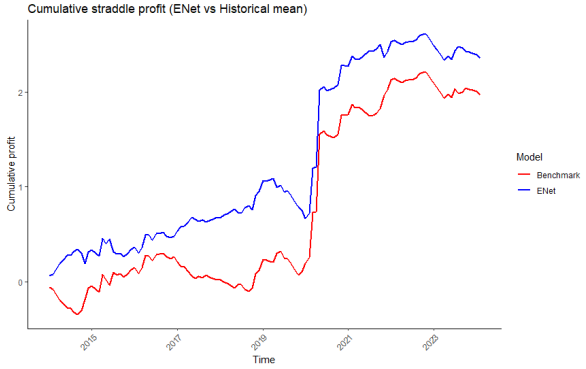
(a) LASSO (BIC): VaR savings.



(b) LASSO (BIC): Cumulative straddle profit.



(c) Elastic Net (60/40): VaR savings.



(d) Elastic Net (60/40): Cumulative straddle profit.

Figure 3.16: VaR being negative means that the strategy is more conservative.

The right column displays the cumulative profit of the straddle strategy. Over the test period, LASSO (BIC) accumulates a total profit of 2.69 and an annualised Sharpe ratio of 0.77, while the benchmark earns 1.97 with a Sharpe of 0.56. Elastic Net (60/40) achieves a total profit of 2.35 and a Sharpe ratio of 0.67, indicating superior economic gains despite its more conservative VaR.

All plots for the LASSO and Elastic Net variants, including heatmaps, CSPE, λ and α evolutions are provided in Appendix C.1 and Appendix C.2, respectively.

3.7 OCMT and BMT

Following the same recursive forecasting procedure as described in Section 3.4 for oil returns, the paper applies the OCMT and BMT to the realised volatility dataset. In our findings, the significance level q was found to be $q^{OCMT} = q^{BMT} = 0.01$.

Table 3.10 reports their out-of-sample forecasting performance.

Table 3.10: OCMT and BMT out-of-sample forecast performance

Model	MSPE	OOS R^2 (%)	SR (%)	CW p	PT p
Benchmark	0.9737	0.00	–	–	–
OCMT	0.735	44.6	51.26	0.000116	0.339
BMT	0.501	48.5	58.82	8.03×10^{-5}	0.0297

Model	Mean model size	Median (Min / Max)
OCMT	13.14	15 (6 / 18)
BMT	1.38	1 (1 / 2)

Although both OCMT and BMT improve point forecast accuracy relative to the benchmark, only BMT exhibits statistically significant directional predictive ability according to the Pesaran-Timmermann test. This suggests that BMT is better at capturing the direction of changes in volatility dynamics, whereas OCMT primarily improves the accuracy of the predicted volatility level.

The variable selection frequencies for OCMT and BMT are summarised in Tables 3.11. OCMT selects a large number of predictors, with an average model size of 13.1, including many uncertainty indices (EMV, SVAR, RA, EU) and the volatility indices log_VIX and log_OVX, all selected in every month. BMT is sparse, selecting on average only 1.4 predictors per month; log_OVX appears in every forecast, while macroeconomic variables such as d_MU and d_FU are occasionally chosen.

Table 3.11: Variable selection frequencies (%)

Predictor	OCMT	BMT
log_OVX	100.0	100.0
EMV	100.0	0.0
SVAR	100.0	0.0
RA	100.0	0.0
EU	100.0	0.0
log_VIX	100.0	0.0
MA_2_12	82.5	0.0
MOM_6	81.7	0.0
MOM_12	79.2	0.0
MA_1_12	77.5	0.0
MOM_2	77.5	0.0
MA_3_12	64.2	0.0
MOM_3	64.2	0.0
MA_2_9	56.7	0.0
MA_1_9	39.2	0.0
d_RU	37.5	0.0
d_MU	37.5	22.5
d_FU	1.7	15.0
MOM_9	14.2	0.0
d_IDEMV	0.8	0.8

Figures 3.17 and 3.18 display the heatmaps for both models. OCMT selects a wide range of predictors throughout the sample, whereas BMT focuses almost exclusively on log_OVX.

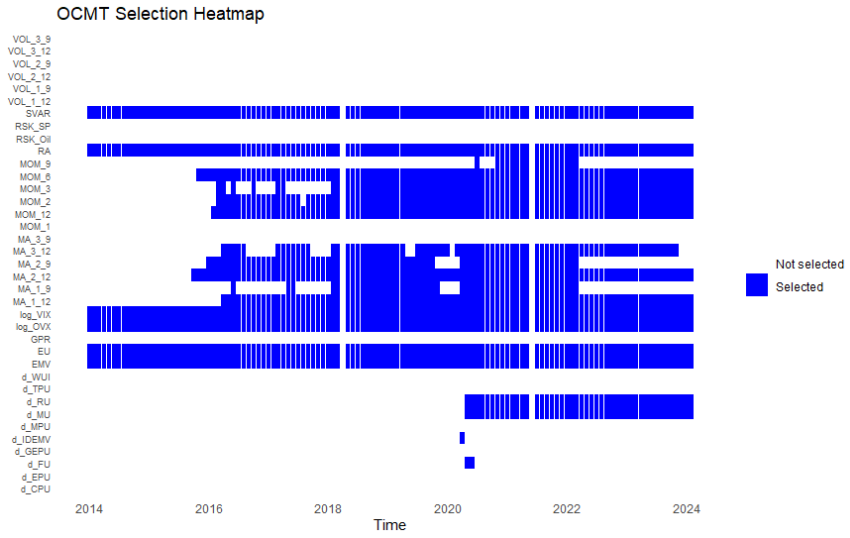


Figure 3.17: OCMT heatmap

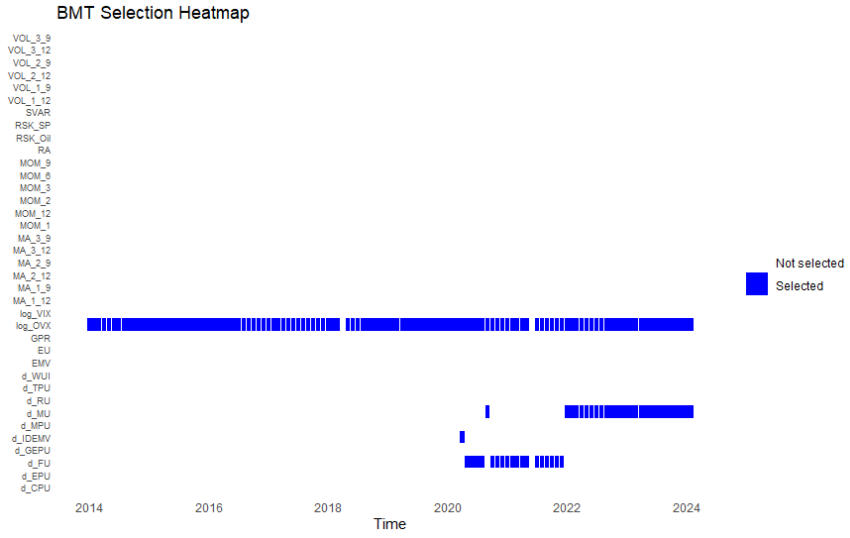
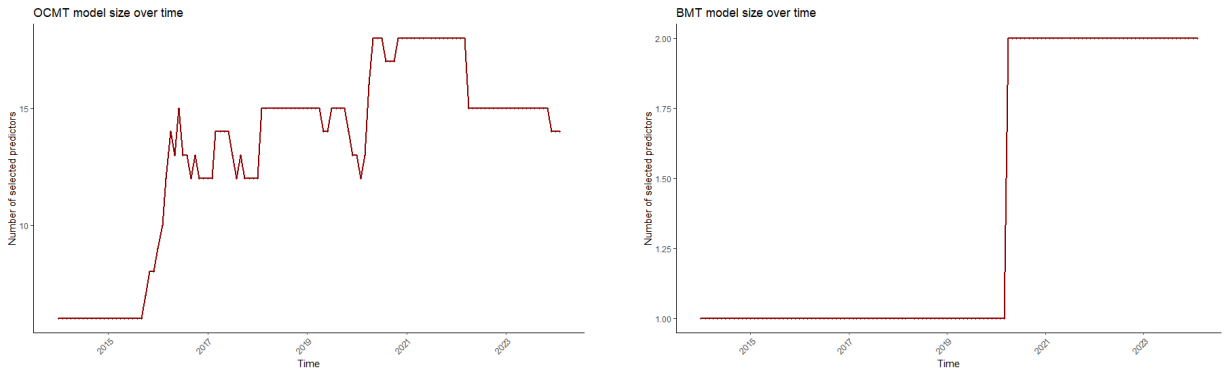


Figure 3.18: BMT heatmap

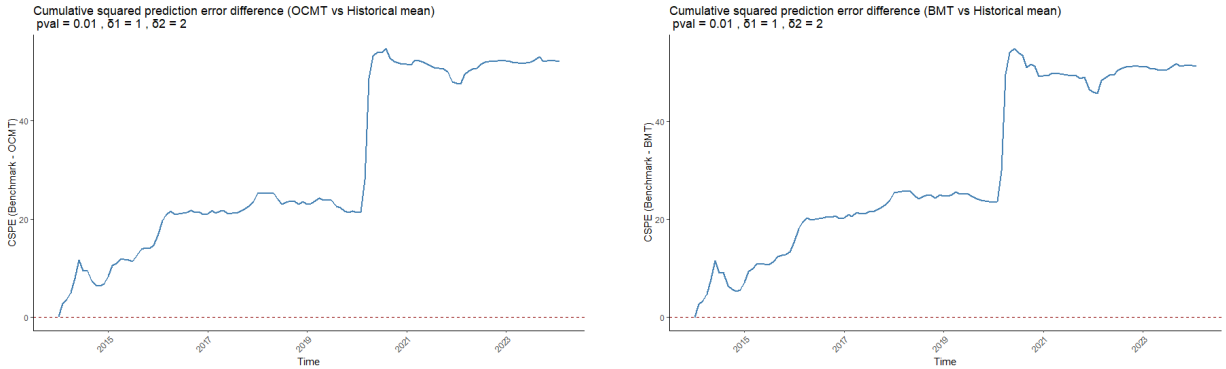
Figure 3.19 display the model size over time. Again, BMT is parsimonous in this case, whereas OCMT allows substantially more predictors to enter simultaneously.



(a) OCMT. (b) BMT.

Figure 3.19: Model size over time plots

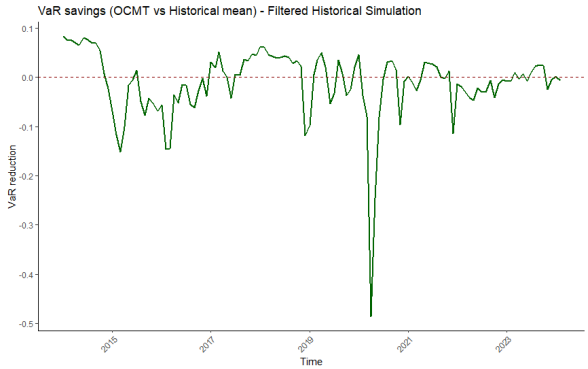
Figure 3.20 shows the CSPE plots. Both models exhibit a generally upward trend, confirming that they consistently outperform the benchmark. The CSPE trajectories are also similar to those obtained for LASSO (BIC) and Elastic Net, indicating consistent forecast improvements across models.



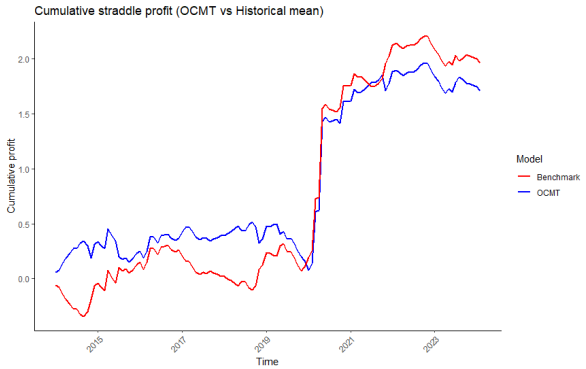
(a) OCMT. (b) BMT.

Figure 3.20: CSPE plots

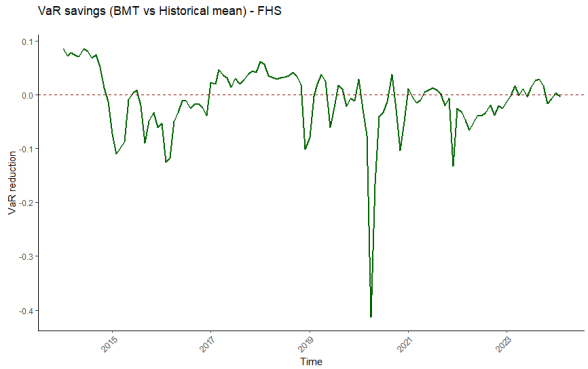
Figure 3.21 presents the economic value through VaR savings and the straddle strategy. For OCMT, the average VaR saving is -9.52%, with a violation rate of 10.8%. The straddle strategy yields a total profit of 1.71 and an annualised Sharpe of 0.48. For BMT, the VaR saving is -7.98% with a violation rate of 10.0%. The straddle profit is 2.16, with a Sharpe ratio of 0.61.



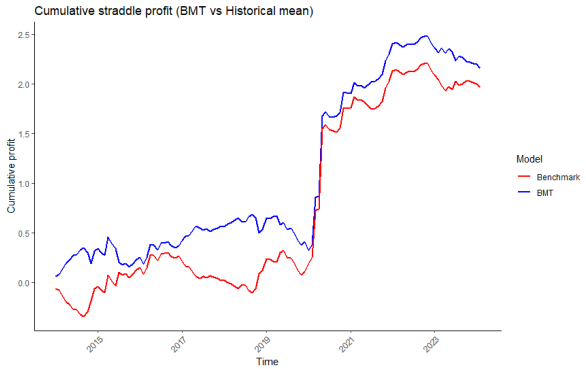
(a) OCMT: VaR savings.



(b) OCMT: Cumulative straddle profit.



(c) BMT: VaR savings.



(d) BMT: Cumulative straddle profit.

Figure 3.21: VaR and cumulative straddle profit plots

All plots for OCMT and BMT are found in Appendix C.3 and Appendix C.4, respectively.

4 Discussion

Table 4.1: Best-performing models across datasets

Dataset	Model	OOS R^2 (%)	Success ratio (%)
Oil returns	OCMT	16.95	48.57
Oil returns	LASSO (BIC)	14.47	68.72
Oil returns	Elastic Net (CV)	14.44	68.25
Oil returns	BMT	11.62	68.72
Realised volatility	BMT	48.50	58.80
Realised volatility	Elastic Net (60/40)	48.30	56.30
Realised volatility	LASSO (BIC)	47.55	60.50

Table 4.1 summarises the main findings across the two forecasting exercises. A central result of the analysis is that no single method dominates across all evaluation criteria. Instead, different models perform best depending on whether the objective is point forecast accuracy or directional prediction.

In the oil return forecasting exercise, OCMT achieved the highest out-of-sample R^2, indicating superior point forecast accuracy. However, its directional forecasting performance was relatively weak compared to the competing methods. By contrast, BMT, LASSO (BIC) and Elastic Net (CV) produced substantially higher success ratios and more significant directional forecasting results.Among these models, LASSO (BIC) appears to provide the strongest overall balance between forecast accuracy and directional performance. Although BMT produces substantially more parsimonious specifications, this aggressive variable reduction may exclude secondary predictors that still contain incremental forecasting information. The findings, therefore, suggest that more restrictive variable selection does not necessarily lead to superior forecasting performance in finite samples, especially when predictors are highly correlated.

A second central finding is the strong predictive power of momentum variables. Across nearly all methods, short-horizon momentum indicators such as MOM_1 and MOM_2 were selected with high frequency, suggesting that recent price movements contain persistent predictive information about future oil returns. This may reflect gradual information diffusion in oil markets or short-run continuation effects following macroeconomic and geopolitical shocks. Similar momentum effects in commodity futures markets have previously been documented by [23], who argue that commodity returns exhibit persistent

continuation patterns over short horizons.

The volatility forecasting exercise produced substantially larger forecasting gains than the return forecasting exercise, which is consistent with [24],which found that realized volatility tends to exhibit substantial persistence over time, making volatility forecasting inherently easier than forecasting returns. Variables related to uncertainty and market risk, particularly OVX and VIX, appear especially important in forecasting realised volatility. This highlights the importance of uncertainty and investor sentiment in oil markets, especially during periods of financial stress.

Among the volatility forecasting models, BMT and Elastic Net (60/40 split) achieved the strongest point forecasting performance, while LASSO (BIC) delivered the best directional accuracy. In particular, log_OVX was selected almost universally across all volatility models, confirming the dominant role of implied oil market uncertainty in predicting future realised volatility.

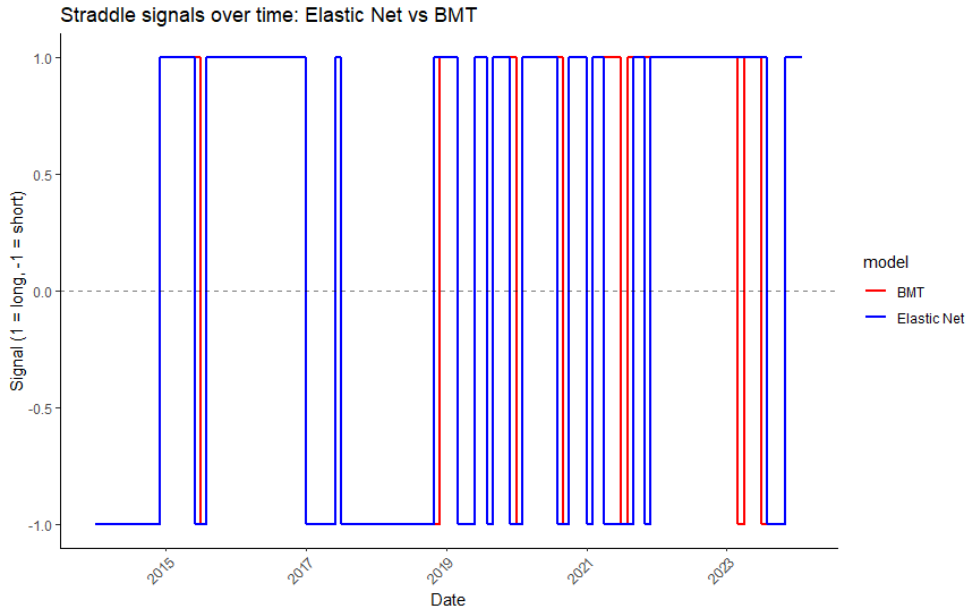


Figure 4.1: Straddle switching over time

Figure 4.1 illustrates the straddle trading signals generated by BMT and Elastic Net during the out-of-sample period. Both strategies switch positions frequently, changing signal in approximately 38.7% of months on average. While this demonstrates that the models respond actively to changing market conditions, the high turnover also raises concerns regarding practical implementation. The analysis assumes frictionless trading and therefore ignores transaction costs associated with repeatedly entering and exiting straddle positions. Under realistic trading costs, profitability and Sharpe ratios would decline materially, implying that the reported portfolio results should be interpreted as optimistic upper bounds.

By contrast, the benchmark strategy never switches position during the evaluation period. This occurs because the benchmark volatility forecast, based on a rolling historical average, remained above the fixed pre-2014 median threshold throughout the out-of-sample period. The result reflects a structural increase in oil market volatility following the 2014 oil price collapse, after which volatility remained persistently elevated. As argued by [1], major oil price shocks are often associated with prolonged periods of macroeconomic uncertainty and heightened financial volatility. Similarly, [2] emphasise that oil market expectations may shift persistently following large supply-demand disruptions. The COVID-19 pandemic further amplified these dynamics, generating unprecedented uncertainty and extreme fluctuations in crude oil markets, as documented by [25].

Another important limitation of the analysis concerns the assumptions underlying the estimation procedures. As mentioned in Section 2.6, Assumption L3 may be difficult to satisfy in practice because many oil market variables, uncertainty indices and macroeconomic indicators share strong common factor structures. Similarly, Assumption B2 may fail when several predictors proxy for the same underlying economic forces, making it difficult to distinguish true signals from highly correlated proxies.

More generally, Assumptions C3-C4 and C6 may be restrictive in our context. The QQ-plots in Figures C.32 and C.33, together with the estimated kurtosis values of 12.61 for oil returns and 6.36 for log realised volatility, indicate substantial heavy-tailed behaviour in both series. In particular, several observations in the extreme quantiles lie far from the reference line, suggesting that large market movements occur more frequently than implied by a Gaussian distribution. Log realised volatility additionally exhibits noticeable asymmetry and persistent periods of elevated volatility. Such characteristics are common in financial time series and have been extensively documented by [26]. Heavy tails, asymmetry and volatility clustering may therefore affect finite-sample performance and the reliability of the asymptotic approximations underlying the estimation procedures. Full-sized versions of the QQ-plots, together with the corresponding time series plots and histograms, are provided in Appendix C.5.

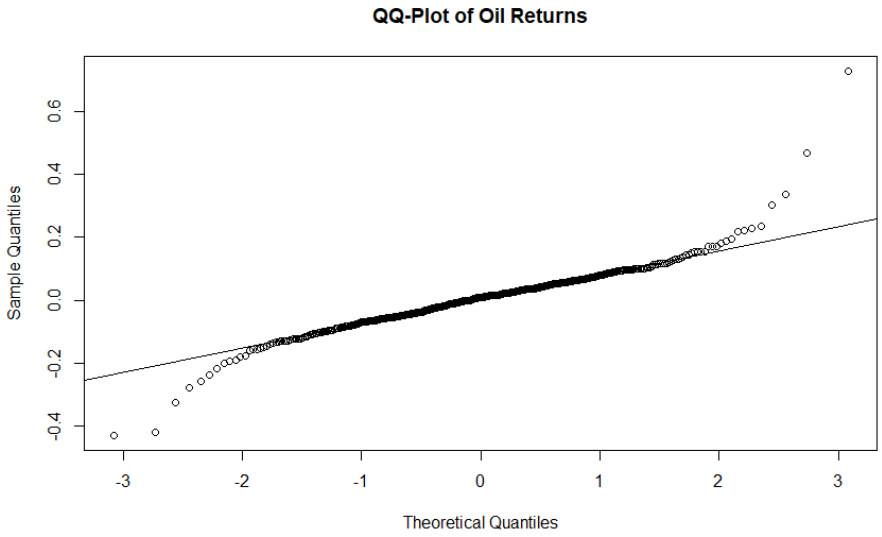


Figure 4.2: QQ-plot of oil returns

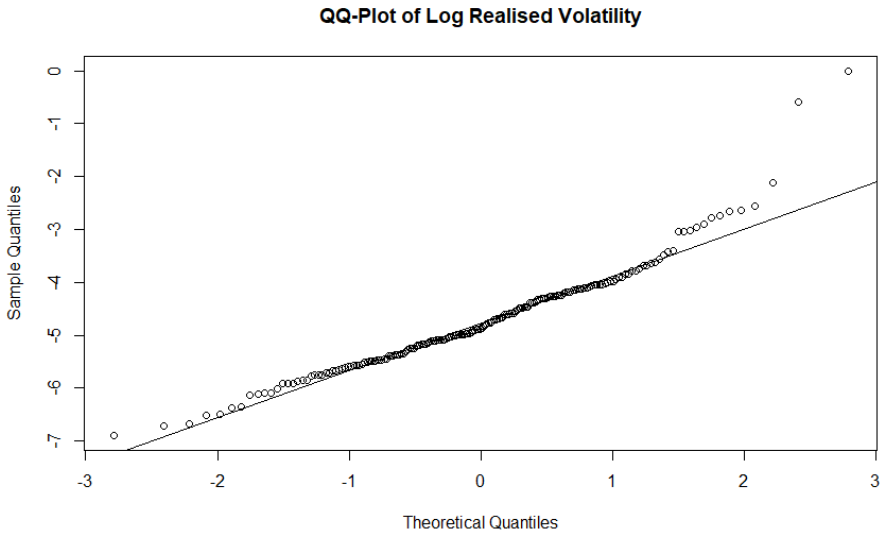


Figure 4.3: QQ-plot of log realised volatility

A natural extension of the present framework would be to incorporate regime-switching dynamics into the forecasting models. The empirical results suggest the presence of structural breaks and persistent shifts in volatility, particularly following the 2014 oil price collapse, where the benchmark volatility forecast remained consistently above its historical threshold. This indicates that the relationship between predictors and the target variable may differ across market regimes.

One possible extension would therefore be to allow both the coefficients and predictor relevance to vary between latent states, such as low- and high-volatility periods. A possible formulation is given by

$$r_{t+1} = \alpha_{S_t} + x_t' \beta_{S_t} + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim N(0, \sigma^2_{S_t}), \tag{4.1}$$

where $S_t \in \{1, 2\}$ denotes an unobserved Markov state representing different market regimes. In this framework, both the predictive coefficients and volatility dynamics are allowed to change across regimes, implying that certain predictors may only possess forecasting power during specific market conditions. The proposed extension is partly motivated by [27], who apply a Markov regime-switching framework to forecast volatility in the Chinese crude oil futures market.

Finally, the results demonstrate that predictive relationships in crude oil markets are highly sensitive to changing market conditions and periods of elevated uncertainty. Although no single forecasting method consistently outperforms the others across all evaluation criteria, both penalised regression methods and greedy selection algorithms provide substantial improvements relative to the benchmark, particularly in the volatility forecasting exercise. The analysis further indicates that predictor importance varies over time, with momentum and uncertainty-related variables emerging as especially influential across several model specifications. Overall, the findings highlight both the strengths and limitations of the four forecasting methods when applied to crude oil markets characterised by structural instability and persistent volatility.

5 Conclusion

This paper examined the extent to which penalised regression methods compare with greedy selection algorithms in forecasting crude oil returns and realised volatility out-of-sample. Using a set of macroeconomic, financial and uncertainty-related predictors, the performance of LASSO, Elastic Net, OCMT and BMT was first evaluated using statistical measures and subsequently assessed in an economic setting.

Across all specifications, the variable selection methods were found to improve forecasting performance relative to the historical benchmark, particularly in the volatility forecasting exercise. Momentum variables emerged as consistently important predictors of oil returns, while uncertainty measures such as OVX and VIX played a central role in forecasting realised volatility. No single method dominated across all evaluation criteria. In the return forecasting exercise, LASSO (BIC) and Elastic Net (CV) provided the strongest overall balance between forecast accuracy and directional performance, with BMT producing similar but slightly weaker results. By contrast, BMT achieved the strongest performance in the volatility forecasting exercise, which may reflect the greater persistence and predictability of realised volatility relative to returns.

Despite the improvements in statistical forecasting performance, these gains did not consistently translate into economically significant portfolio results. In particular, the generally weak CER gains and the high turnover of the trading strategies suggest that practical profitability may be substantially reduced once transaction costs are taken into account. Moreover, the VaR evaluation indicates that both the benchmark and model-based volatility forecasts tend to underestimate tail risk, as the observed violation rates exceed the nominal 5% level.

Several limitations of the analysis suggest directions for future research. The prevalence of highly correlated predictors challenges the theoretical assumptions underlying the selection methods, while structural breaks and persistent shifts in oil market volatility indicate that the underlying relationships may not be stable over time. Incorporating regime-switching dynamics may therefore improve both forecasting performance and economic relevance.

Overall, the findings demonstrate that both penalised regression methods and greedy selection algorithms are capable of extracting useful predictive information from a set of crude oil market predictors. While the methods differ in terms of sparsity, forecast accuracy and directional performance, no single approach uniformly dominates across all evaluation criteria. The results therefore highlight the trade-offs inherent in variable selection for financial forecasting, particularly in environments characterised by structural instability, elevated uncertainty and persistent volatility.

6 Bibliography

[1] Kilian L, Park C. The impact of oil price shocks on the U.S. stock market. International Economic Review; 2009.

[2] Baumeister C, Kilian L. Forecasting the Real Price of Oil in a Changing World: A Forecast Combination Approach. Journal of Business and Economics Statistics; 2013.

[3] Zhang Y, Ma F, Wang Y. Forecasting crude oil prices with a large set of predictors: Can LASSO select powerful predictors? Journal of Empirical Finance; 2019.

[4] Wen D, He M, Wang Y, Zhang Y. Forecasting crude oil market volatility: A comprehensive look at uncertainty variables. International Journal of Forecasting; 2024.

[5] Chudik, Kapetanios. A One-Covariate at a Time, Multiple Testing Approach to Variable Selection in High-Dimensional Linear Regression Models. Federal Reserve Bank of Dallas; 2018.

[6] Kapetanios G, Sarafidis V, Ventouri A. Model Selection in High-Dimensional Linear Regression using Boosting with Multiple Testing; 2026.

[7] Bühlmann P, Geer S. Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer; 2011.

[8] Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society; 1996.

[9] Zou H, Hastie T. Regularization and Variable Selection via the Elastic Net. Journal of the Royal Statistical Society: Series B; 2005.

[10] Fan J, Li R. Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. Journal of the American Statistical Association; 2001.

[11] Pesaran M, Timmermann A. A Simple Nonparametric Test of Predictive Performance. Journal of Business & Economic Statistics; 1992.

[12] Clark T, West K. Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. Journal of Econometrics; 2007.

[13] He F, Dong X, Qian Y, Xu X. Dynamic interactions among trade policy uncertainty, climate policy uncertainty, and crude oil prices. International Review of Economics Finance; 2024.

[14] Zhang G, Wang J, Liu Y. Energy transition in China: Is there a role for climate policy uncertainty? Journal of Environmental Management; 2024.

[15] Yao CZ, Liu Y. Crude oil volatility forecasting: New evidence from world uncertainty index. Finance Research Letters; 2023.

[16] Liu T, Zhang Y, Zhang W, Hamori S. Quantile Connectedness of Uncertainty Indices, Carbon Emissions, Energy, and Green Assets: Insights from Extreme Market Conditions. Energies; 2024.

[17] Gao R, Zhao Y, Zhang B. Baltic dry index and global economic policy uncertainty: evidence from the linear and nonlinear Granger causality tests. Taylor and Francis; 2021.

[18] Campos I, Cortazar G, Reyes T. Modeling and predicting oil VIX: Internet search volume versus traditional mariables. Energy Economics; 2017.

[19] Chen X, Feng L, Wang Y. Pricing VIX futures: A framework with random level shifts. Finance Research Letters; 2022.

[20] Gong X, Ye X, Zhang W, Zhang Y. Predicting energy futures high-frequency volatility using technical indicators: The role of interaction. Energy Economics; 2023.

[21] Gurrola-Perez P, Murphy D. Filtered historical simulation Value-at-Risk models and their competitors. Bank of England; 2015.

[22] Sheu HJ, Wei YC. Options Trading Based on the Forecasting of Volatility Direction with the Incorporation of Investor Sentiment. Emerging Markets Finance and Trade; 2011.

[23] Fuertes AM, Miffre J, Rallis G. Tactical allocation in commodity futures markets: Combining momentum and term structure signals. Journal of banking and finance; 2010.

[24] Andersen T, Bollerslev T, Diebold F, Labys P. Modeling and Forecasting Realized Volatility. Econometrica; 2003.

[25] Sharif A, Aloui C, Yarovaya L. COVID-19 pandemic, oil prices, stock market, geopolitical risk and policy uncertainty nexus in the US economy: Fresh evidence from the wavelet-based approach. International Review of Financial Analysis; 2020.

[26] Cont R. Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. Taylor & Francis Online; 2010.

[27] Qiao G, Pan Y, Liang C. Forecasting Volatility in Chinese Crude Oil Futures: Insights from Volatility-of-Volatility and Markov Regime-Switching Approaches. Taylor & Francis; 2024.

[28] Hall P, Heyde C. Martingale Limit Theory and Its Application. Academic Press; 1980.

[29] White H. Asymptotic Theory for Econometricians. Academic Press; 1984.

[30] Billingsley P. Convergence of probability measures. John Wiley Sons; 1999.

[31] der Vaart V. Asymptotic Statistics. Cambridge University Press; 2000.

A | Theorems used

This appendix collects the theorems invoked in the paper.

Theorem A.0.1 (Martingale CLT)**.** Let $\{X_{n,i}, \mathcal{F}_{n,i}\}$ be a martingale difference array with $\mathbb{E}[X_{n,i}^2] < \infty$ for all n, i. Assume that:

(i) $\max_{1 \le i \le k_n} |X_{n,i}| \xrightarrow{p} 0$;

(ii) $\sup_n \mathbb{E}\big[\max_{1 \le i \le k_n} X_{n,i}^2\big] < \infty$;

(iii) $\sum_{i=1}^{k_n} X_{n,i}^2 \xrightarrow{p} \eta^2$ for some random variable η^2 that is finite a.s. and constant.

Then

$$\sum_{i=1}^{k_n} X_{n,i} \xrightarrow{d} Z,$$

where $Z = \eta N$ with $N \sim \mathcal{N}(0,1)$ independent of η. In particular, if η^2 is a constant $\sigma^2 > 0$, then

$$\frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} X_{n,i} \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

Thus, the normalized sum is $O_p(1)$.

Theorem 3.2 in [28]

Theorem A.0.2 (Weak LLN for mixing processes)**.** Let $\{X_t\}_{t=1}^{\infty}$ be a stationary and strongly mixing sequence with mixing coefficients $\alpha(t)$ satisfying $\sum_{t=1}^{\infty} \alpha(t)^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$. Assume $\mathbb{E}|X_t|^{2+\delta} < \infty$. Then

$$\frac{1}{n} \sum_{t=1}^{n} (X_t - \mathbb{E}[X_t]) \xrightarrow{p} 0$$

as $n \to \infty$.

Under Assumption C3, the mixing coefficients decay exponentially, i.e., $\alpha(t) \le C\xi^t$ for some $0 < \xi < 1$. For such processes, the condition $\sum_{t=1}^{\infty} \alpha(t)^{\delta/(2+\delta)} < \infty$ is satisfied for any $\delta > 0$.

Theorem 3.47 in [29]

Theorem A.0.3 (Mixing CLT for martingale diff.)**.** Let $\{Z_t\}_{t=1}^{\infty}$ be a stationary, strongly mixing sequence of martingale differences with respect to an increasing filtration $\mathcal{F}_t$ (i.e., $\mathbb{E}[Z_t \mid \mathcal{F}_{t-1}] = 0$). Suppose the mixing coefficients $\alpha(t)$ decay exponentially fast and $\mathbb{E}[Z_t^2] < \infty$. Define $\sigma^2 = \mathbb{E}[Z_t^2]$. Then

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_t \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

as $n \to \infty$.

Theorem 5.24 in [29]

Theorem A.0.4 (CMT)**.** Let X_n and X be random vectors taking values in a metric space D, and let $g : D \to \mathbb{R}^m$ be a continuous function. If $X_n \xrightarrow{d} X$ then $g(X_n) \xrightarrow{d} g(X)$. If $X_n \xrightarrow{p} X$ then $g(X_n) \xrightarrow{p} g(X)$.

Theorem 2.1 in [30]

Theorem A.0.5 (Slutsky's)**.** Let X_n and Y_n be sequences of random vectors, and let a and b be constants. If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} a$, then:

1. $X_n + Y_n \xrightarrow{d} X + a$;
2. $Y_n X_n \xrightarrow{d} aX$;
3. $Y_n^{-1} X_n \xrightarrow{d} a^{-1} X$ whenever a is invertible.

Lemma 2.8 in [31]

The following theorem is used for convergence in distribution on an event with probability tending to one.

Theorem A.0.6. Let X_n and Y_n be random variables such that $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} 0$. Then $X_n + Y_n \xrightarrow{d} X$. More generally, if $X_n \xrightarrow{d} X$ and $\Pr(\|X_n - Y_n\| > \varepsilon) \to 0$ for every $\varepsilon > 0$, then $Y_n \xrightarrow{d} X$.

Lemma 2.12 in [31].

Lemma A.0.7 (Growth rate of OCMT critical values)**.** Let

$$c_p(q, \delta) = \Phi^{-1}\left(1 - \frac{q}{2c_q p^{\delta}}\right),$$

where $q \in (0, 1)$, $c_q > 0$, and $\delta > 0$. Then

$$c_p(q, \delta) = O(\sqrt{\log p}) \qquad \text{as } p \to \infty.$$

Lemma A2 in the Online Supplement of [5].

Lemma A.0.8 (Tail bound for null t-statistics)**.** Suppose Assumptions C2–C6 hold. Then there exist constants $C, \omega > 0$ such that for any null predictor $j \in \mathcal{N}$,

$$\Pr\big(|t_j - \nu_j| > x\big) \leq Ce^{-\omega x^2},$$

for all sufficiently large x, where ν_j denotes the non-centrality parameter of the statistic.

Lemma A10 in the Online Supplement of [5].

Lemma A.0.9 (Hidden signals)**.** Suppose Assumptions C1-C6 hold. If a true signal $i \in S$ satisfies

$$\theta_i = 0$$

in an early stage of OCMT, then there exists a later stage $k \geq 2$ such that the conditional net effect becomes nonzero:

$$\theta_i^{(k)} \neq 0.$$

Hence every hidden signal becomes detectable after conditioning on previously selected predictors.

Lemma A1 in the Online Supplement of [5].

Proposition A.0.10 (Finite termination of OCMT)**.** Suppose Assumptions C1-C6 hold. Then the OCMT procedure terminates after a finite number of stages with probability approaching one.

Proposition 1 in the Online Supplement of [5].

Lemma A.0.11 (Bound on the number of stages)**.** Suppose Assumptions C1-C6 hold. Let K_n denote the total number of OCMT stages. Then

$$K_n = O_p(1).$$

Hence the total number of stages is bounded in probability.

Lemma A20 in the Online Supplement of [5].

B First dataset plots

B.1 LASSO

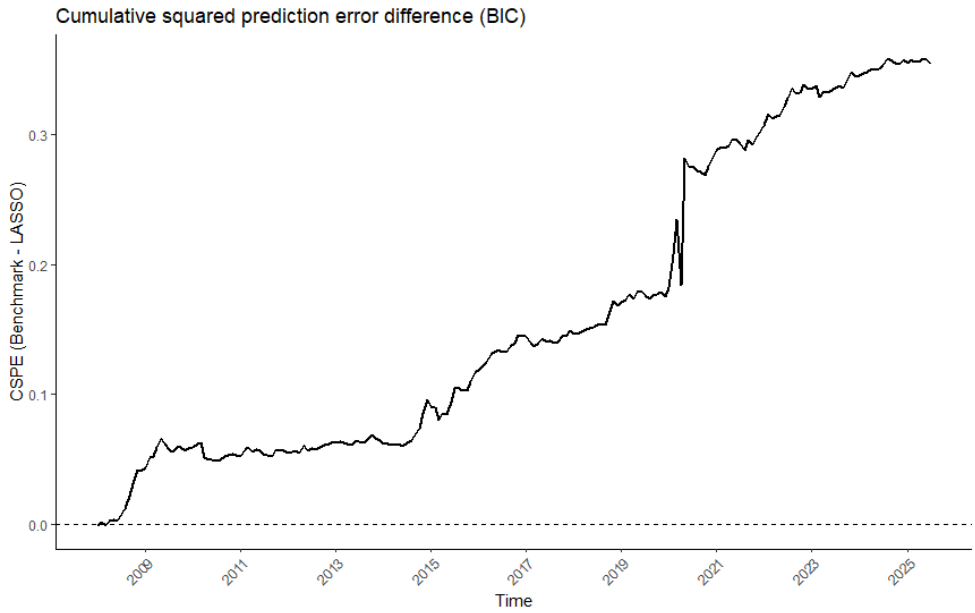


Figure B.1: LASSO (BIC): CSPE.

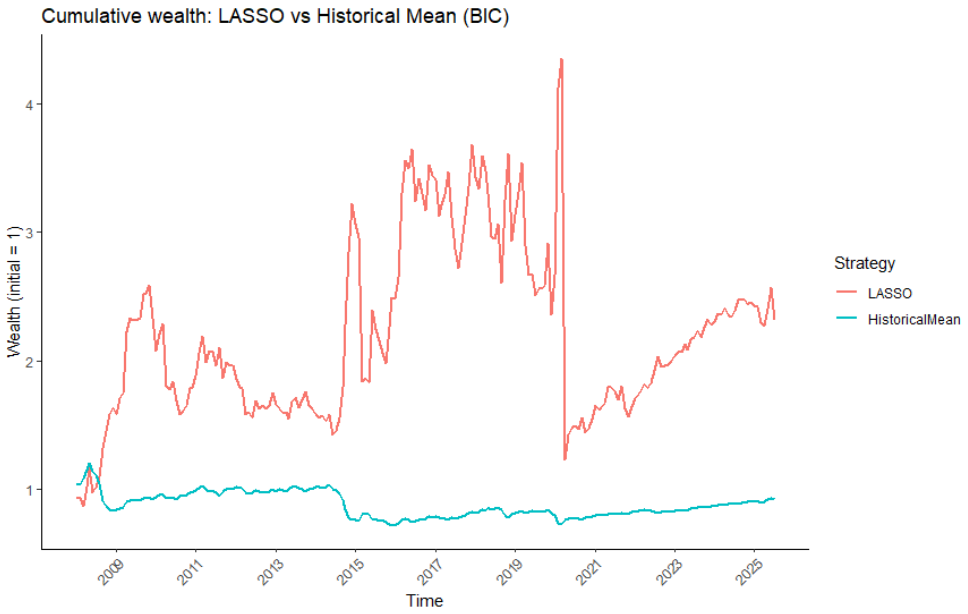


Figure B.2: LASSO (BIC): cumulative wealth.

Variable selection heatmaps

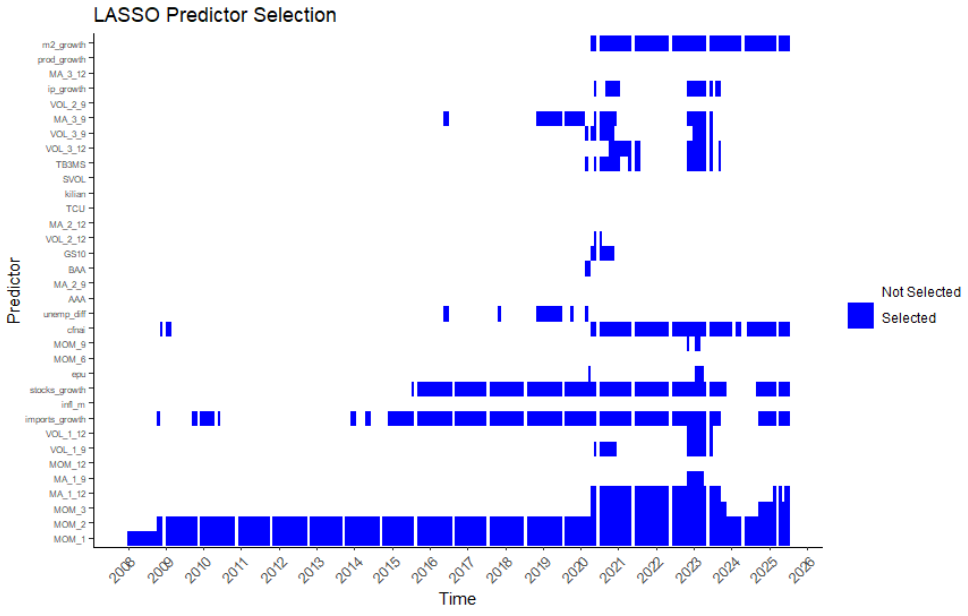


Figure B.3: LASSO (standard): heatmap.

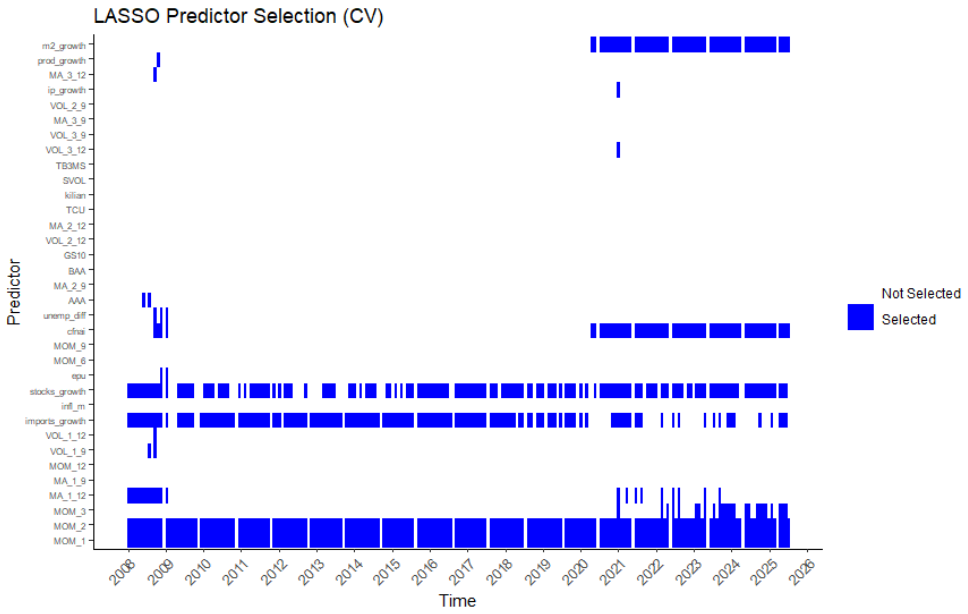


Figure B.4: LASSO (CV): heatmap.

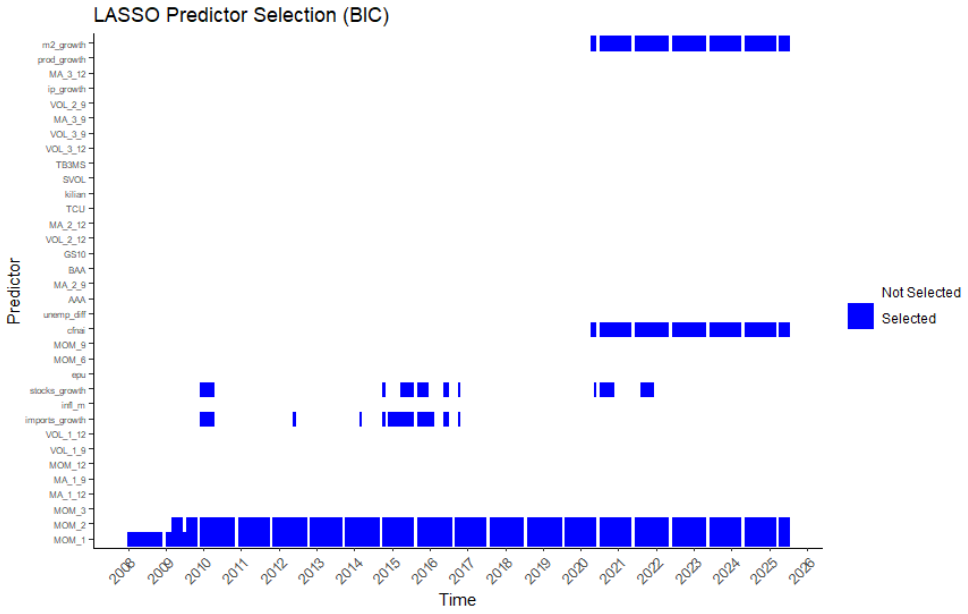


Figure B.5: LASSO (BIC): heatmap.

λ evolution

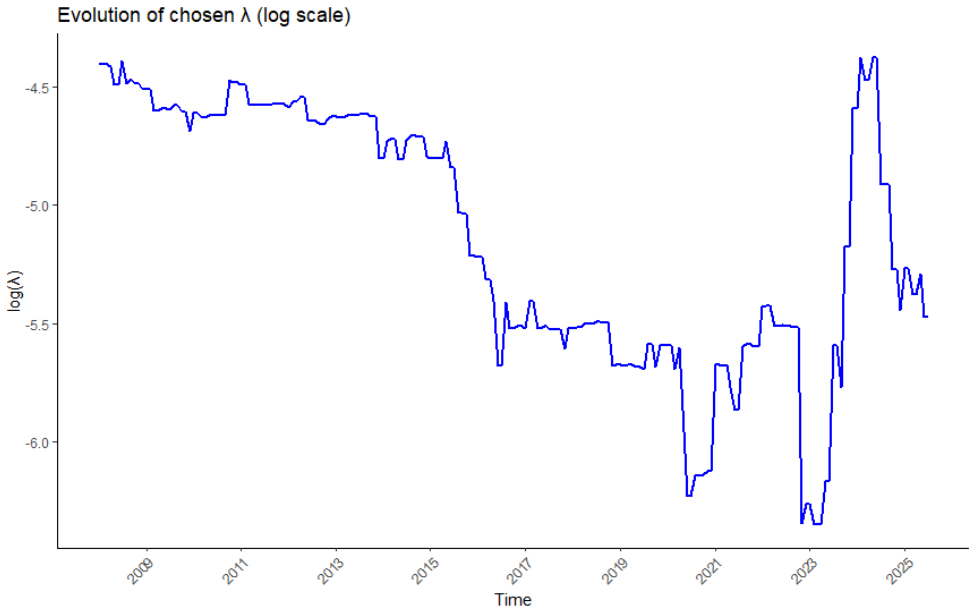


Figure B.6: LASSO (standard): λ evolution.

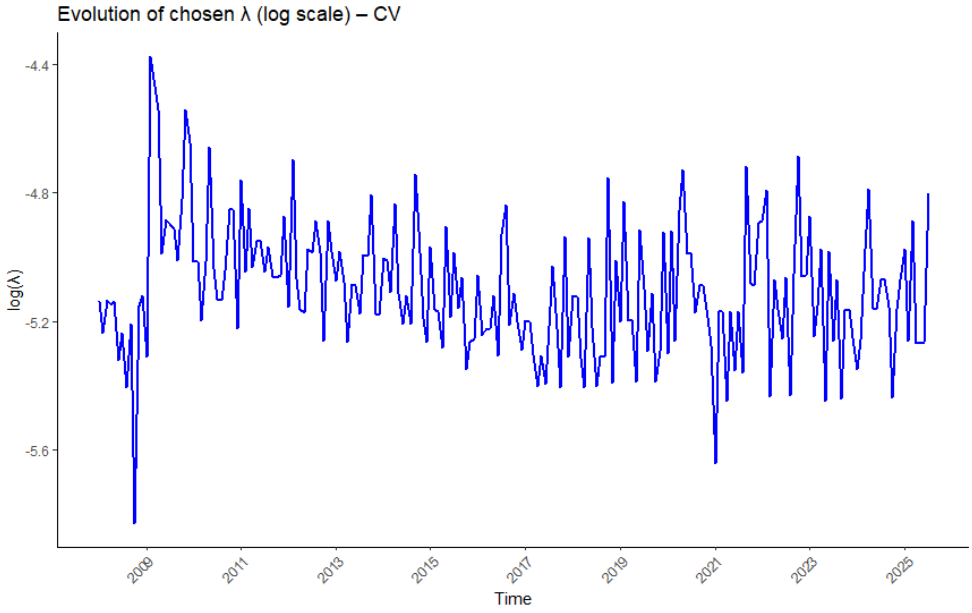


Figure B.7: LASSO (CV): λ evolution.

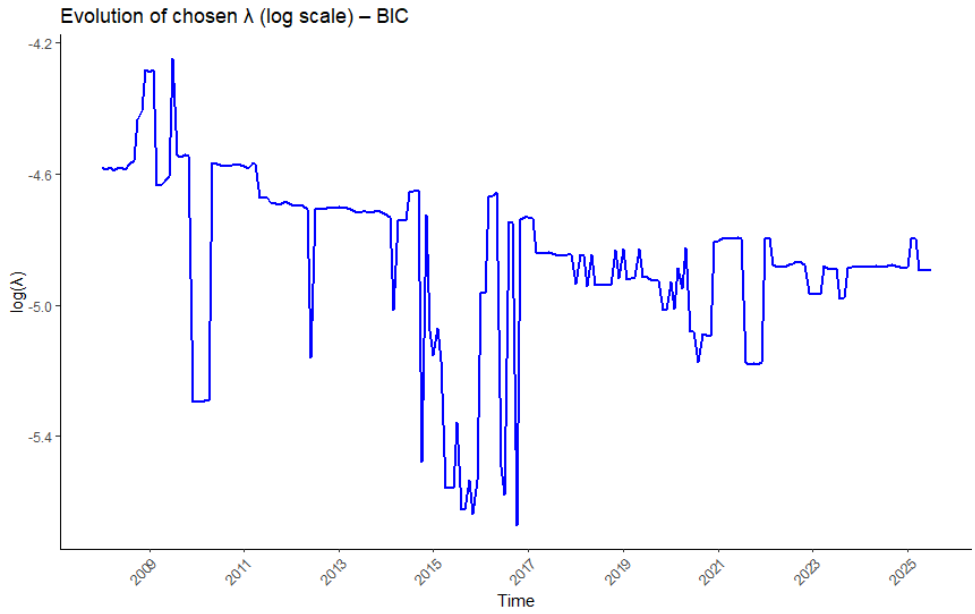


Figure B.8: LASSO (BIC): λ evolution.

B.2 Elastic Net

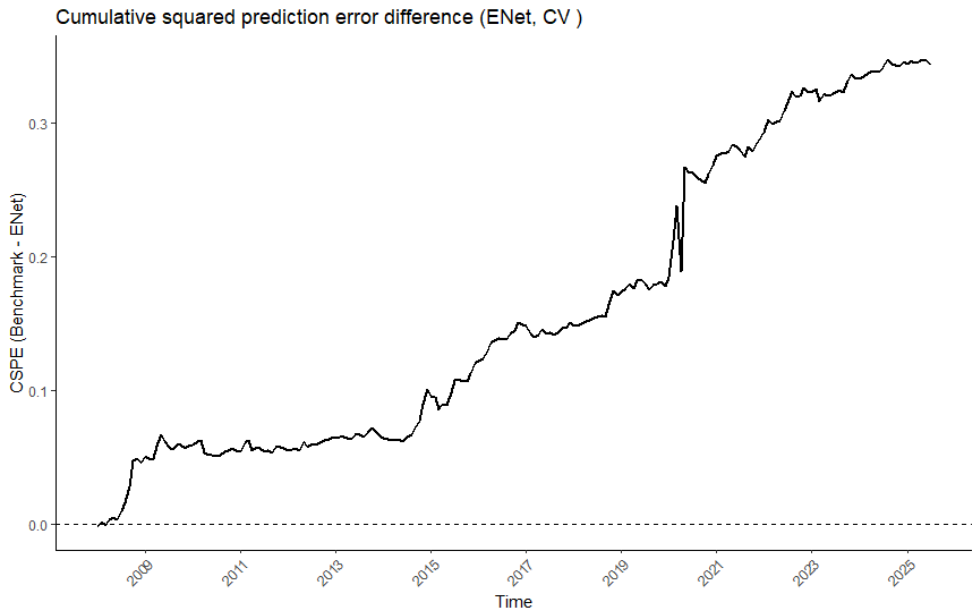


Figure B.9: Elastic Net (CV): CSPE.

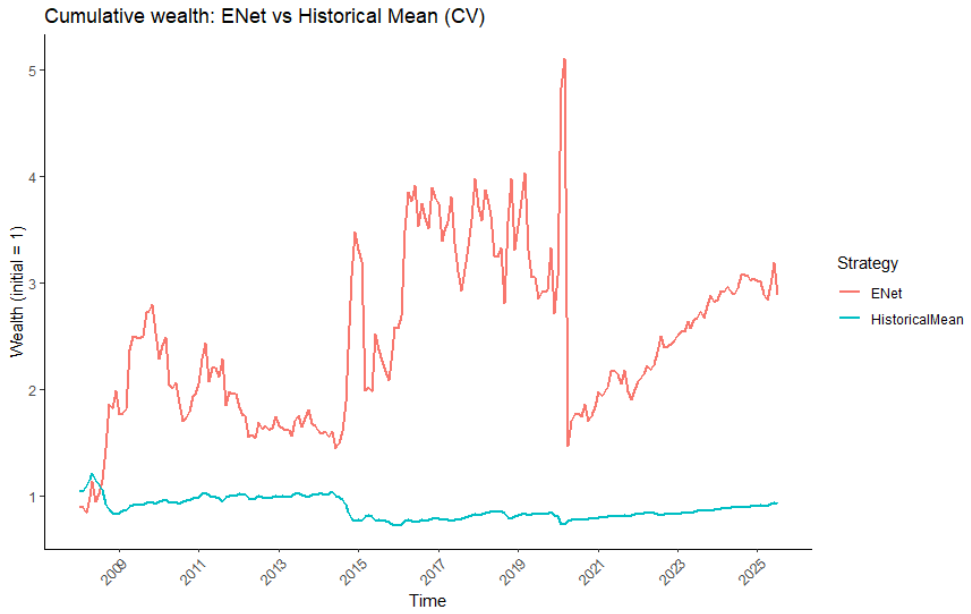


Figure B.10: Elastic Net (CV): cumulative wealth.

Variable selection heatmaps

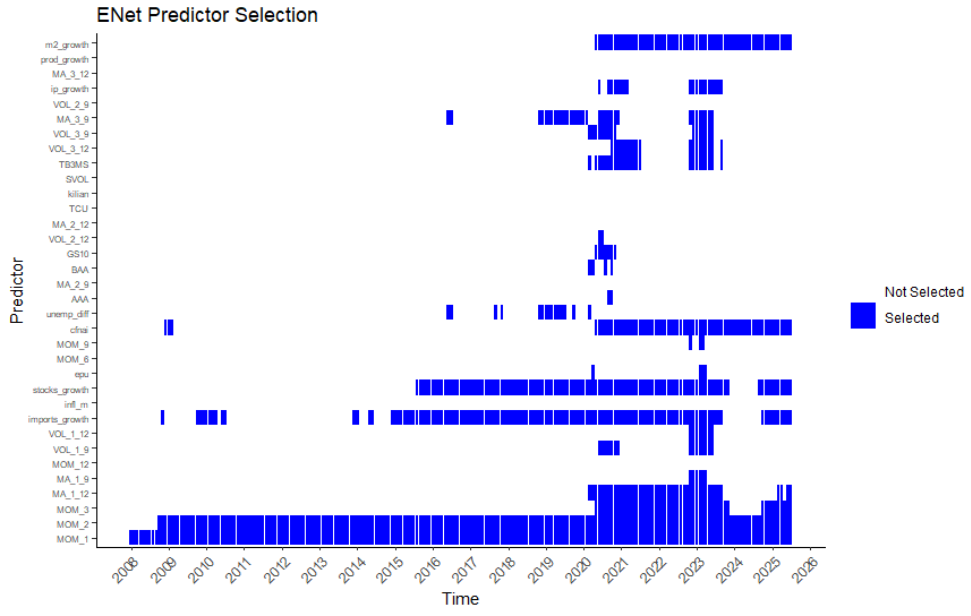


Figure B.11: Elastic Net (standard): heatmap.

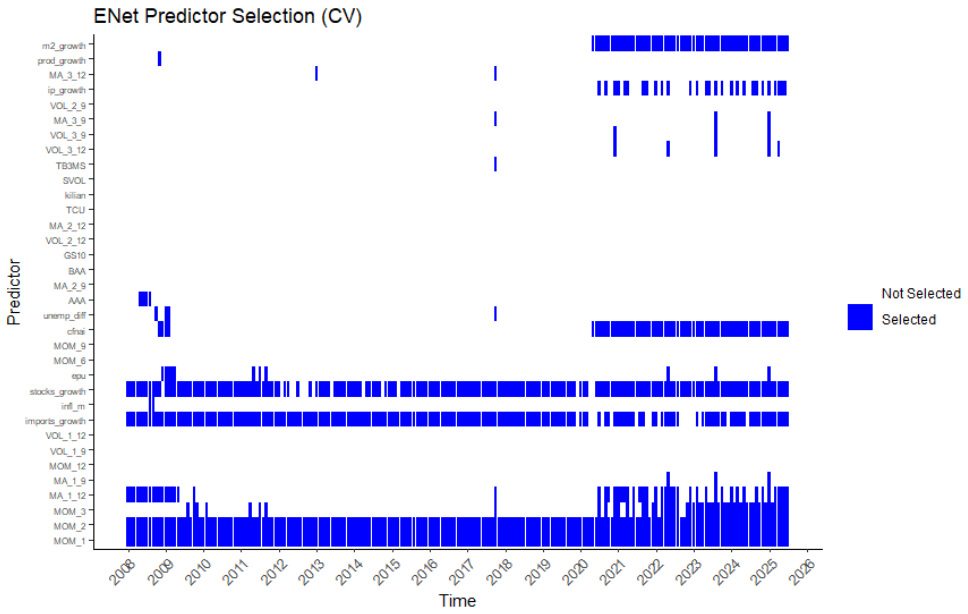


Figure B.12: Elastic Net (CV): heatmap.

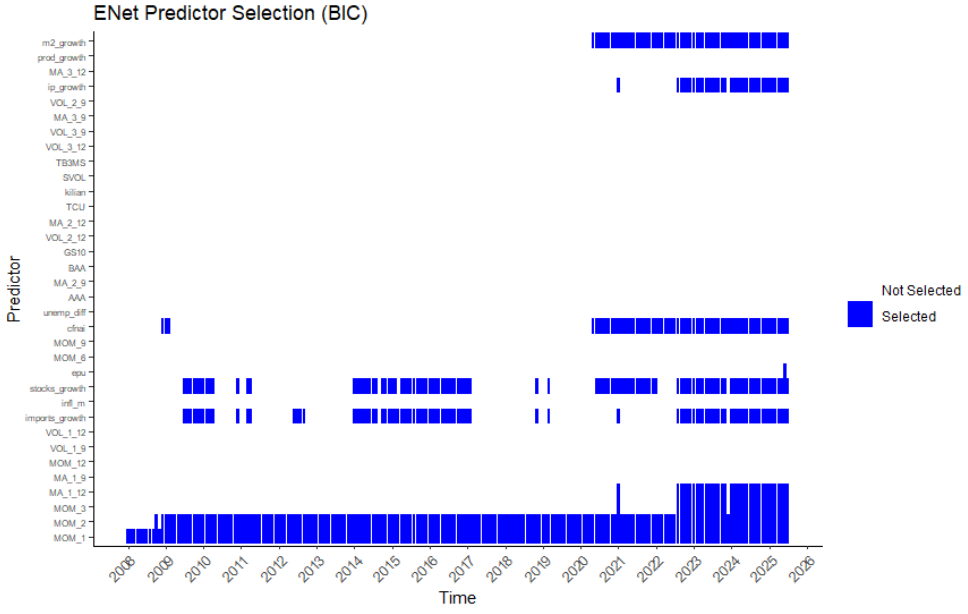


Figure B.13: Elastic Net (BIC): heatmap.

λ evolution

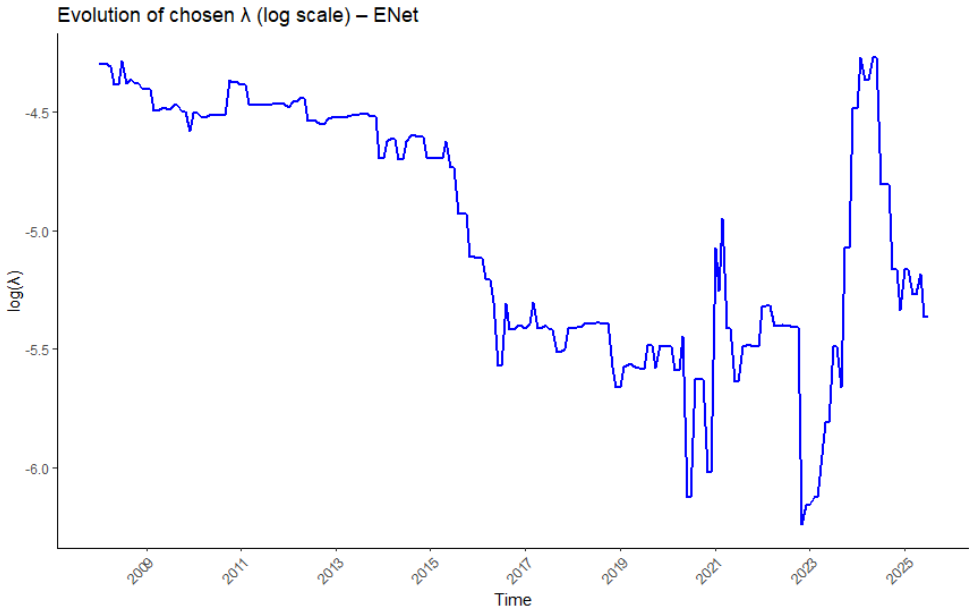


Figure B.14: Elastic Net (standard): λ evolution.

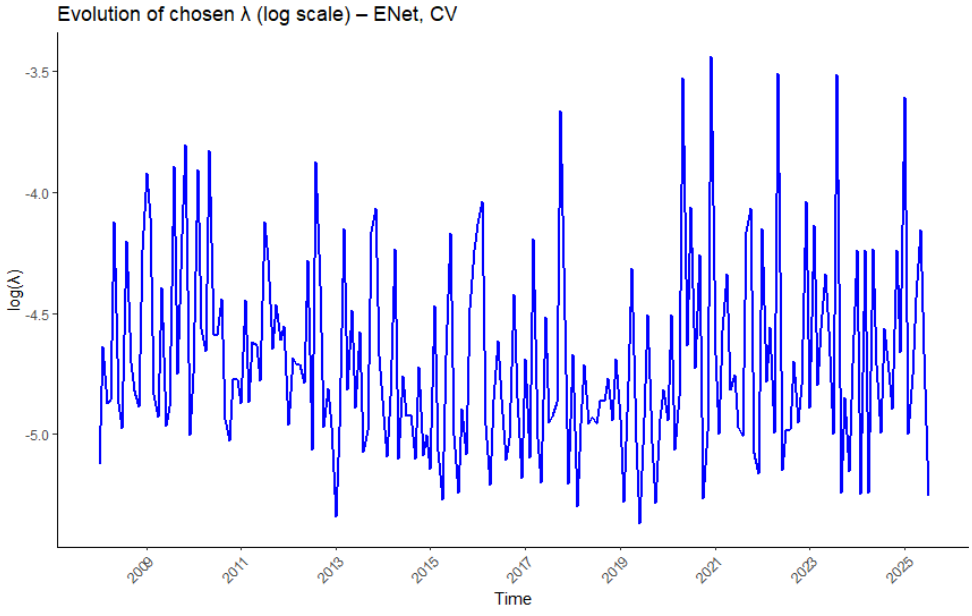


Figure B.15: Elastic Net (CV): λ evolution.

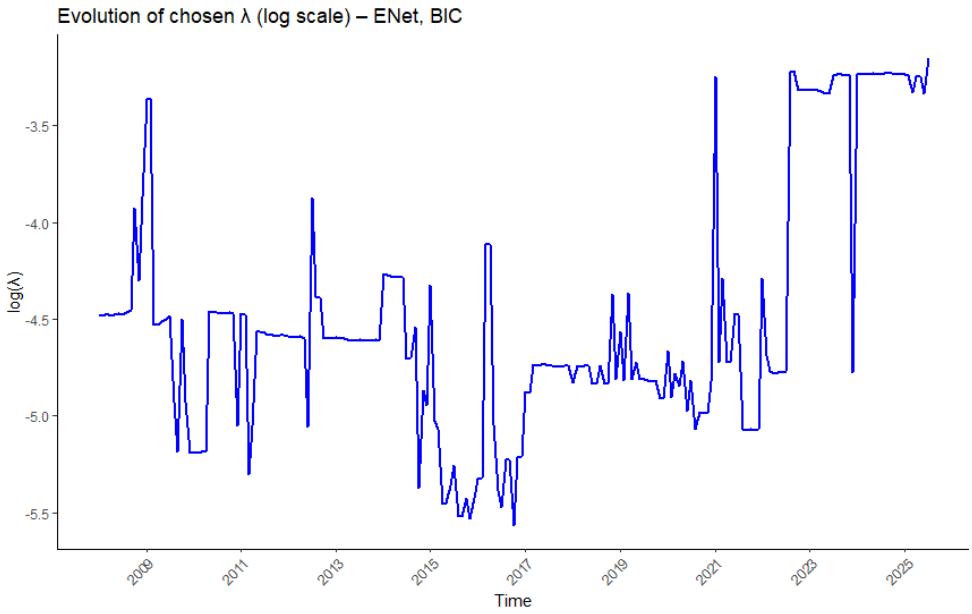


Figure B.16: Elastic Net (BIC): λ evolution.

α evolution

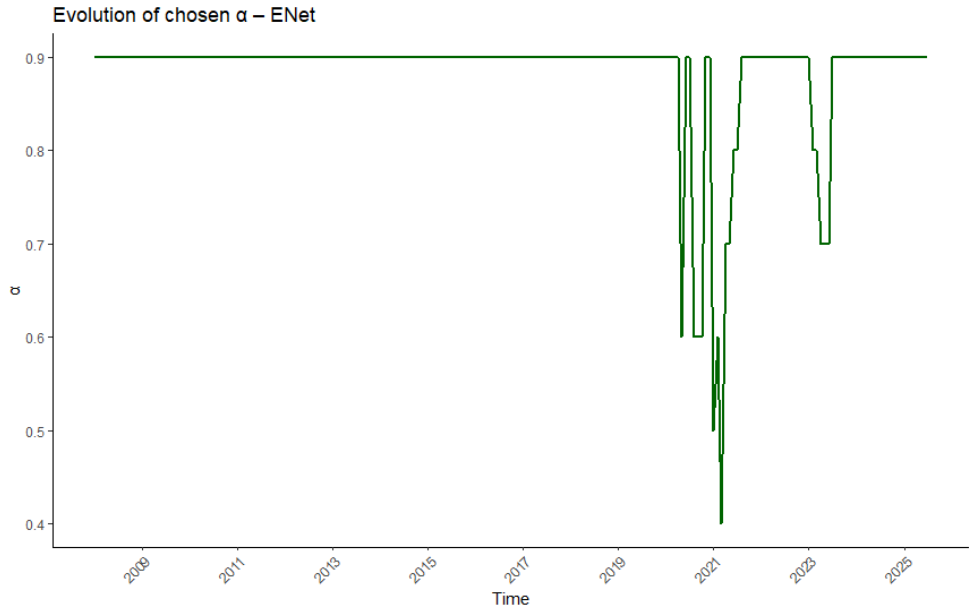


Figure B.17: Elastic Net (standard): α evolution.

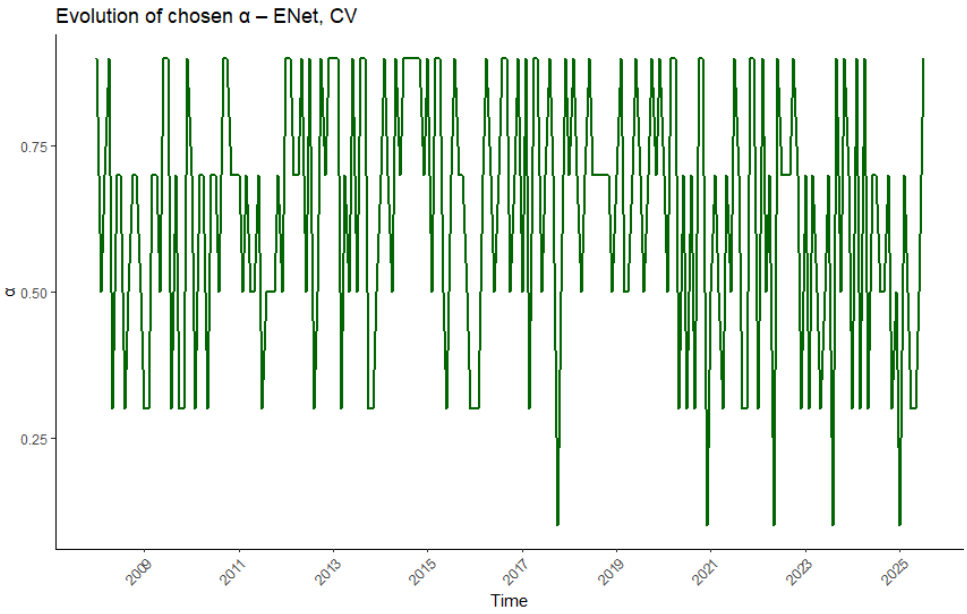


Figure B.18: Elastic Net (CV): α evolution.

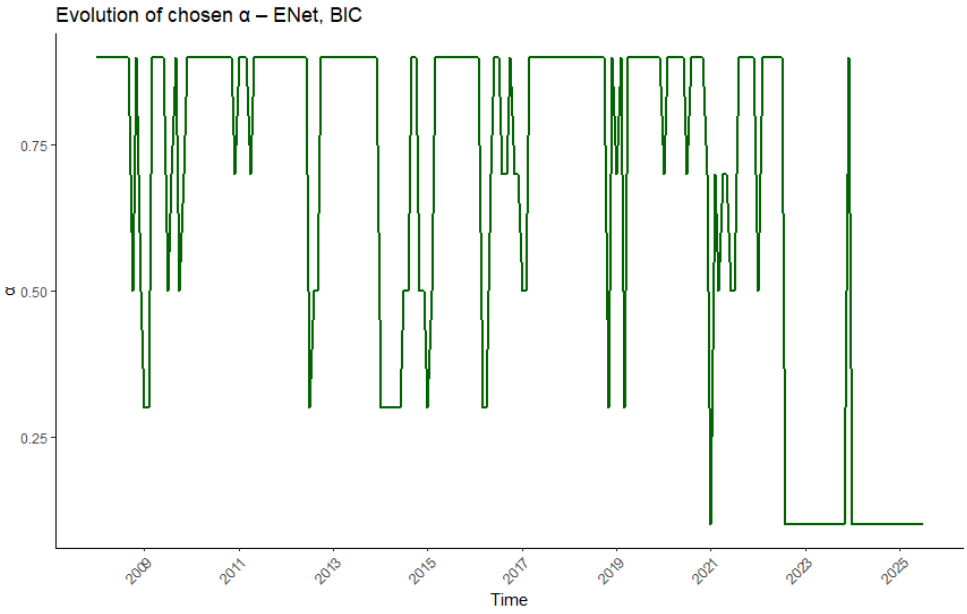


Figure B.19: Elastic Net (BIC): α evolution.

B.3 OCMT

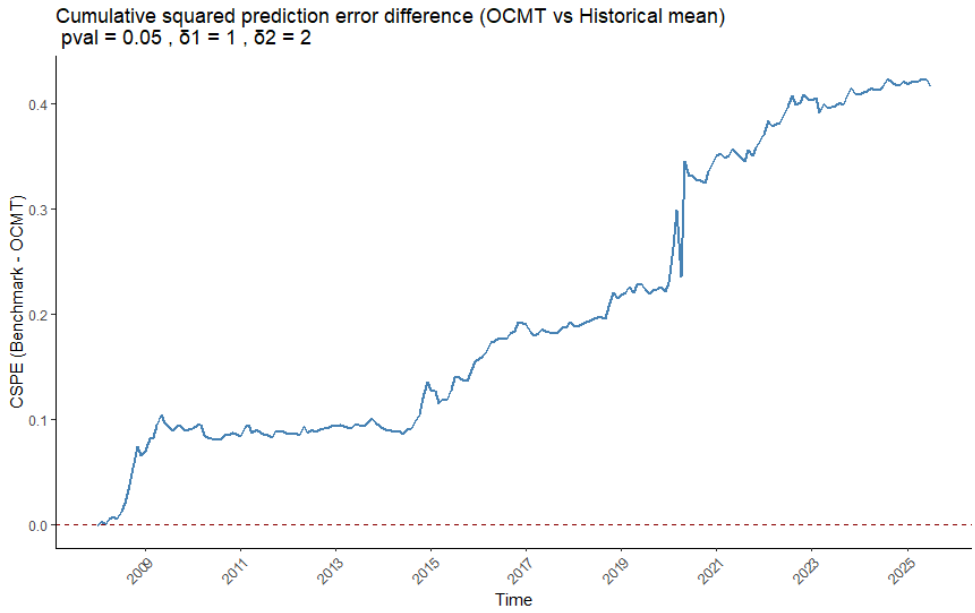


Figure B.20: OCMT: CSPE.

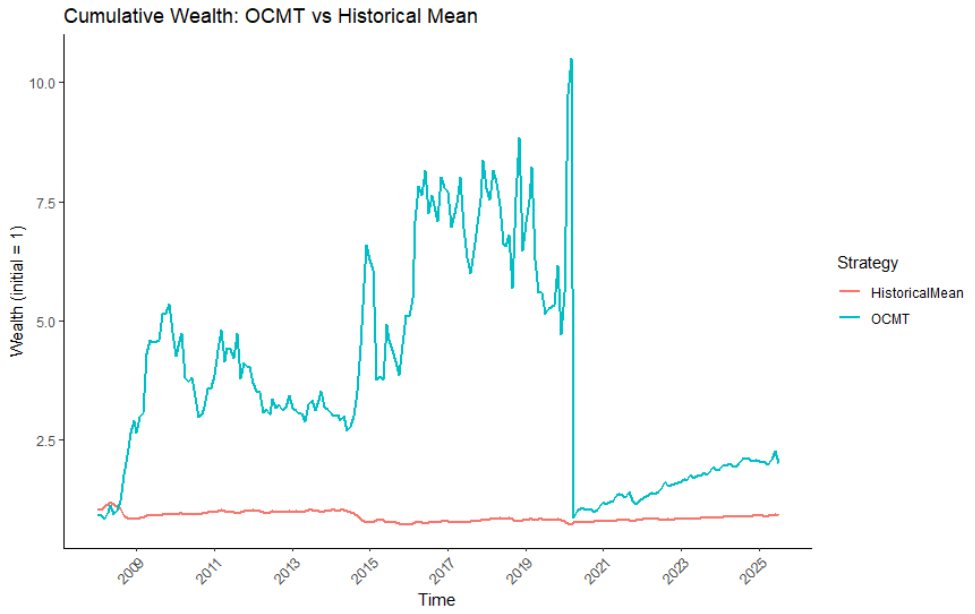


Figure B.21: OCMT: cumulative wealth.

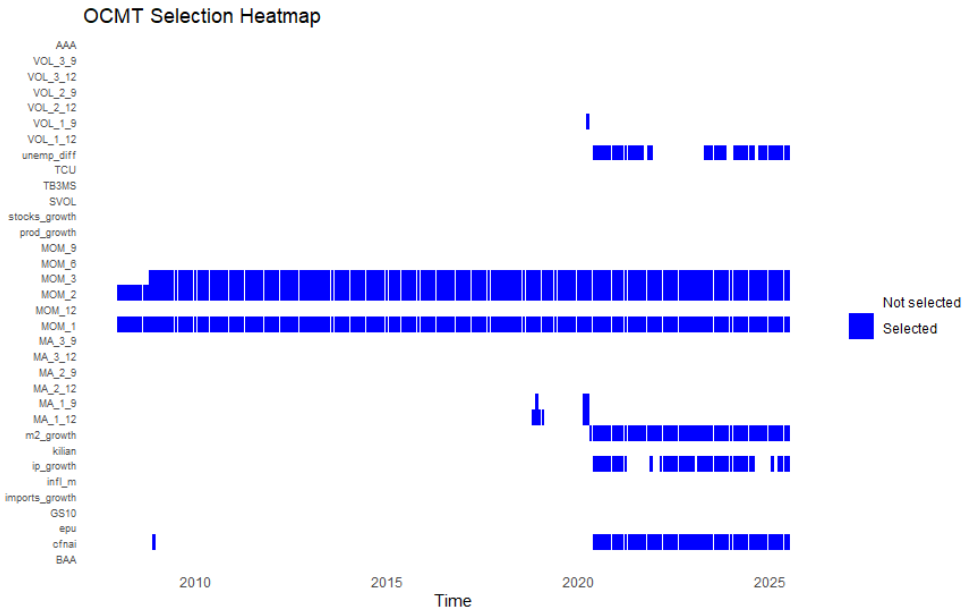


Figure B.22: OCMT: variable selection heatmap.

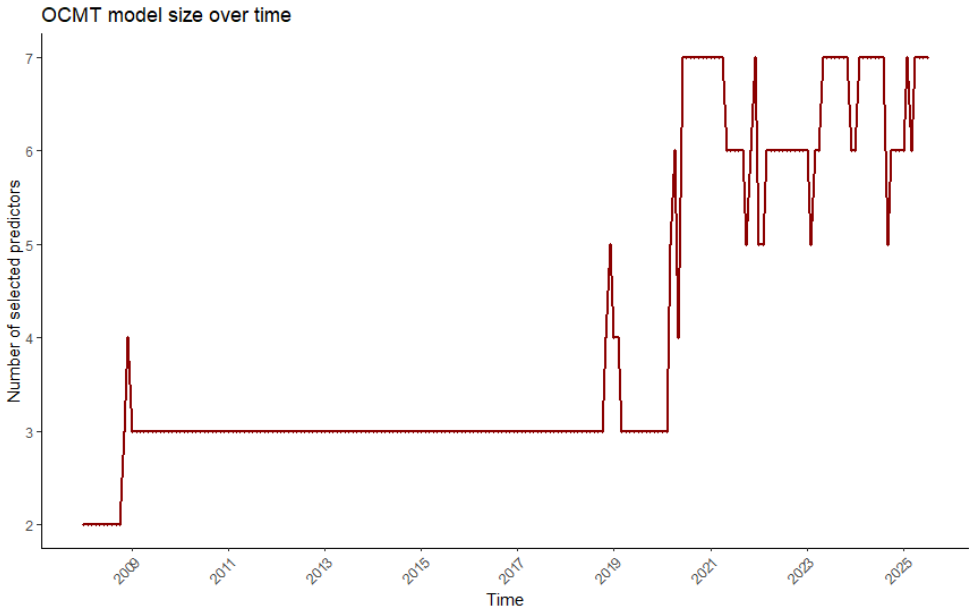


Figure B.23: OCMT: model size evolution.

B.4 BMT

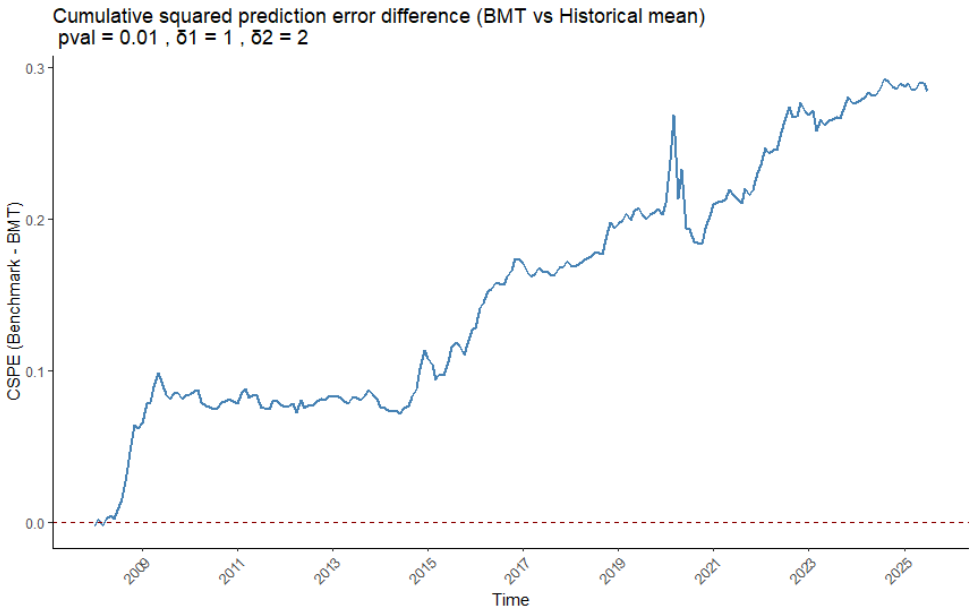


Figure B.24: BMT: CSPE.

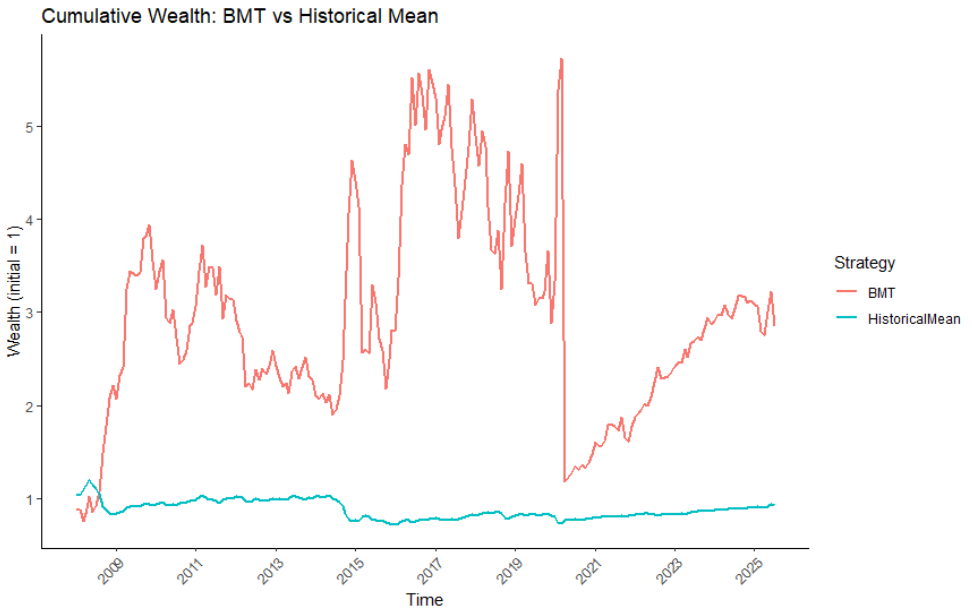


Figure B.25: BMT: cumulative wealth.

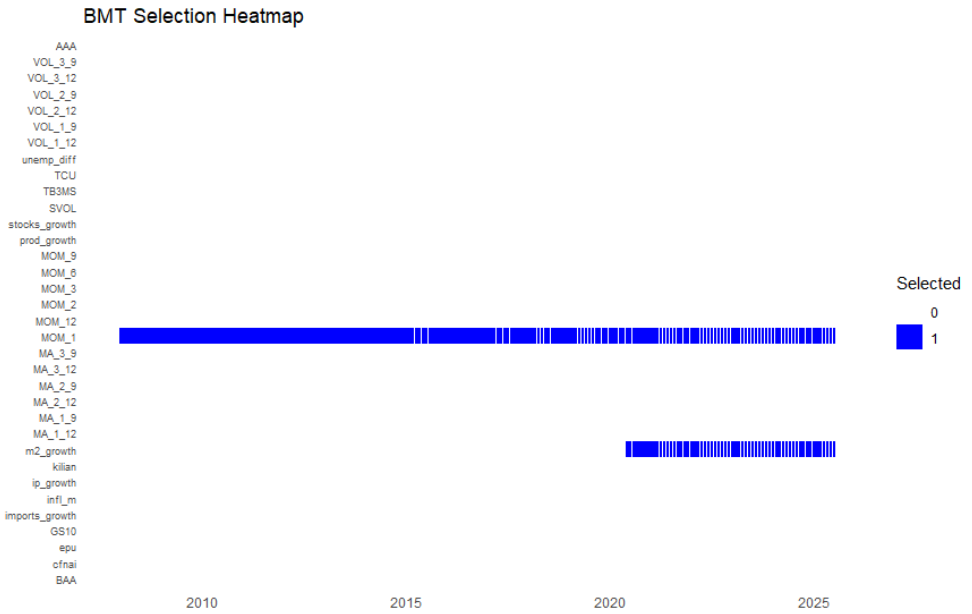


Figure B.26: BMT: variable selection heatmap.

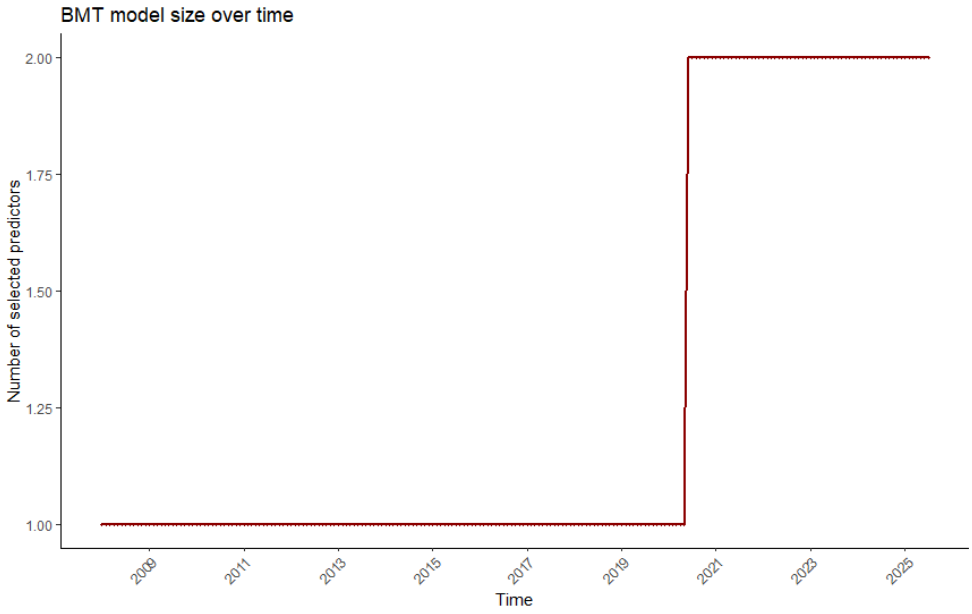


Figure B.27: BMT: model size evolution.

C Second dataset plots

C.1 LASSO

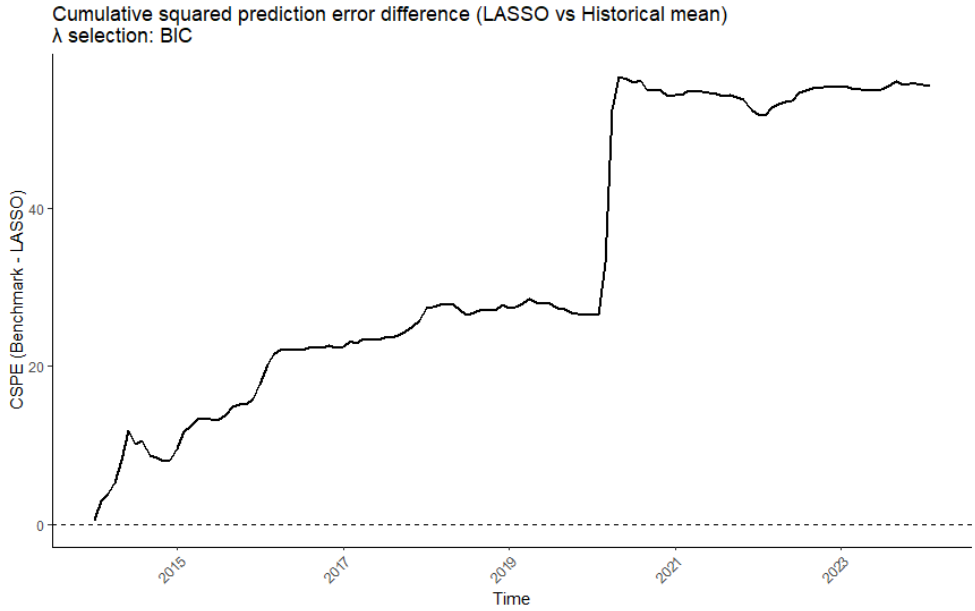


Figure C.1: LASSO (BIC): CSPE.

Variable selection heatmaps

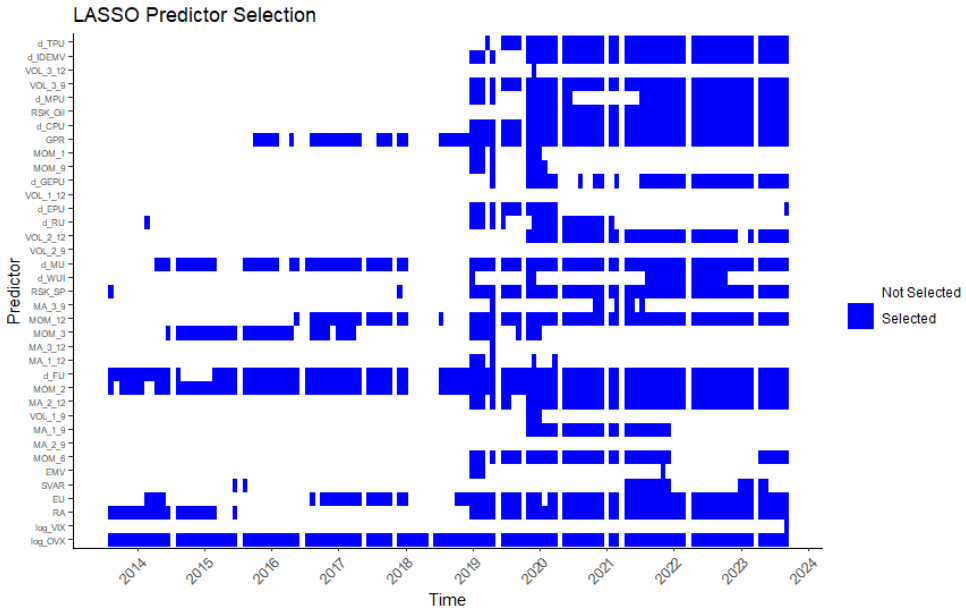


Figure C.2: LASSO (standard): heatmap.

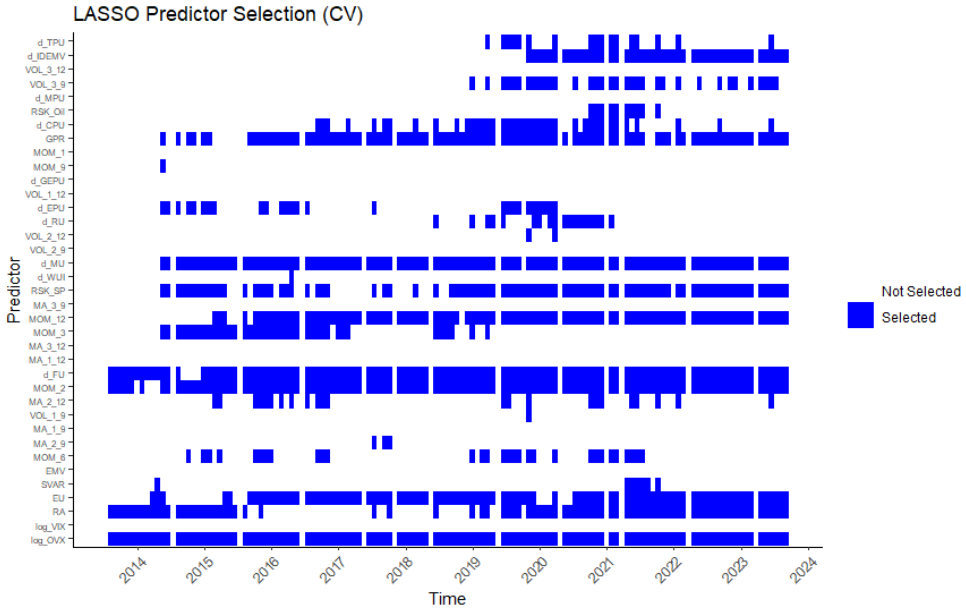


Figure C.3: LASSO (CV): heatmap.

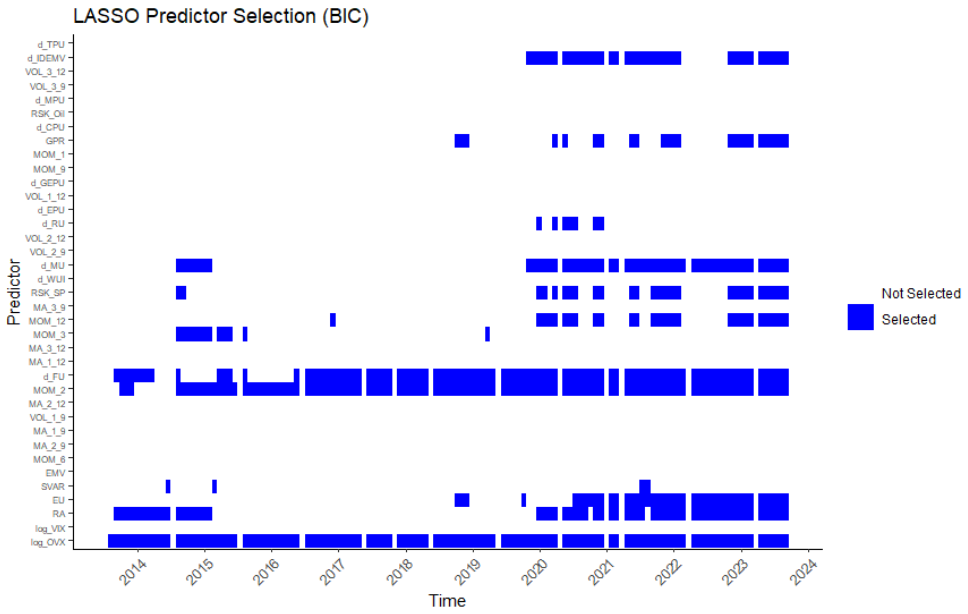


Figure C.4: LASSO (BIC): heatmap.

λ evolution

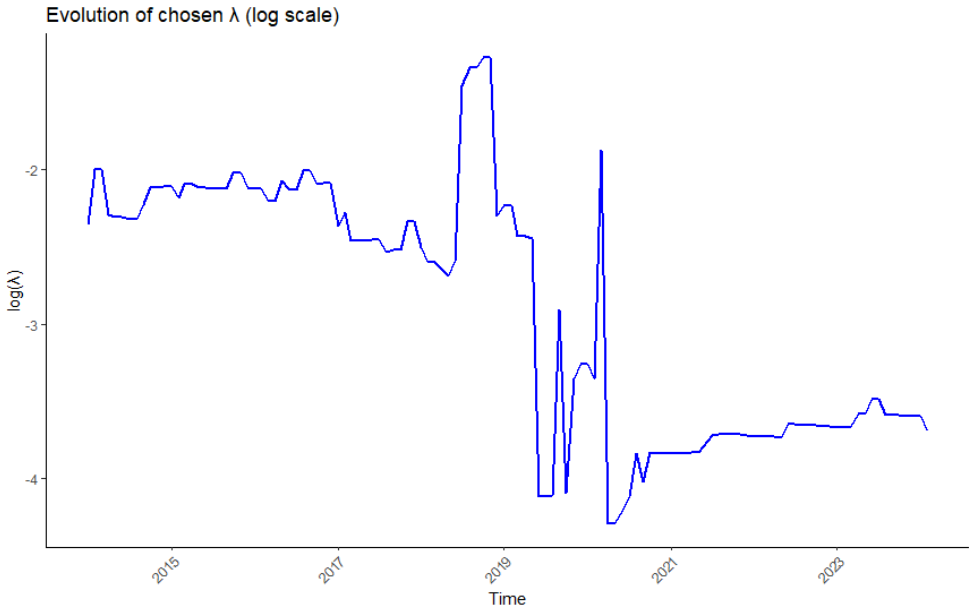


Figure C.5: LASSO (standard): λ evolution.

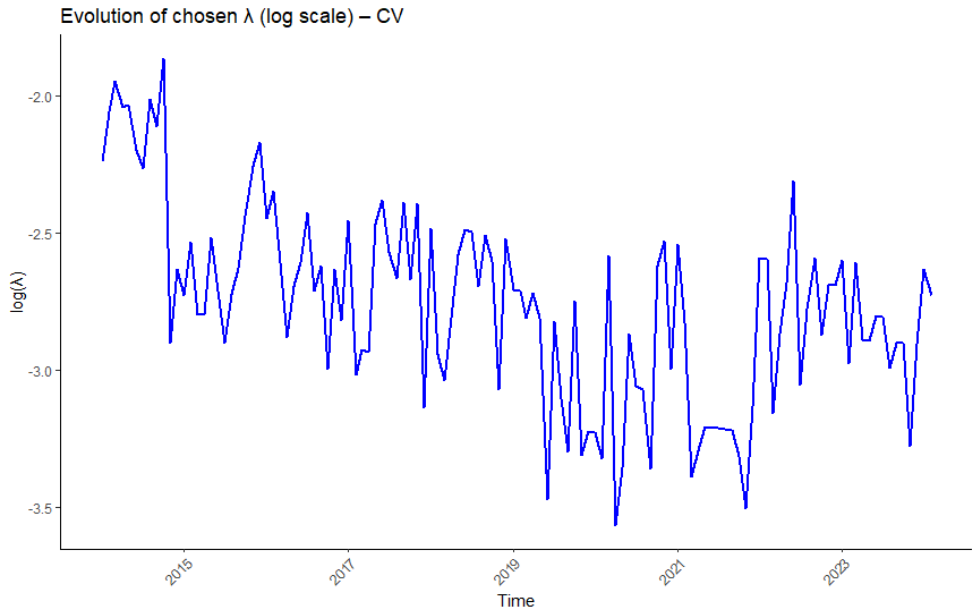


Figure C.6: LASSO (CV): λ evolution.

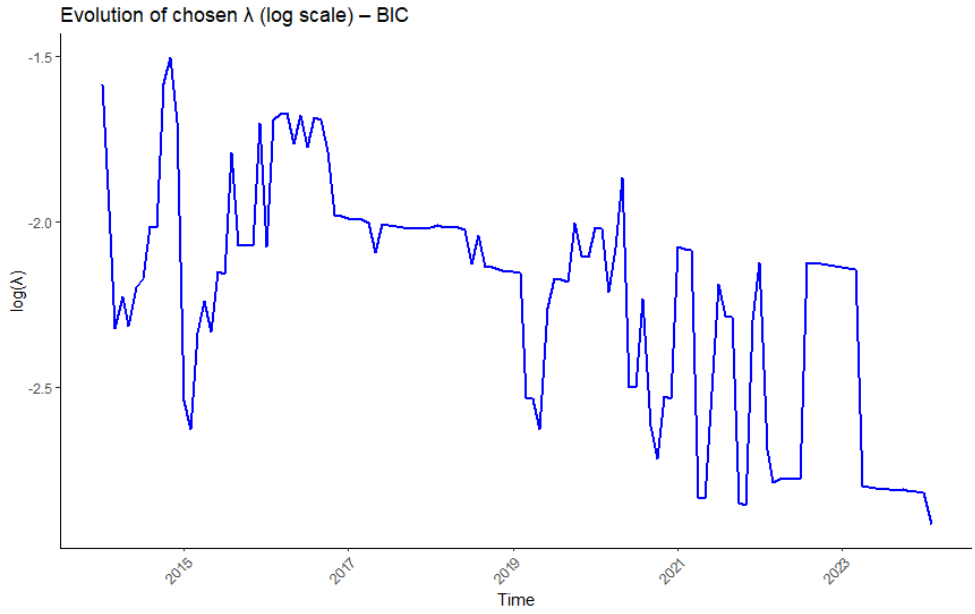


Figure C.7: LASSO (BIC): λ evolution.

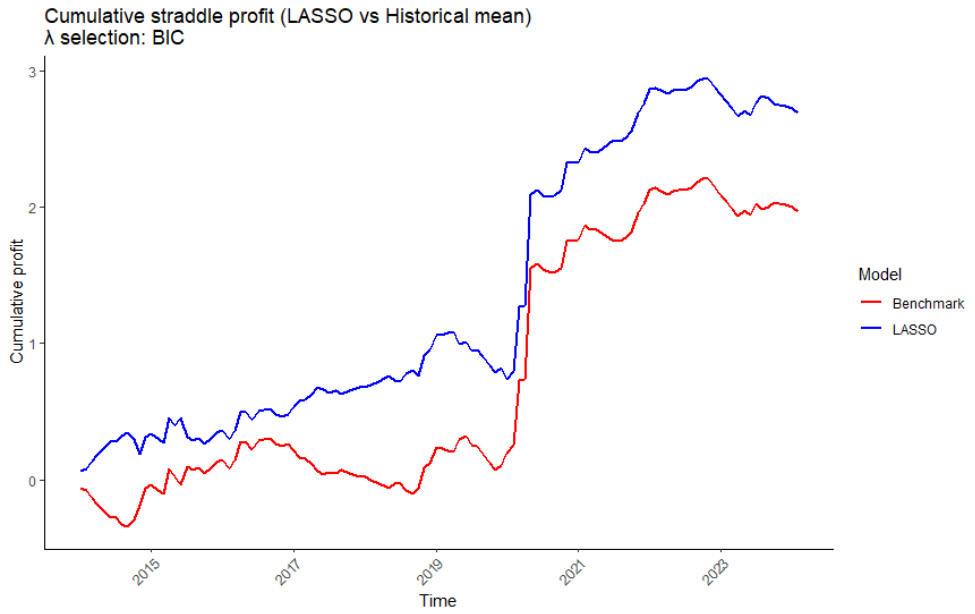


Figure C.8: LASSO (BIC): straddle profit.

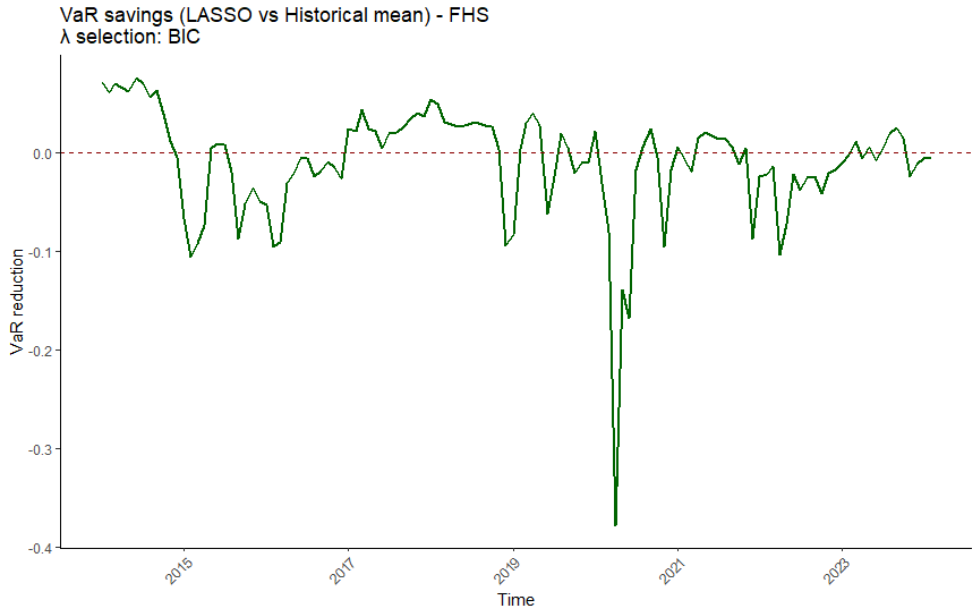


Figure C.9: LASSO (BIC): VaR savings.

C.2 Elastic Net

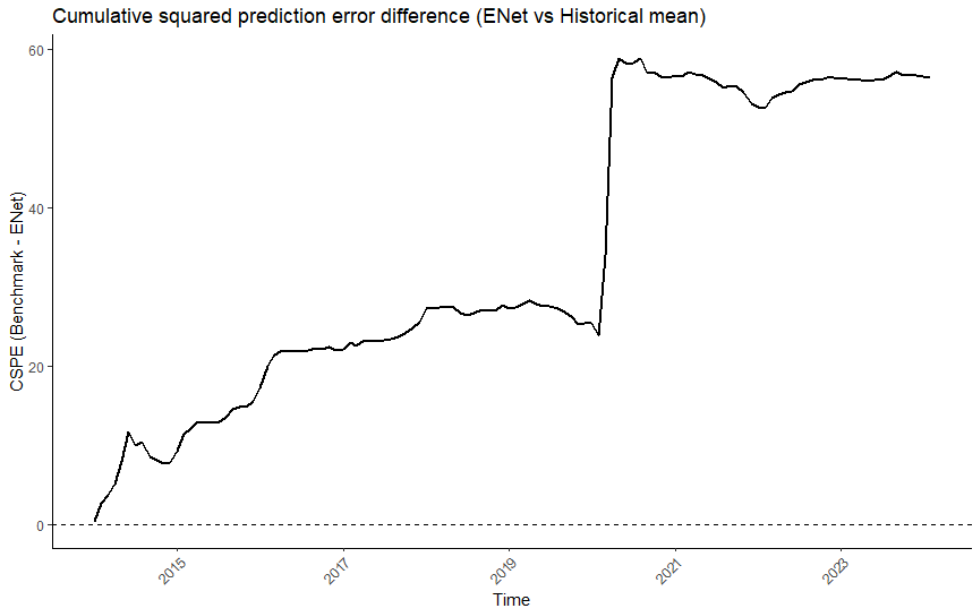


Figure C.10: Elastic Net (standard): CSPE.

Variable selection heatmaps

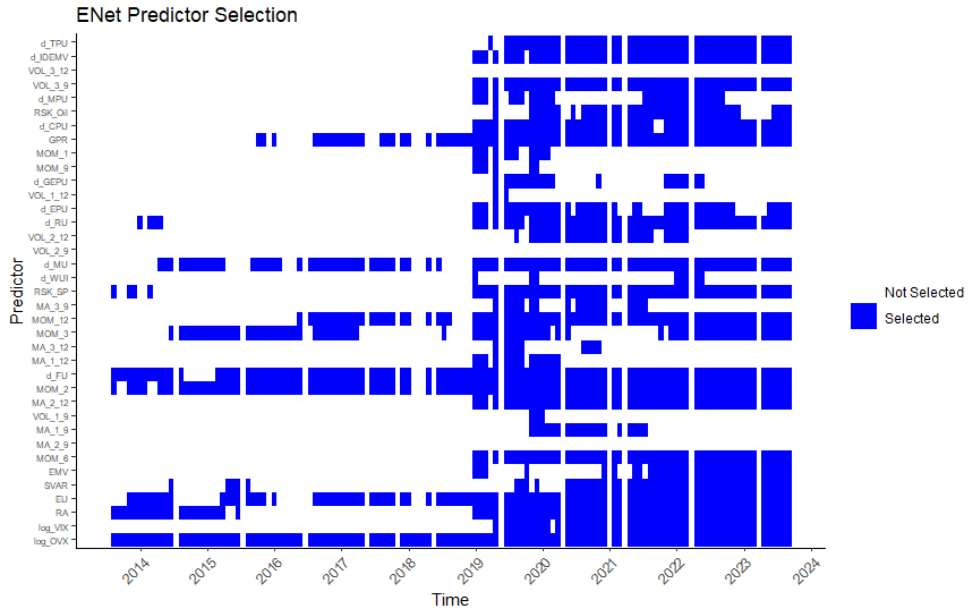


Figure C.11: Elastic Net (standard): heatmap.

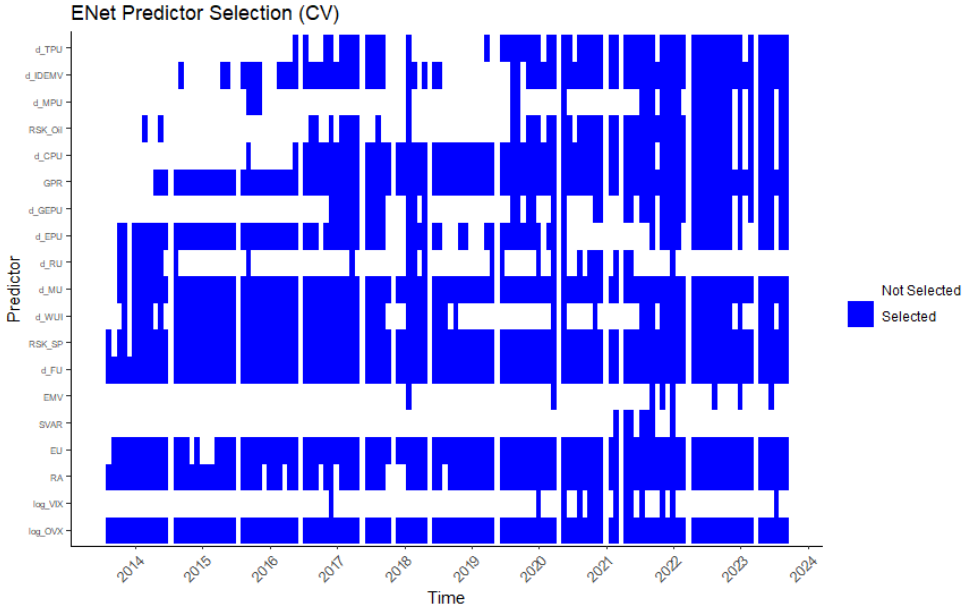


Figure C.12: Elastic Net (CV): heatmap.

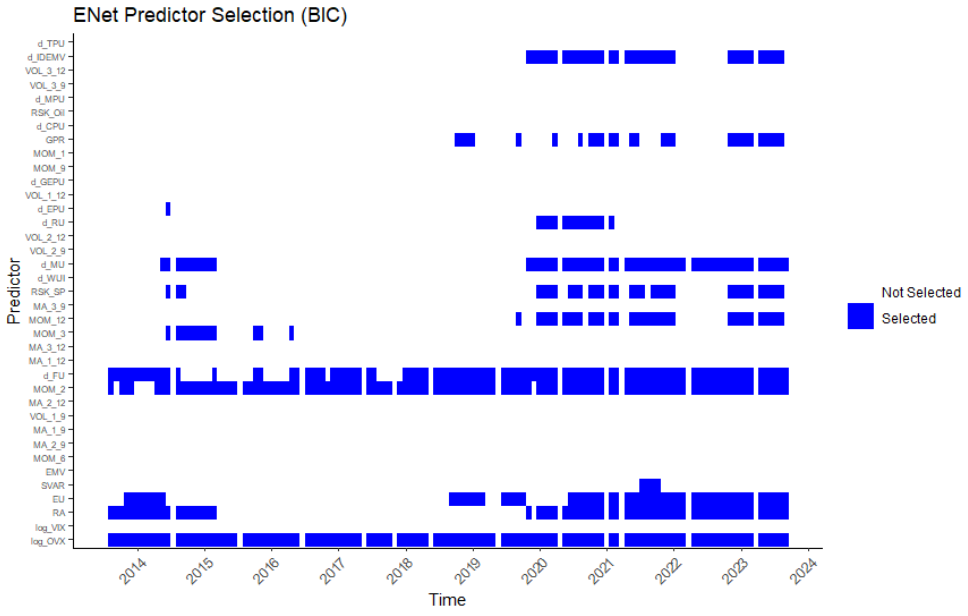


Figure C.13: Elastic Net (BIC): heatmap.

λ evolution

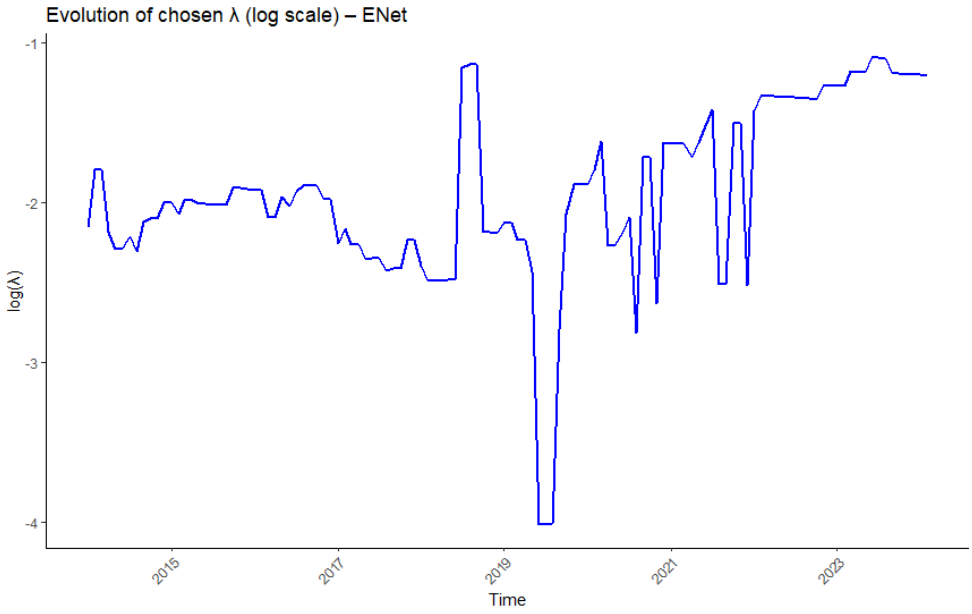


Figure C.14: Elastic Net (standard): λ evolution.

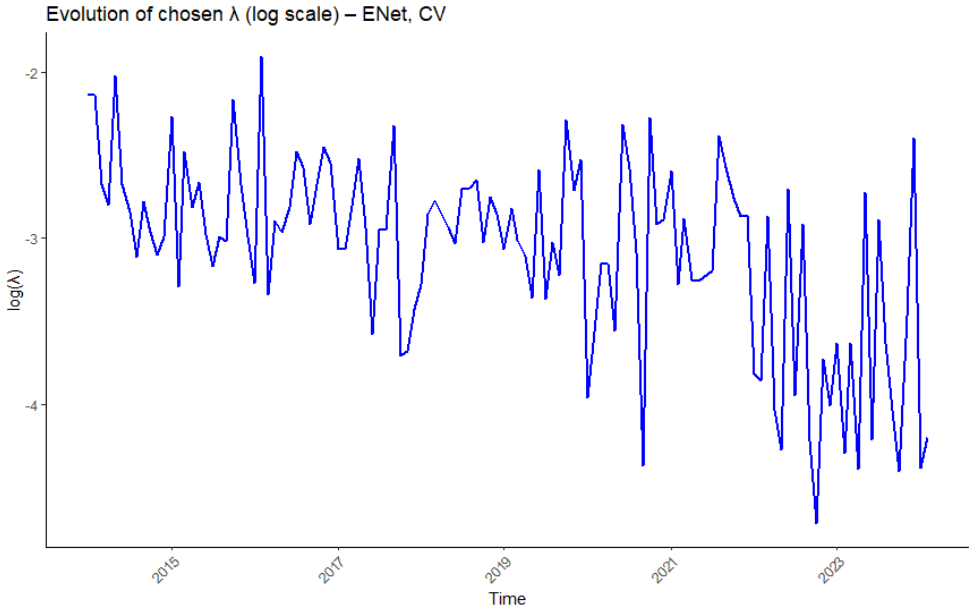


Figure C.15: Elastic Net (CV): λ evolution.

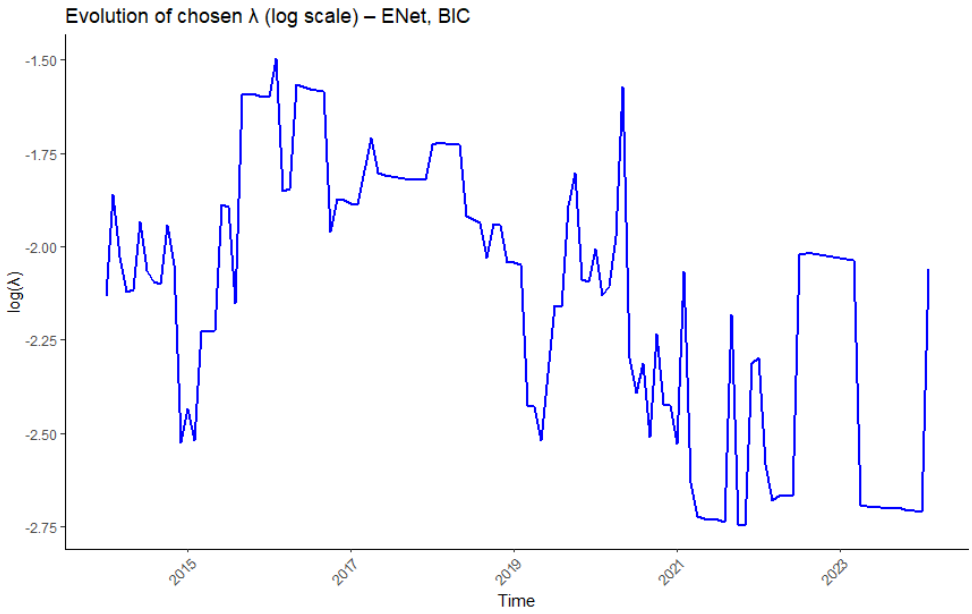


Figure C.16: Elastic Net (BIC): λ evolution.

α evolution

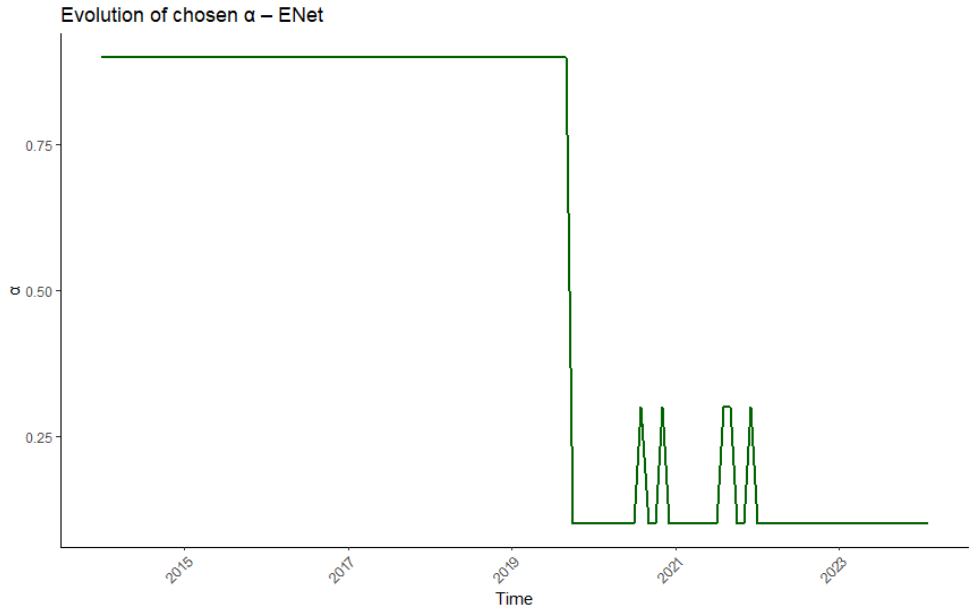


Figure C.17: Elastic Net (standard): α evolution.

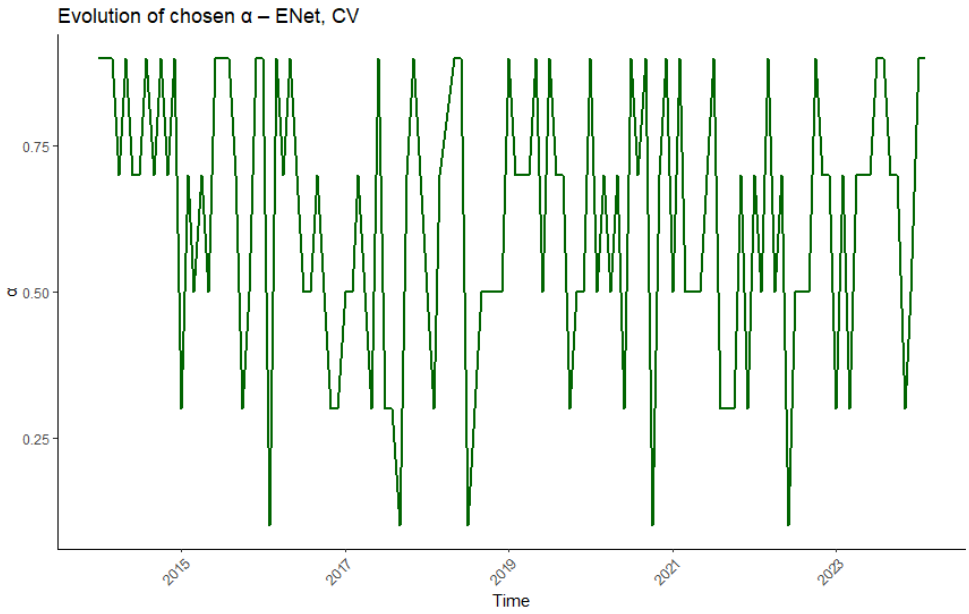


Figure C.18: Elastic Net (CV): α evolution.

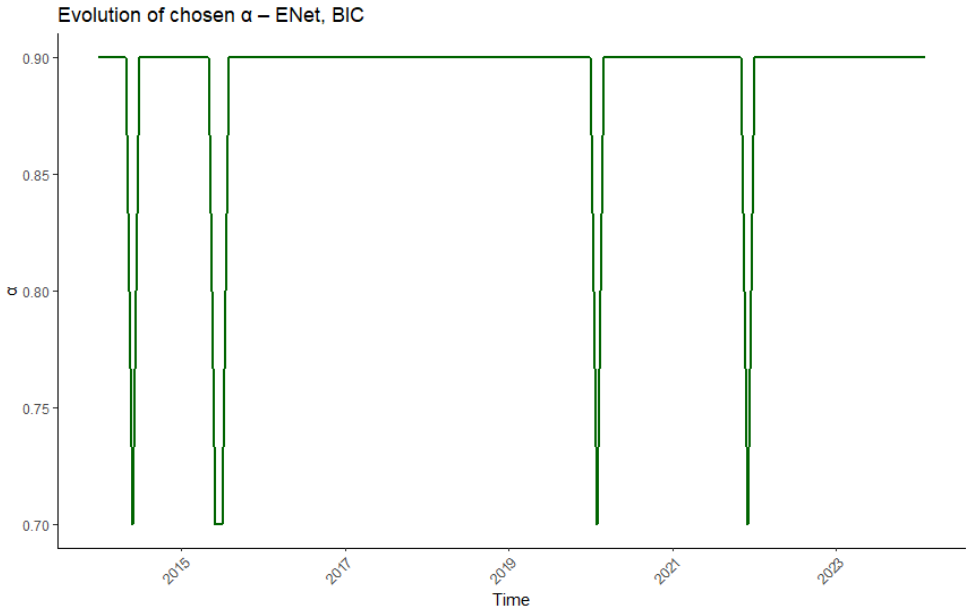


Figure C.19: Elastic Net (BIC): α evolution.

Straddle cumulative profit

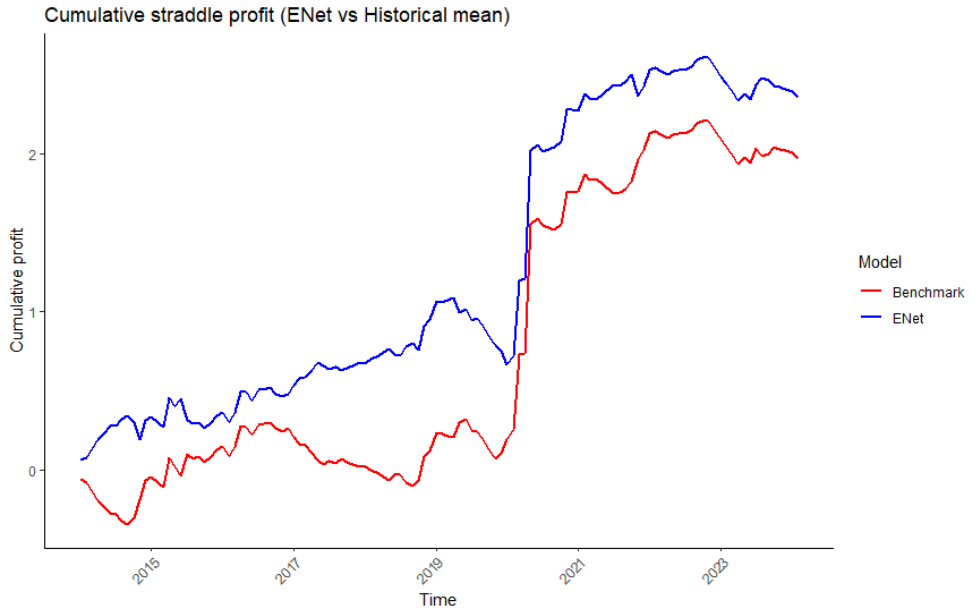


Figure C.20: Elastic Net (standard): straddle profit.

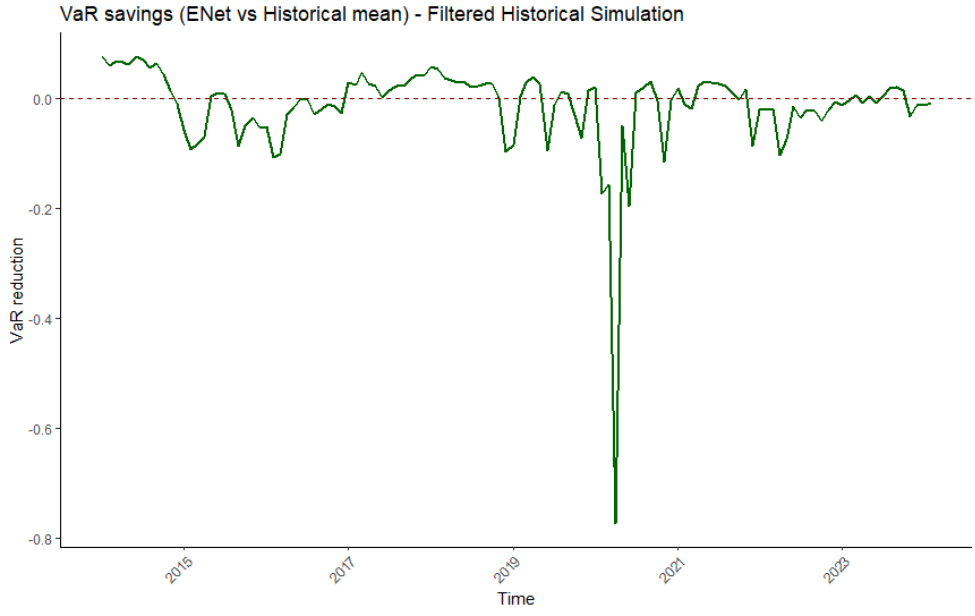


Figure C.21: Elastic Net (standard): VaR savings.

C.3 OCMT

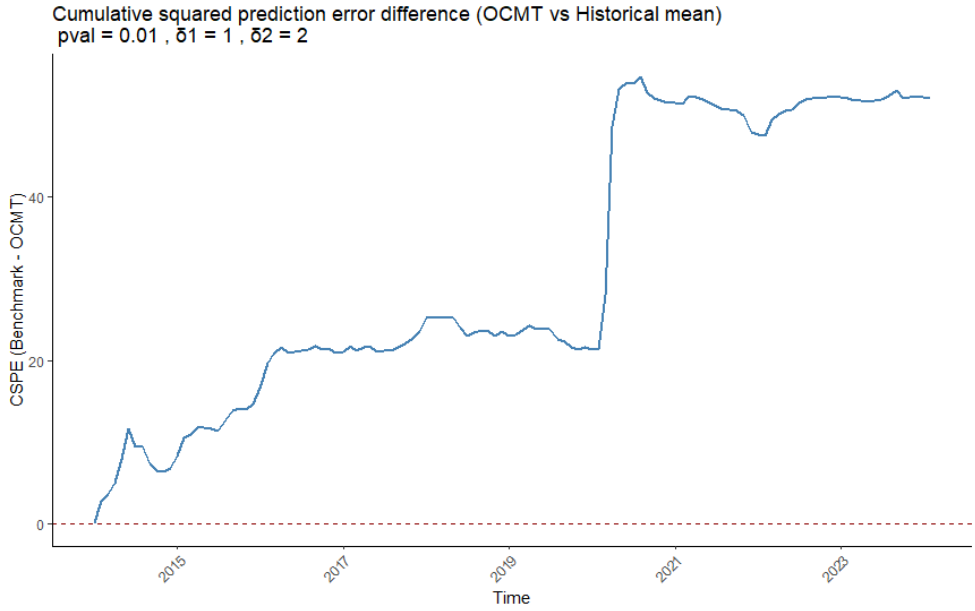


Figure C.22: OCMT: CSPE.

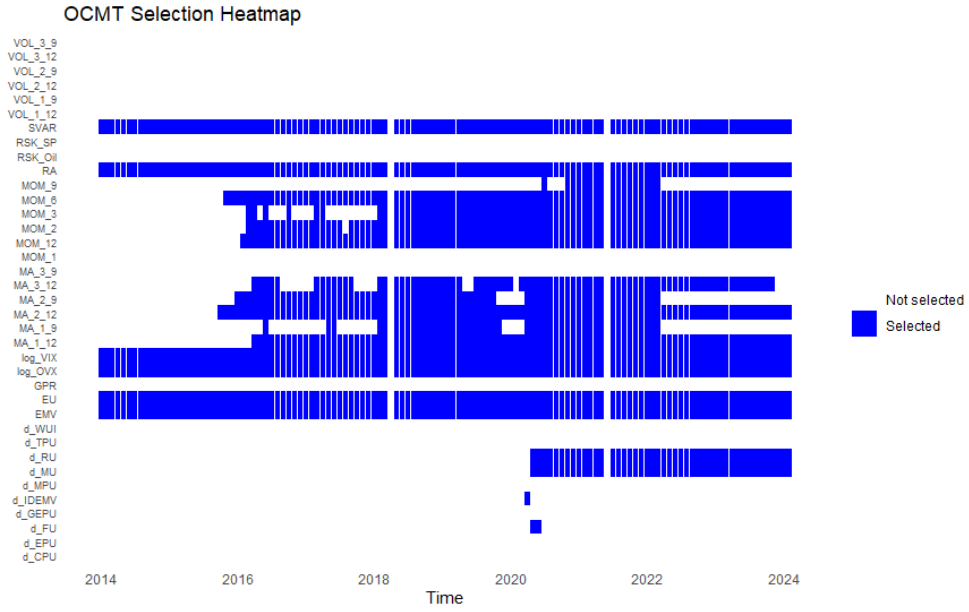


Figure C.23: OCMT: variable selection heatmap.

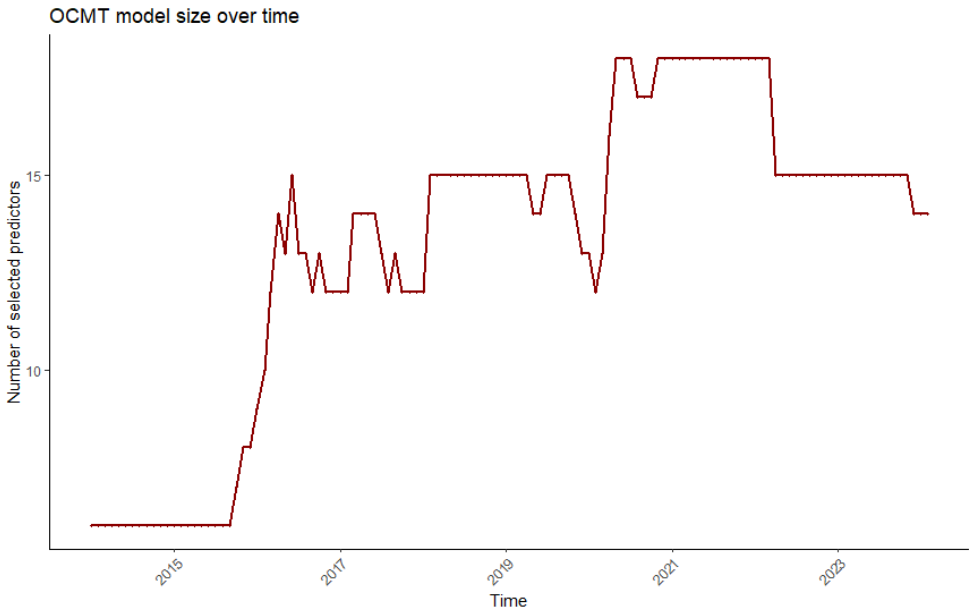


Figure C.24: OCMT: model size evolution.

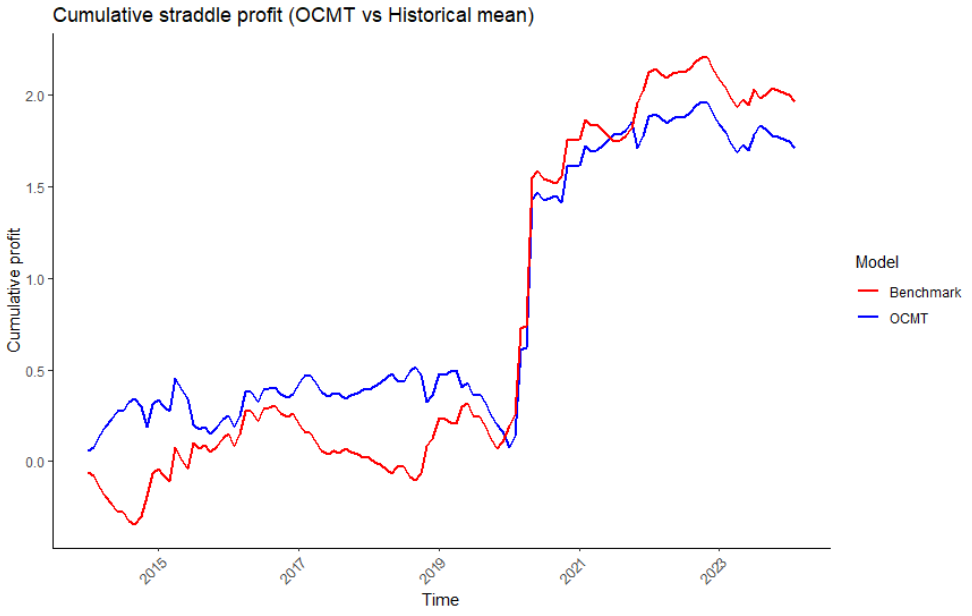


Figure C.25: OCMT: straddle profit.

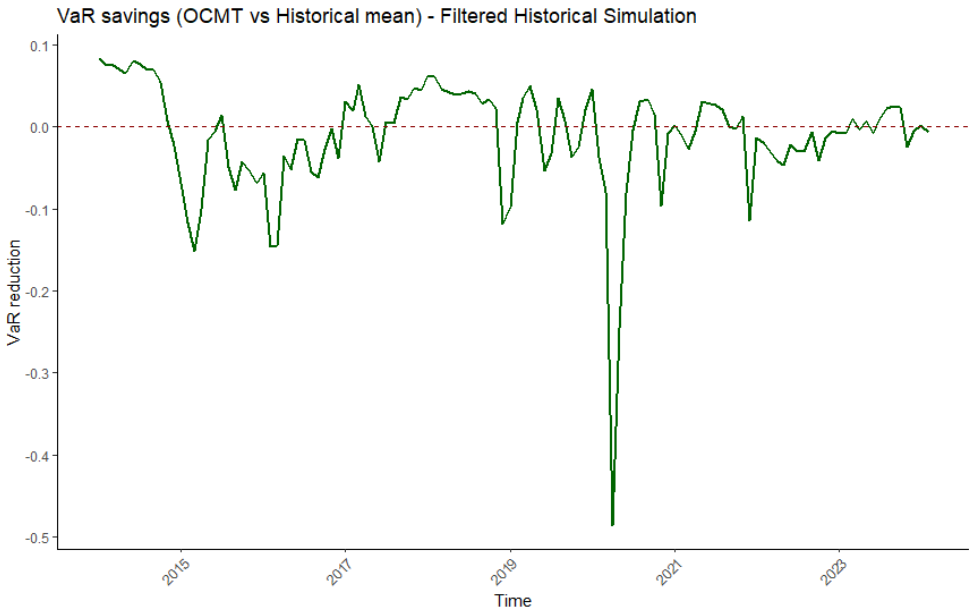


Figure C.26: OCMT: VaR savings.

C.4 BMT

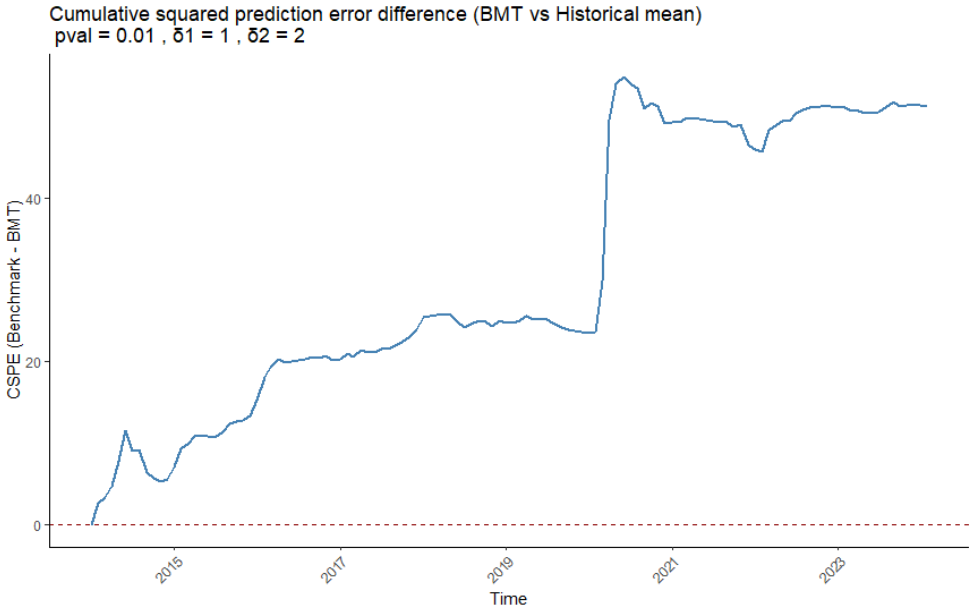


Figure C.27: BMT: CSPE.

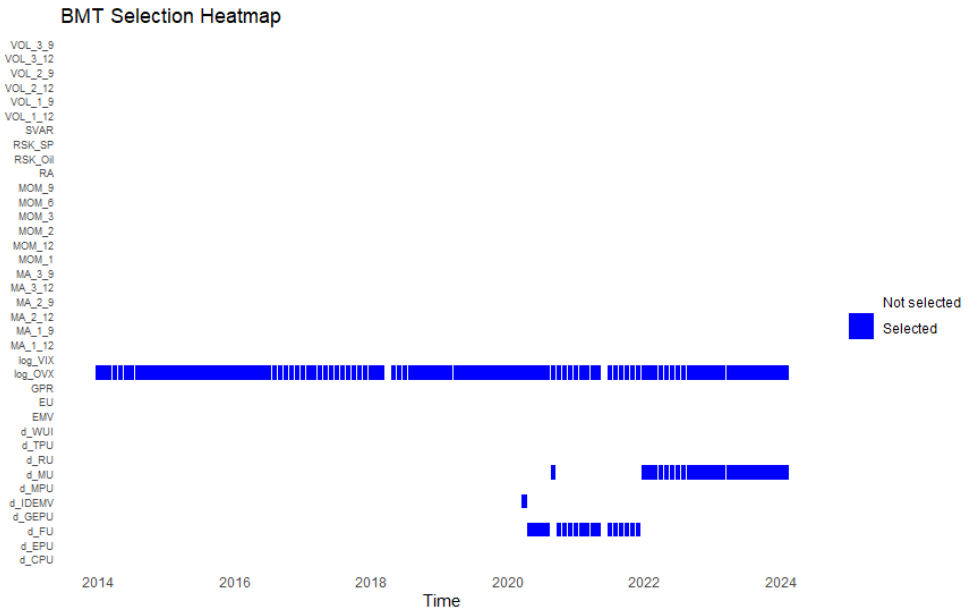


Figure C.28: BMT: variable selection heatmap.

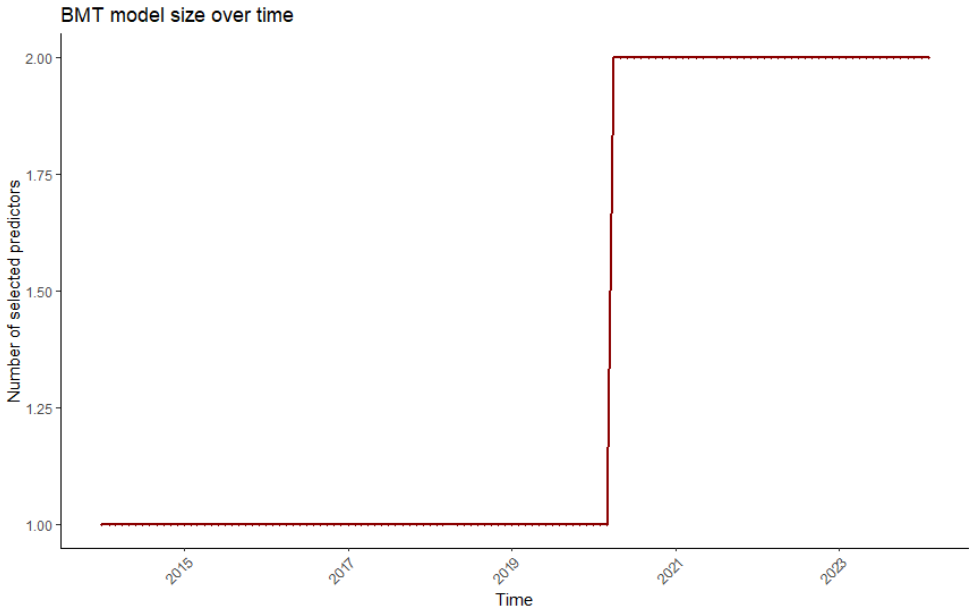


Figure C.29: BMT: model size evolution.

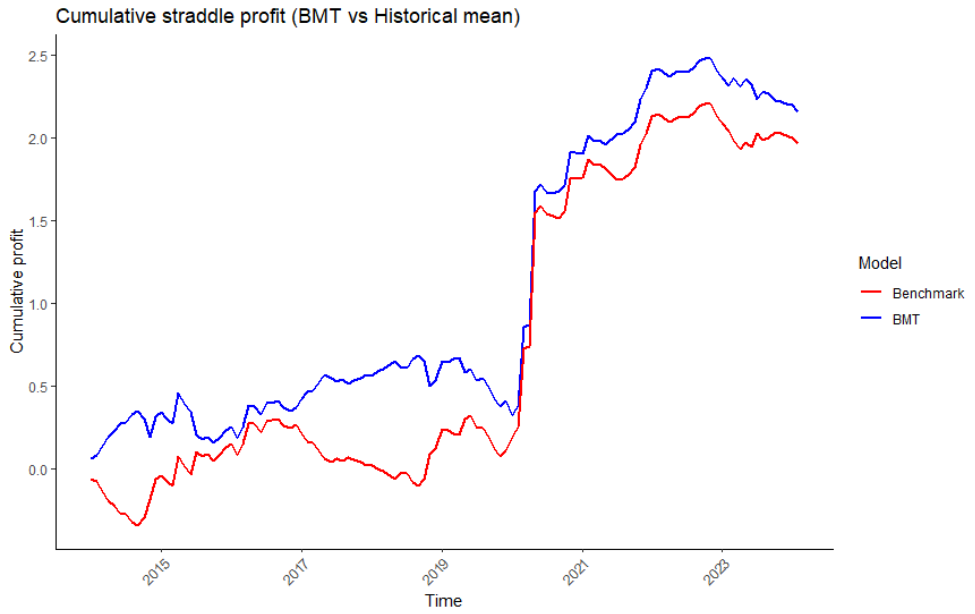


Figure C.30: BMT: straddle profit.

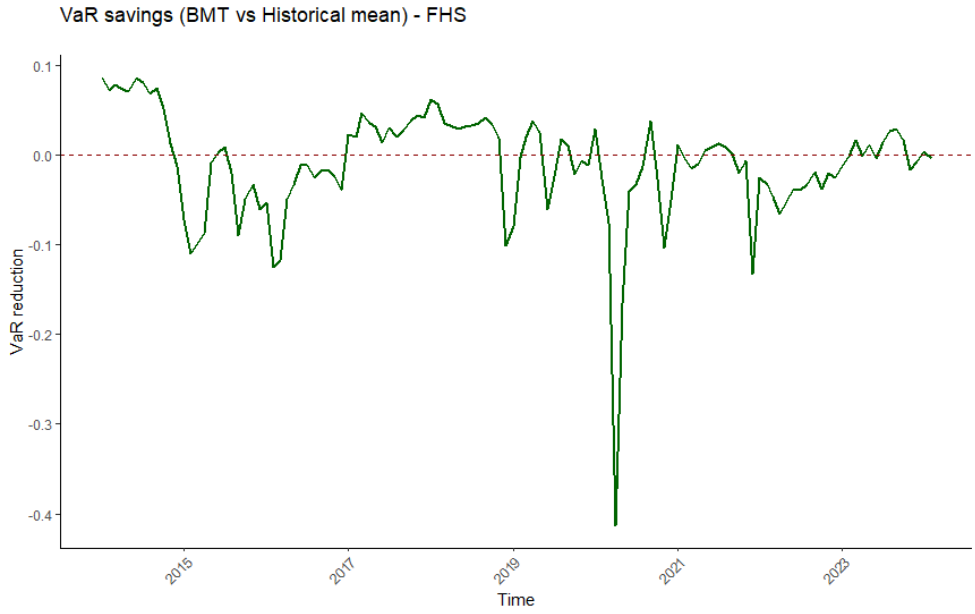


Figure C.31: BMT: VaR savings.

C.5 Diagnostics

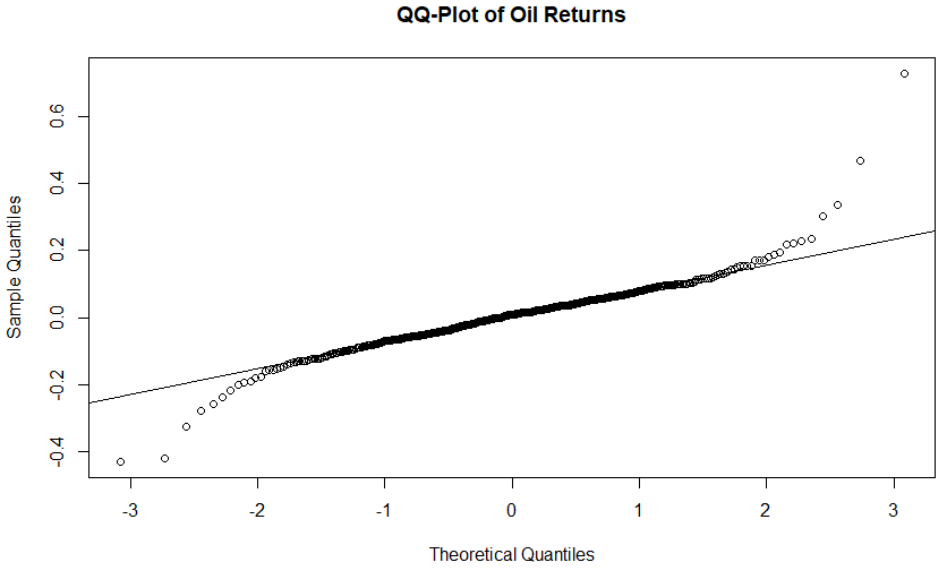


Figure C.32: QQ-plot of oil returns.

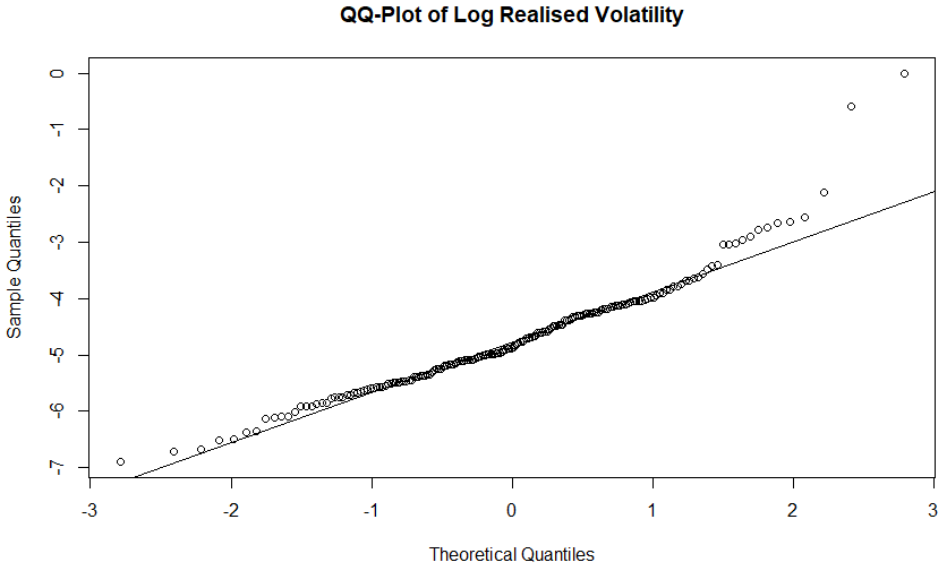


Figure C.33: QQ-plot of log realised volatility.

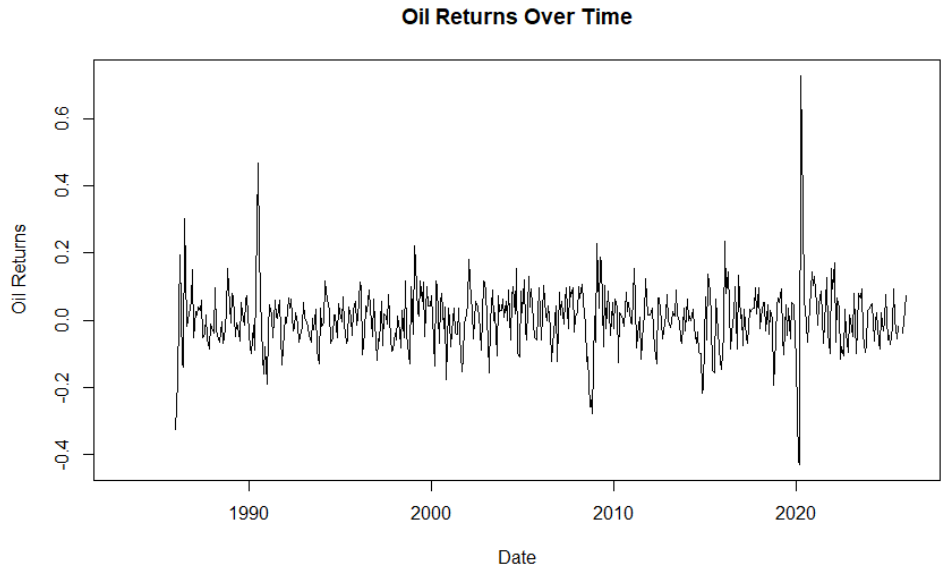


Figure C.34: Oil returns over time.

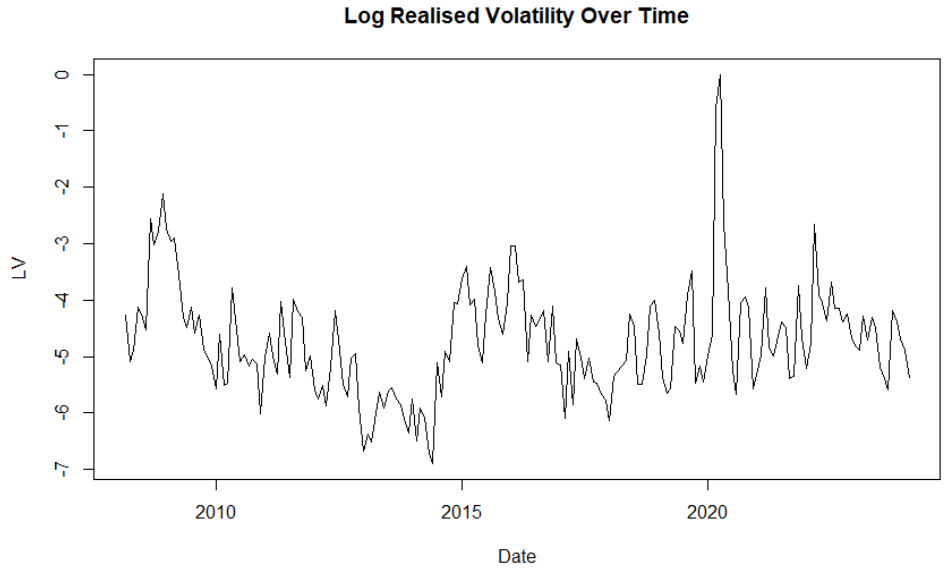


Figure C.35: Log realised volatility over time.

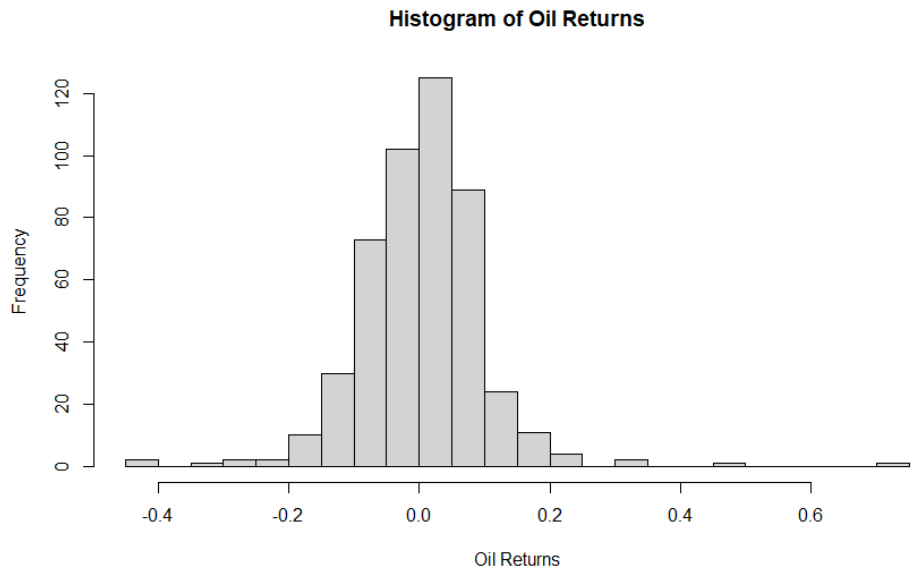


Figure C.36: Histogram of oil returns.

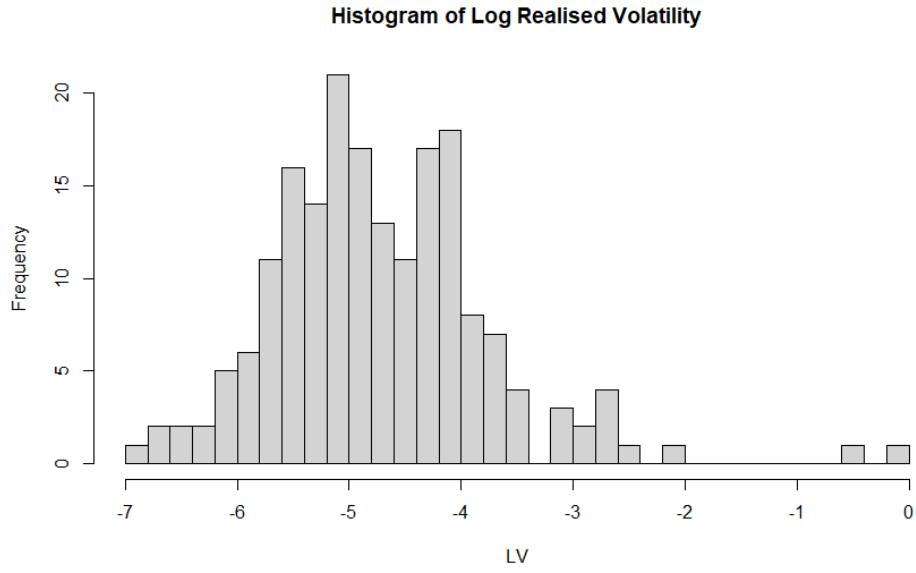


Figure C.37: Histogram of log realised volatility.