Imperial College
Business School

Imperial means
Intelligent Business

Equity Premium Prediction with Machine Learning Models

Prof. Tarun Ramadorai

Introduction

- In this study we use the Goyal and Welch (2008) data and framework to estimate the OOS performance of several machine learning models - LASSO, Decision Tree and Random Forest, and compare our results with Goyal and Welch (2008).
- We show that LASSO improves the OOS performance of 7 out of the 14 single variable models in Goyal and Welch (2008).
- Both LASSO and Forest also improve the OOS performance of the multiple variable model which includes all 14 predictors together in Goyal and Welch (2008).
- We show that forecasts of the equity premium can be further improved by: using a moving window instead of an expanding window; running the Random Forest on the features selected by the LASSO; and imposing theoretically motivated positivity constraints on forecasts of the equity premium.

Data

- We use the monthly data, which is updated to December 2015, provided on Amit Goyal's website. The sample period is December 1927 to December 2015.
- The equity premium is calculated as the simple return (including dividends) on the S&P 500 index minus the prevailing Treasury-bill rate. The equity premium has an annualized mean of 7.59% and an annualized standard deviation of 0.19.
- We use 14 predictors from Goyal and Welch (2008), and provide a summary of the monthly forecasts on the next slide. Note that **all** is a kitchen sink model which includes all 14 predictors.

Table 1. Forecasts at monthly frequency

This table replicates Table 3 in Goyal and Welch (2008) with updated data till December 2015. This table presents statistics on forecast errors in-sample (IS) and out-of-sample (OOS) for equity premium forecasts at the monthly frequency. The sample period is from December 1927 to December 2015. All columns are based on predicting simple returns (adjusted with dividends) for correspondence with Goyal and Welch (2008) and Campbell and Thompson (2008). "T" means "truncated" to avoid a negative equity premium prediction. "U" means "unconditional" to avoid a forecast that is based on a coefficient that is inverse to what the theory predicts.

		IS	OOS		Campbell and Thompson (2005) OOS	
	Variable	$\bar{R}^2$	$\bar{R}^2$	ΔRMSE	$\bar{R}^2$	ΔRMSE
					TU	TU
de	Dividend payout ratio	-0.12	-0.65	-0.01	-0.21	-0.0018
svar	Stock variance	-0.12	-0.91	-0.02	-0.89	-0.0161
dfr	Default return spread	0.00	-0.38	-0.01	-0.41	-0.0060
lty	Long term yield	0.05	-0.51	-0.01	0.36	0.0102
ltr	Long term return	0.10	-0.54	-0.01	0.10	0.0047
infl	Inflation	0.12	0.28	0.01	0.32	0.0092
tms	Term spread	0.11	0.08	0.00	0.08	0.0042
tbl	Treasury-bill rate	0.20	0.05	0.00	0.30	0.0089
dfy	Default yield spread	0.21	-0.60	-0.01	-0.56	-0.0091
dp	Dividend price ratio	0.30	-0.35	0.00	0.11	0.0050
ep	Earning price ratio	0.30	-1.37	-0.03	-0.62	-0.0105
bm	Book to market	0.69	-2.79	-0.06	-1.95	-0.0380
ntis	Net equity expansion	0.37	-1.06	-0.02	-1.06	0.0005
dy	Dividend yield	0.42	-1.02	-0.02	-0.10	48.9296

Modelling Approach
LASSO

- We fit the following model with LASSO in each window :

$$EQ_{t+1} = \beta_0 + \beta X_t + \epsilon_{t+1}$$

where EQ is the equity premium, X are the predictors used in Goyal and Welch (2008), and β is a K*1 vector of coefficients. Both single variable models and multiple variable models are estimated.

- The LASSO estimator is the solution to the following penalized likelihood problem:

$$argmin \sum_{i=1}^{N} \left(Y_i - \beta_0 - \sum_{k=1}^{K} X_{ik}\beta_k \right)^2 + \lambda \sum_{k=1}^{K} |\beta_k|$$
$$= RSS + \lambda \sum_{k=1}^{K} |\beta_k|$$

Modelling Approach
LASSO

- LASSO performs coefficient shrinkage and variable selection simultaneously. The λ parameter controls for the degree of sparsity of the coefficients, shrinking some coefficients, and setting others to zero.
- When λ is close to zero, the constraint has no effect and the LASSO estimators will be close to OLS.
- We use 10-fold cross-validation to select the best λ from 30 values along a log space from 10^{-5} to 10^{30} .
- Each estimation window is split into 10 smaller sets. A model is trained using 9 folds as training data, and the 10th fold is used as a test set to select the best λ, and this process is done 10 times, each with a different fold. The λ used for the model is the average of the best λs selected by cross validation.

Modelling Approach
Decision Tree

- The decision tree is a non-parametric supervised learning method used for classification and regression. We use regression trees to create models that predict the value of equity premium by learning decision rules from the features of the data.

- The key parameters in the model are set as follows:

  - **Maximum features** is the maximum number of features allowed to try when looking for the best split. We let the model automatically select all the features.
  - **Maximum depth** of the tree is set such that nodes are expanded until all leaves contain less than 2 observations.
  - The **minimum number of observations** required to split **an internal node** is set to 2.
  - The **minimum number of observations** required to be at **a leaf node** is chosen by 10-fold cross-validation.
  - **Maximum leaf nodes.** We allow a potentially unlimited number of leaf nodes.

Modelling Approach
Random Forest

- The random forest is an ensemble model that consists of a large number of separately grown decision trees. The model averages predictions from bootstrapped trees to improve predictive accuracy. It usually provides better prediction results than tree regressions.

- In a random forest model, trees are fit on various sub-samples of the dataset. The sub-sample size is the same as the original input sample size and bootstrap samples are used in our modelling.

- The random forest predictor is

$$\hat{f}_{rf}^{B}(x) = \frac{1}{B}\sum_{B}^{b=1} T(x;\theta_b)$$

where B is the number of trees, θ_b characterises the b^{th} random forest tree in terms of split variables, cut-points at each node and leaf nodes.

Modelling Approach
Random Forest

- As with the approach in Decision Tree, we use 10-fold cross-validation to select the best minimum samples in each leaf in each regression.

- A key parameter in the forest regression is the number of trees to be built in the forest. A large number of trees gives a better prediction, but makes computing slower. We set this parameter to 500.

- The other parameters are set at the same values as in the Decision Tree.

Empirical Results

LASSO with single predictor

- We first compare the OOS performance of LASSO with OLS.
- In each model, we use only one predictor from the 14 predictors.
- For each model, both an expanding window and a 240-month rolling window are used for comparison. The first prediction starts in January 1948.
- We show the following results from the single predictor models:
  - LASSO improves the OOS performance of 7 out of the 14 predictors in Goyal and Welch (2008), i.e. **tbl**, **ntis**, **ep**, **bm**, **dfy**, **svar** and **tms**.

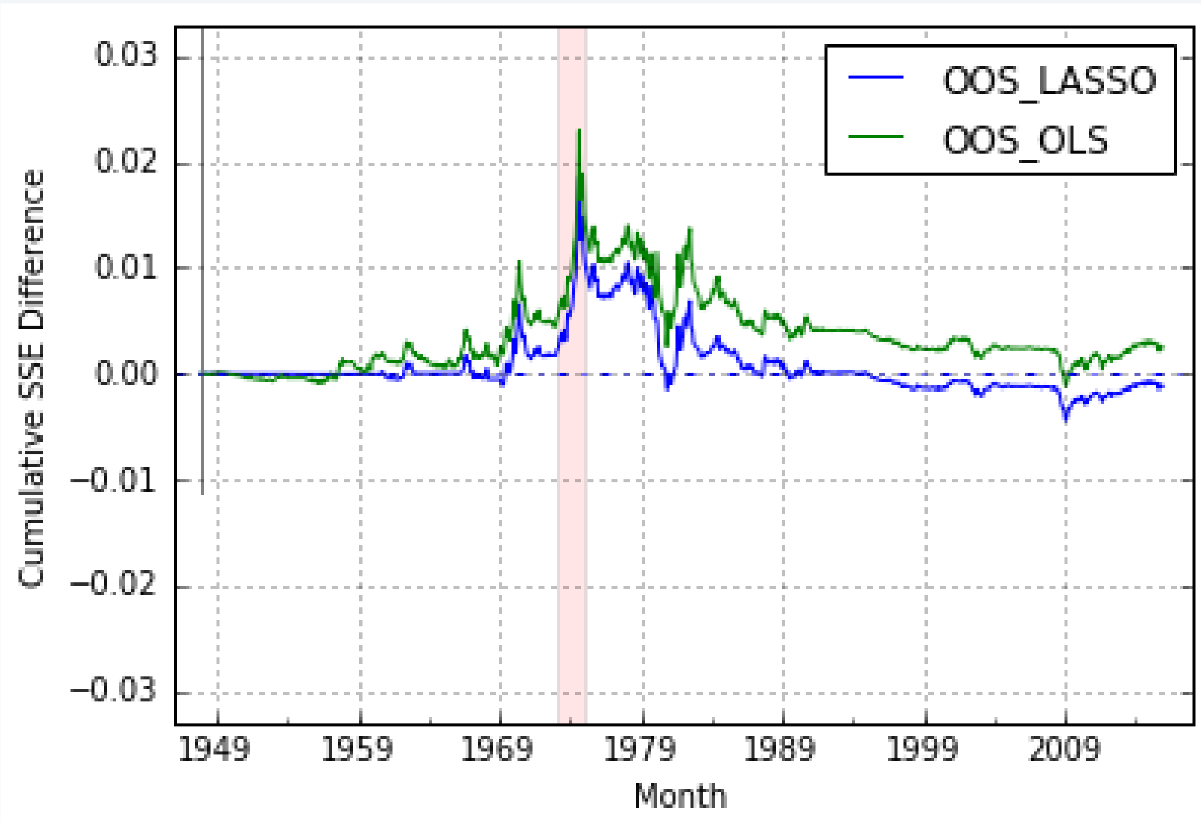
Empirical Results
LASSO with single predictor

- LASSO dynamically chooses λ based on historical information. When λ is large enough, LASSO generates close results to the historical mean model. When λ is close to zero, LASSO generates similar results to unconstrained OLS.

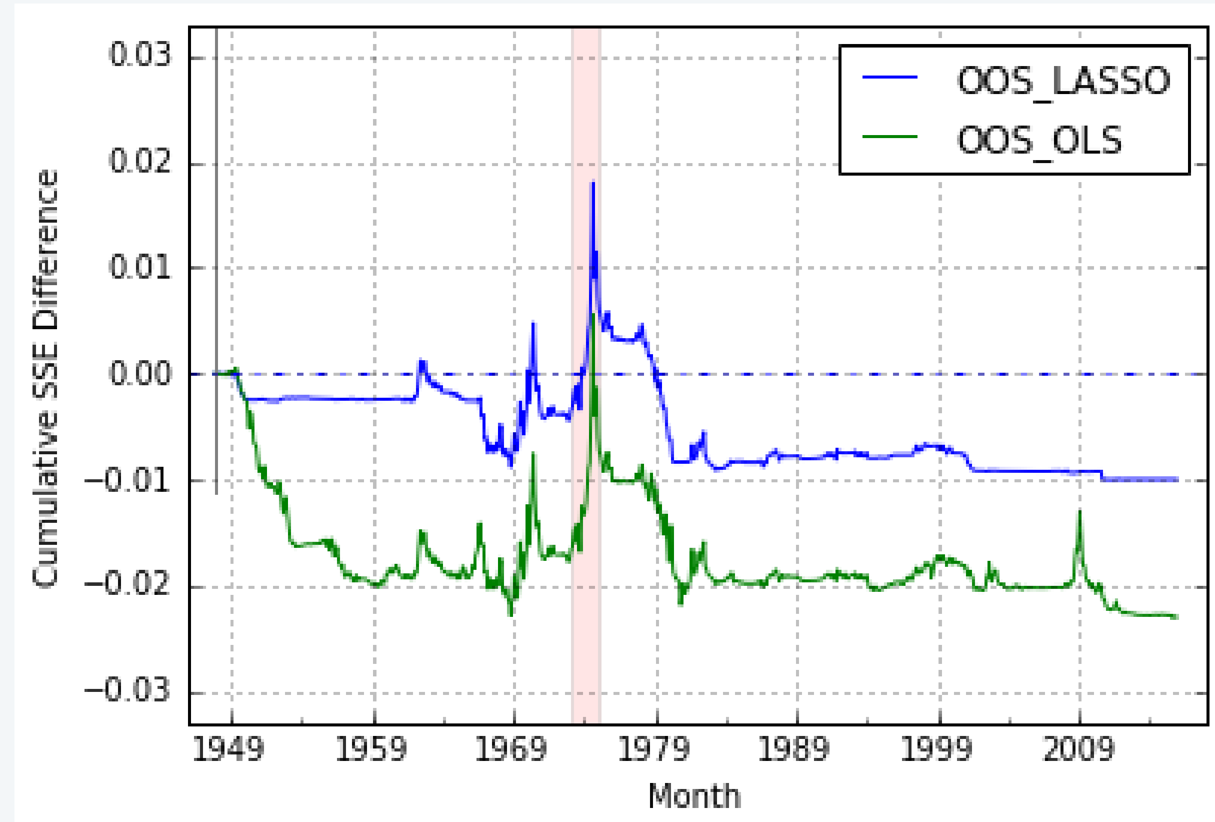
- LASSO prediction improves when a rolling window is used instead of an expanding window. The rolling window takes recent information into estimation, while the expanding window also accounts for historical information.

- LASSO performance can be further improved by imposing Campbell and Thompson (2008) constraints, and we show the results arrived at by imposing these restrictions in the Appendix.

Empirical Results
LASSO with single predictor



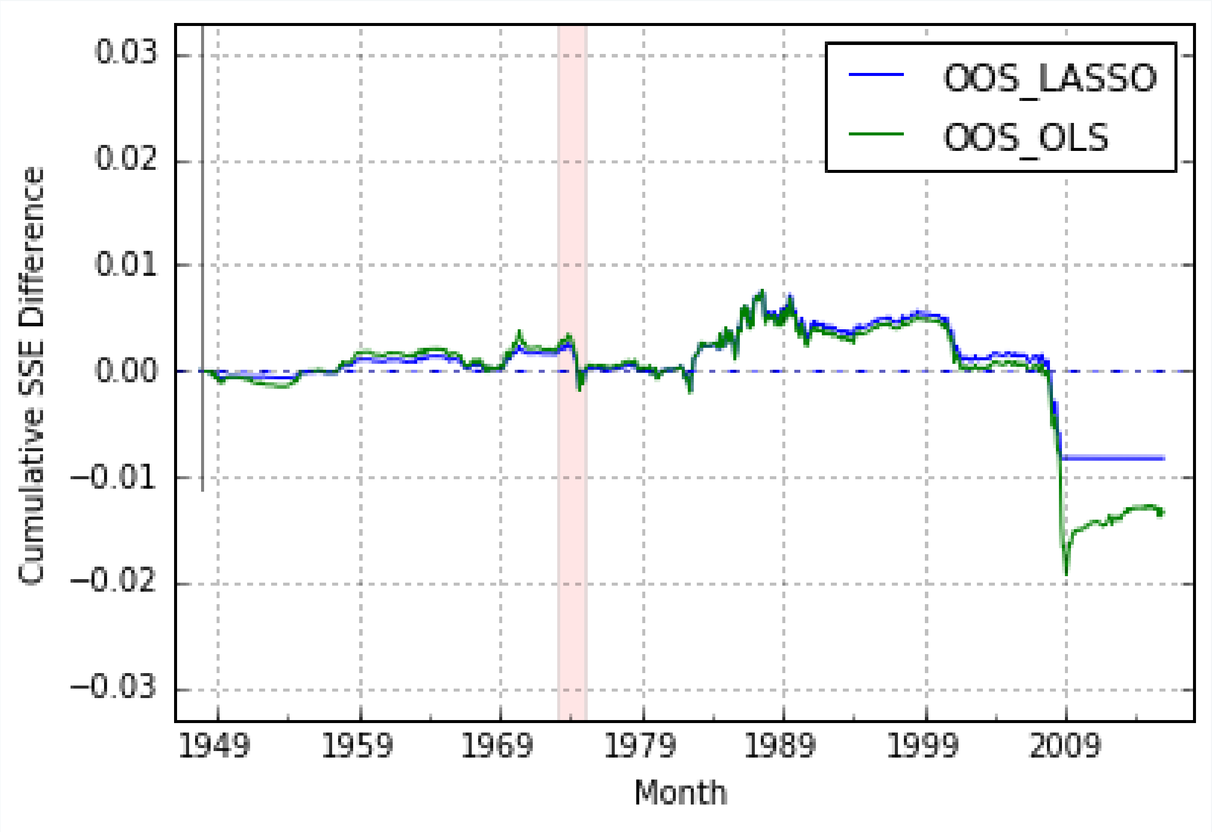
(a) expanding window



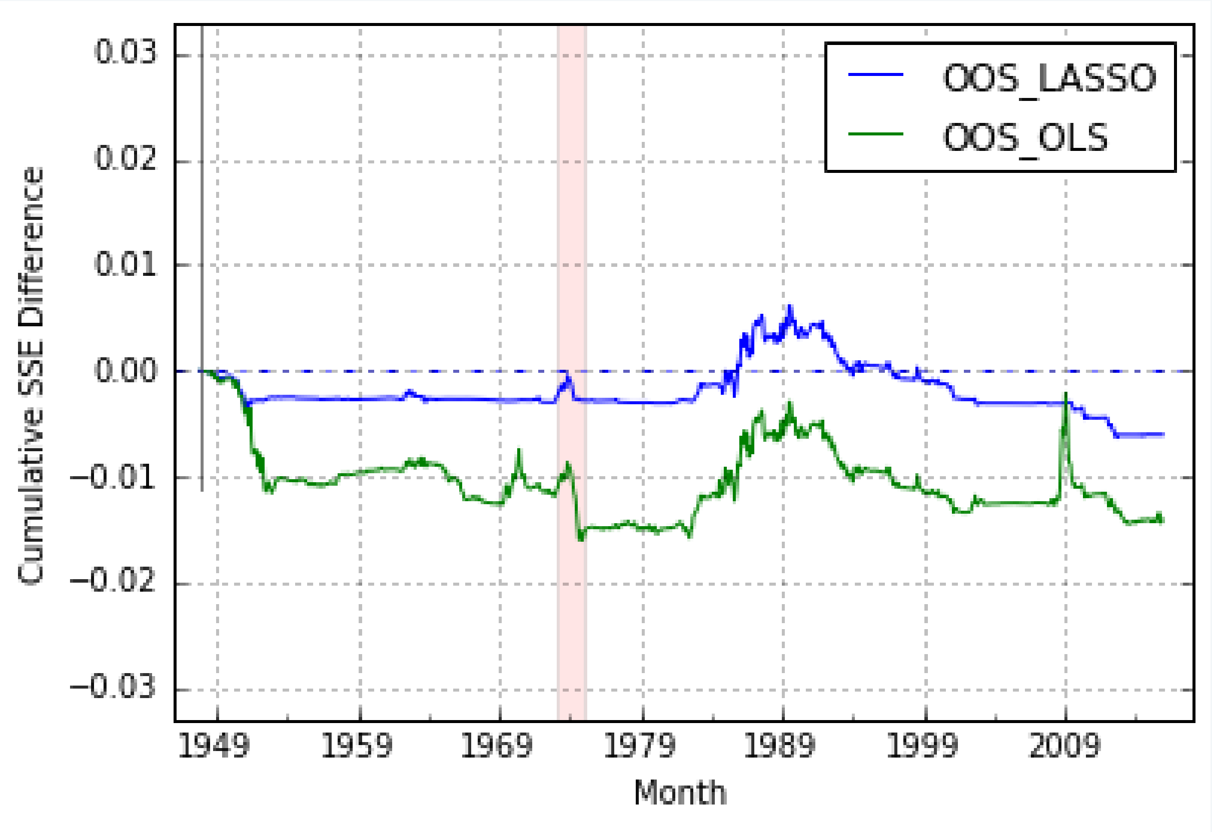
(b) rolling window

Figure: Monthly performance of tbl

Empirical Results
LASSO with single predictor



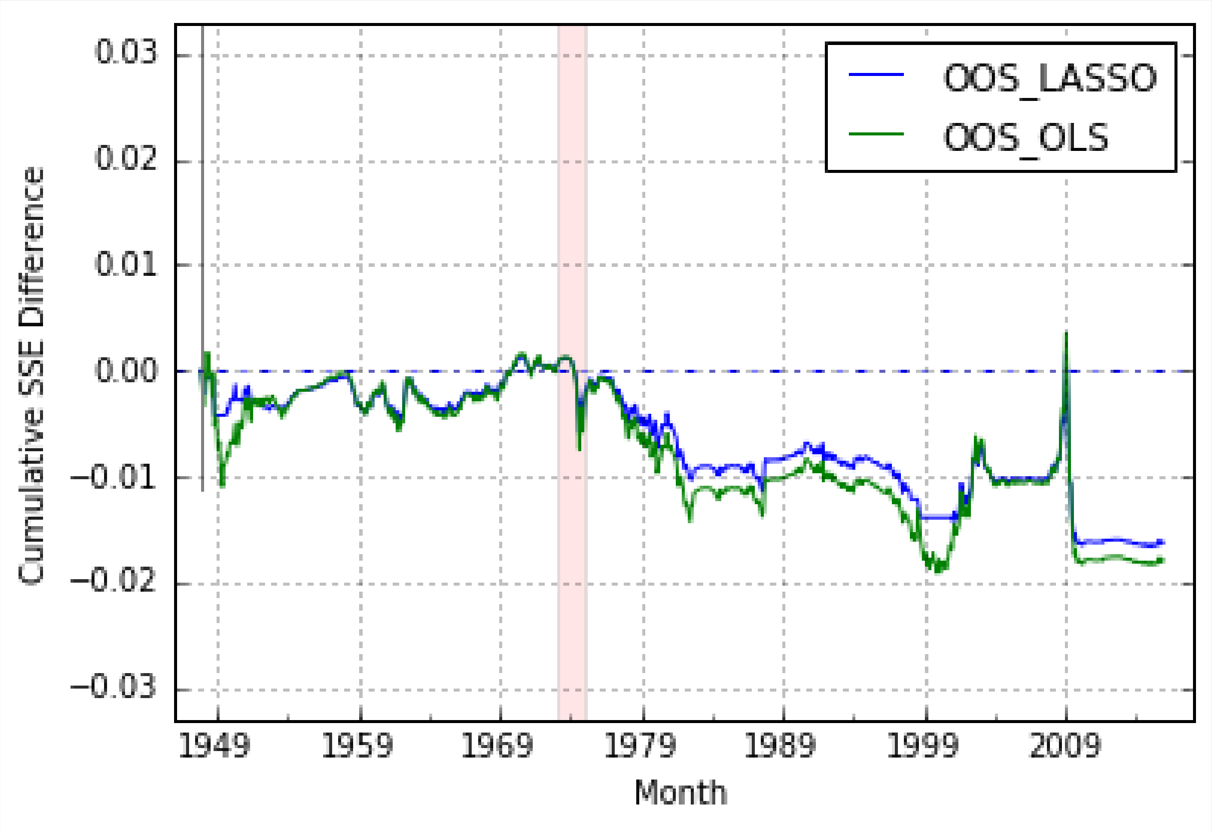
(a) expanding window



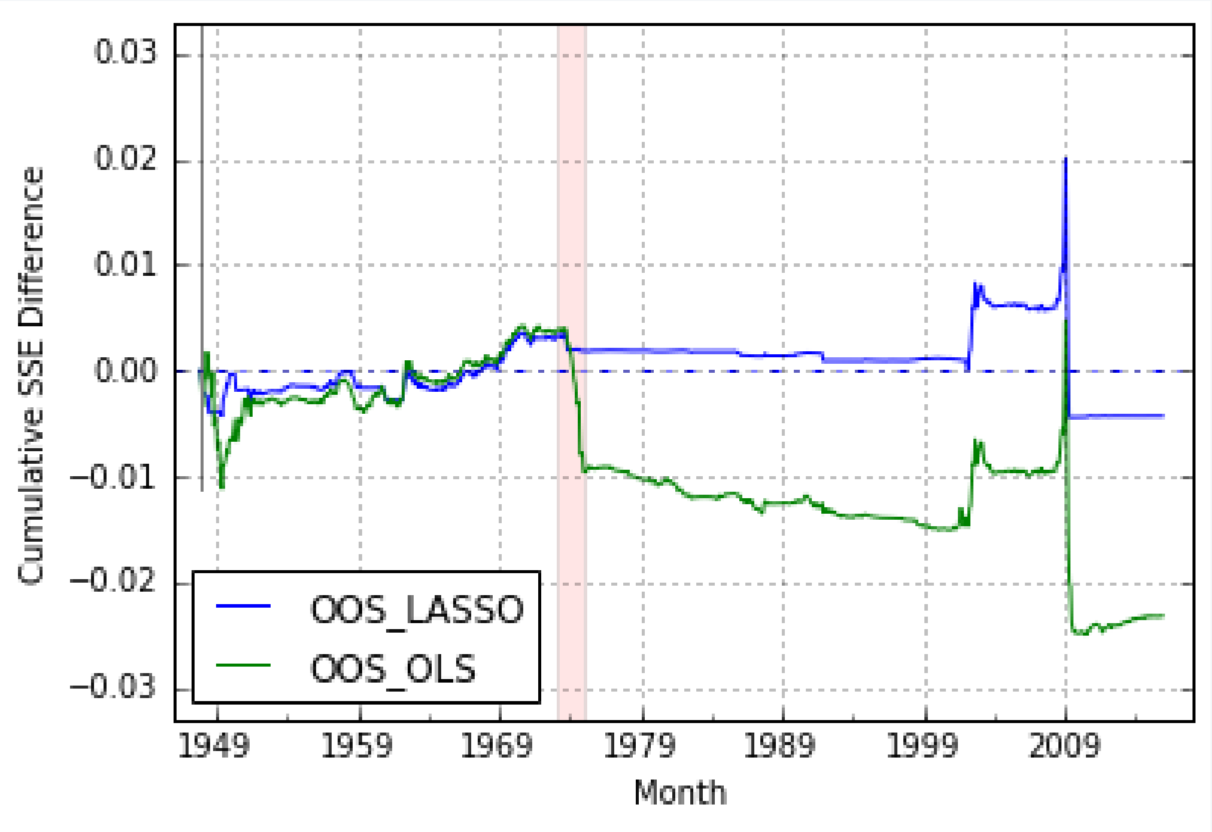
(b) rolling window

Figure: Monthly performance of ntis

Empirical Results
LASSO with single predictor



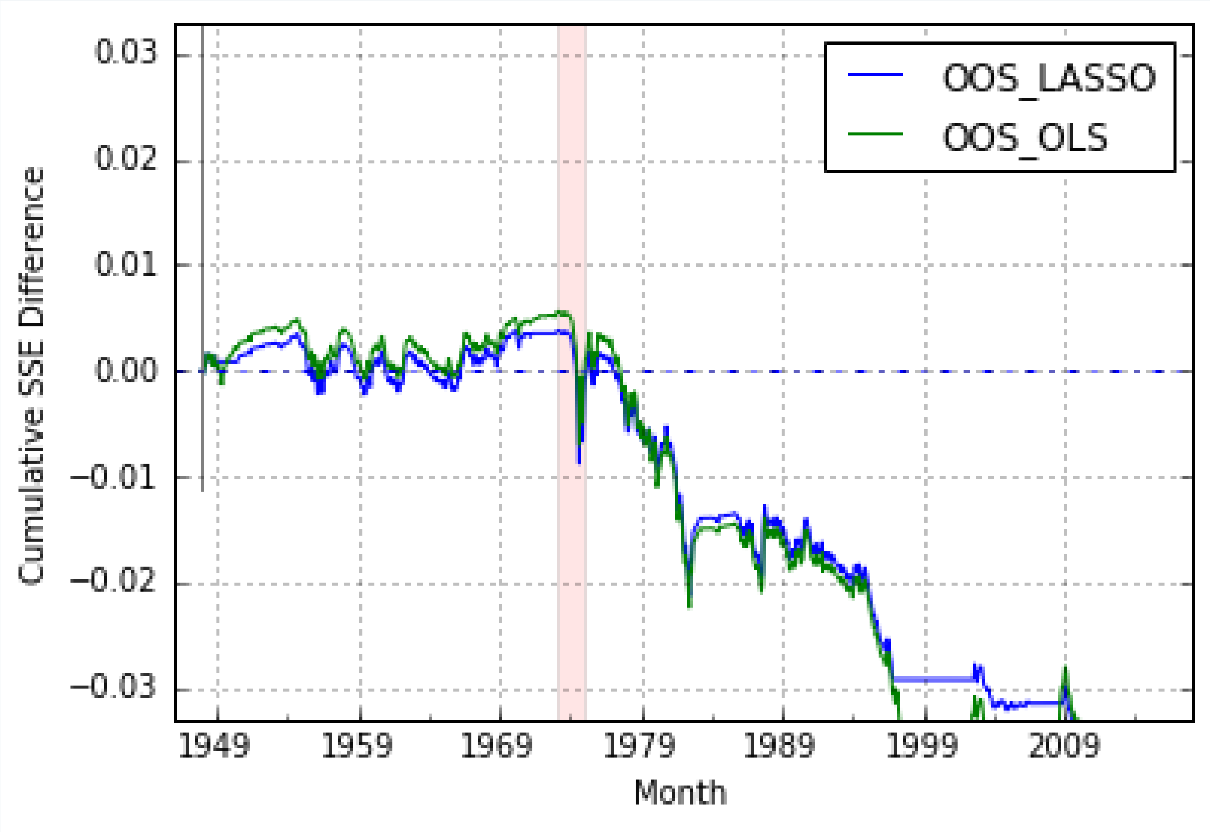
(a) expanding window



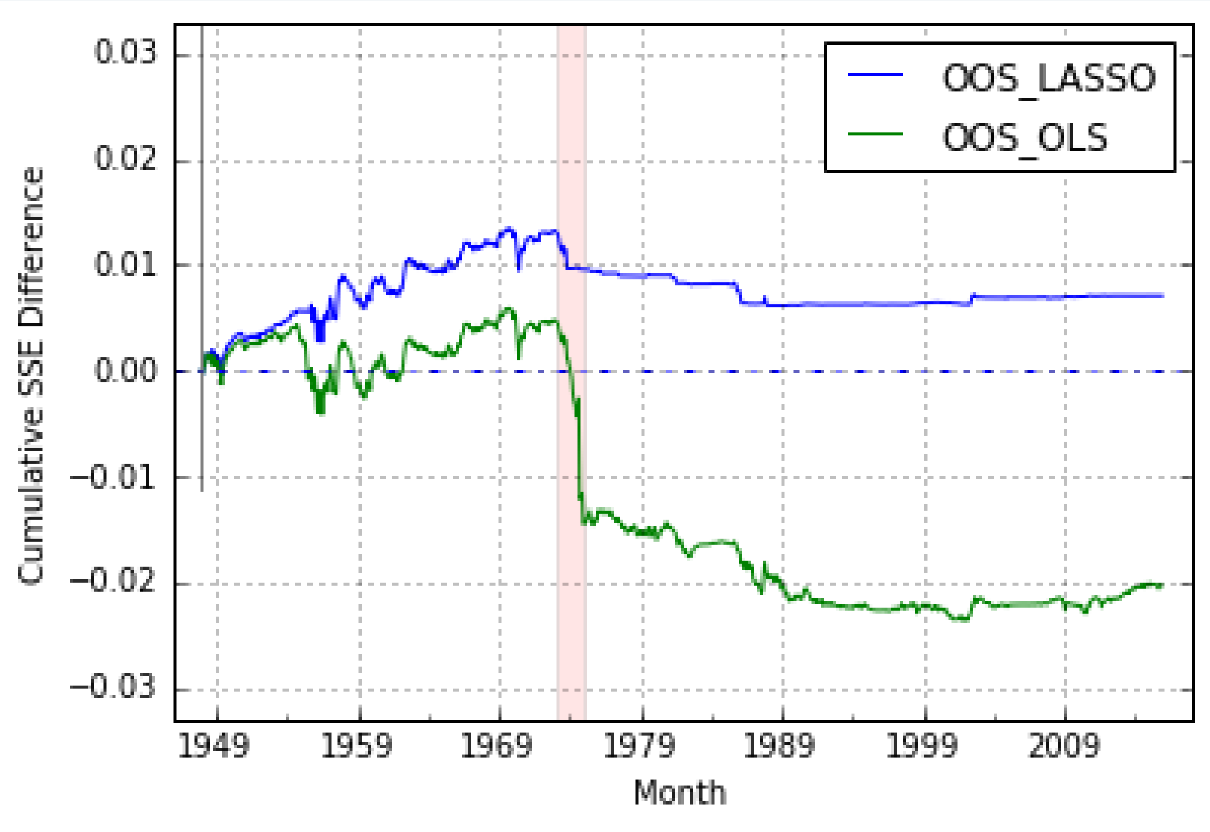
(b) rolling window

Figure: Monthly performance of ep

Empirical Results
LASSO with single predictor



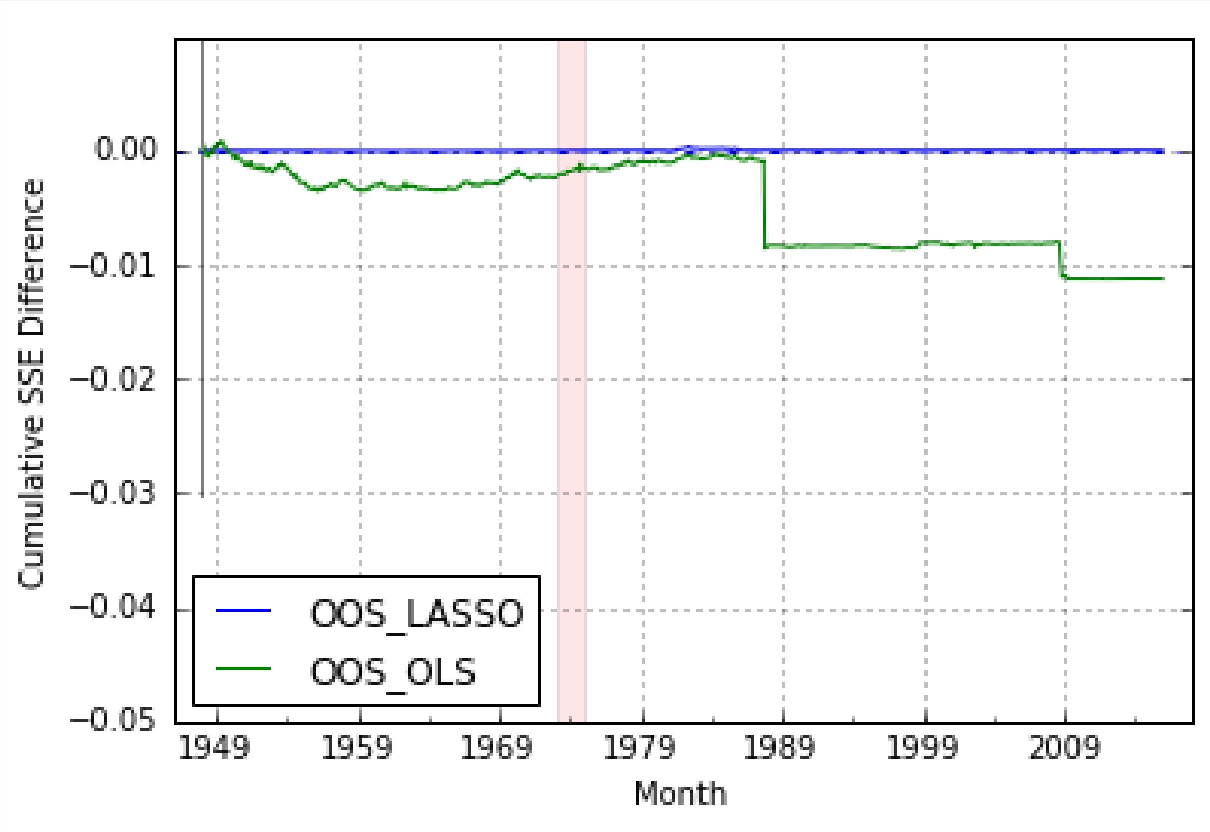
(a) expanding window



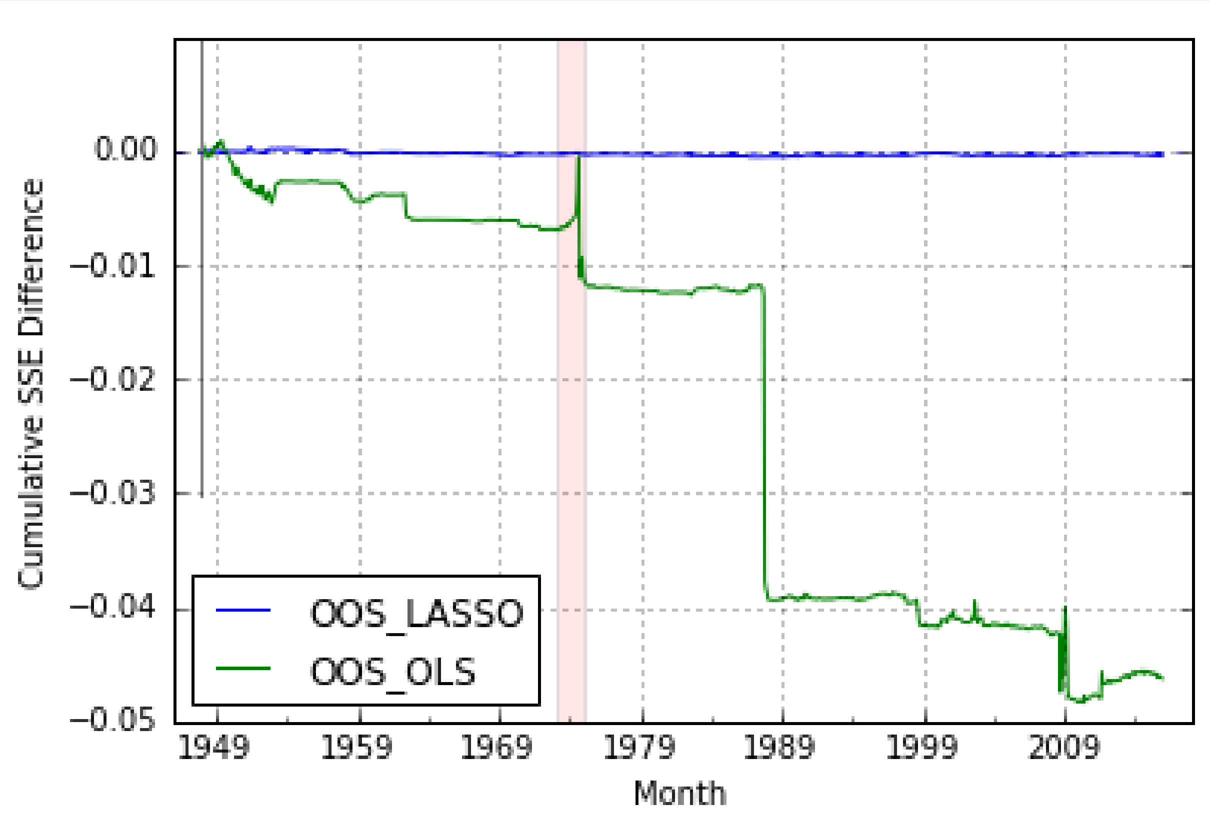
(b) rolling window

Figure: Monthly performance of bm

Empirical Results
LASSO with single predictor



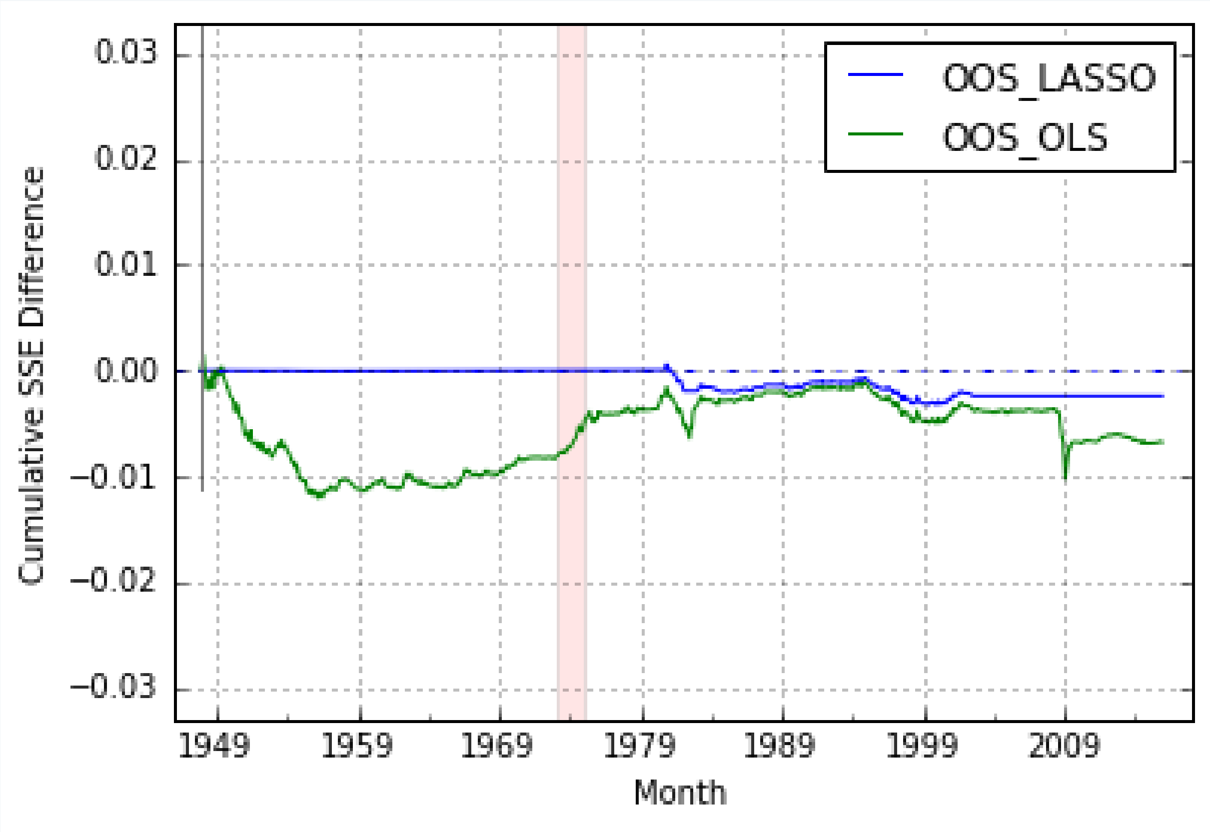
(a) expanding window



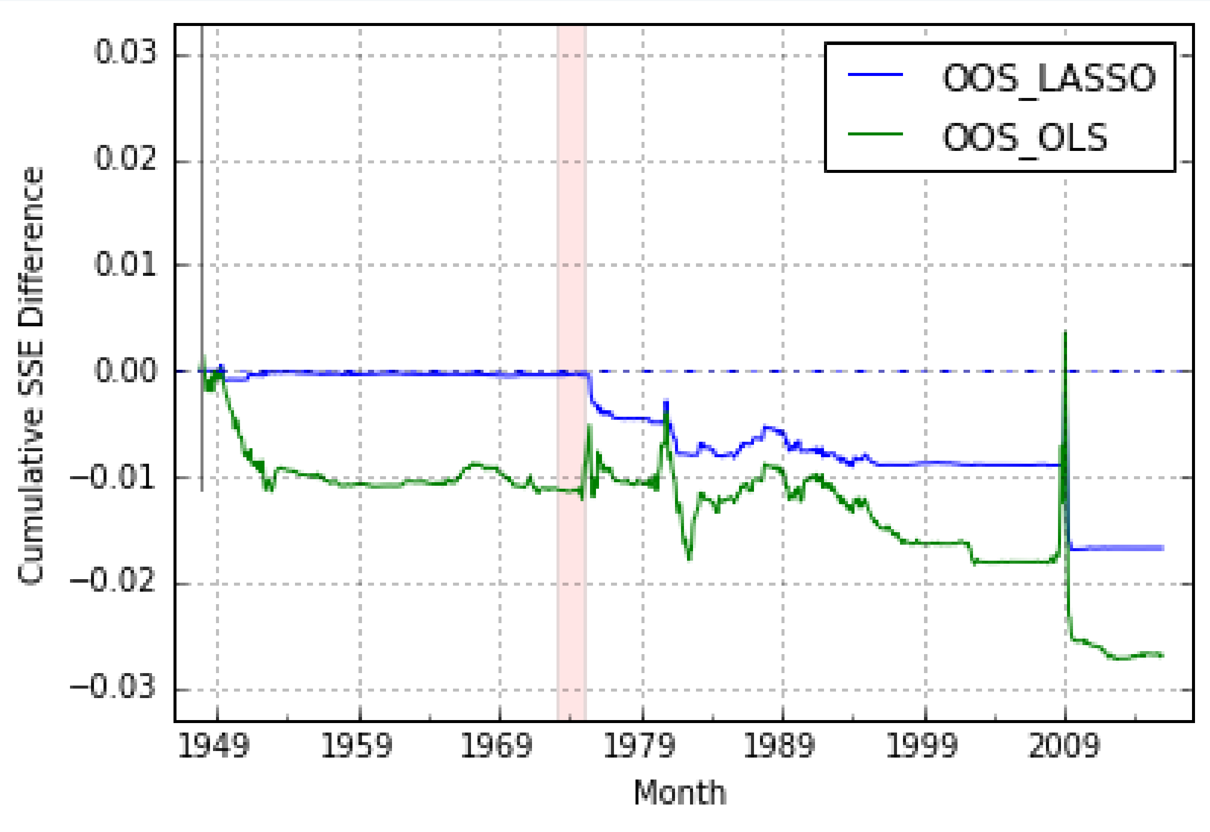
(b) rolling window

Figure: Monthly performance of svar

Empirical Results
LASSO with single predictor



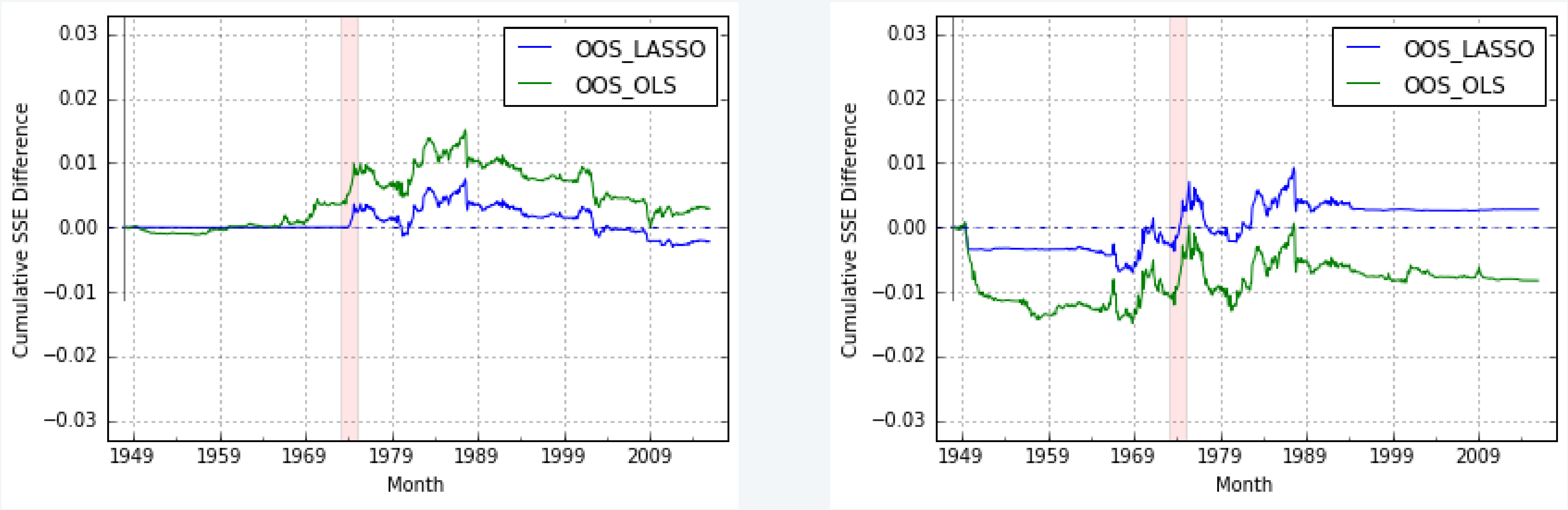
(a) expanding window



(b) rolling window

Figure: Monthly performance of dfy

Empirical Results
LASSO with single predictor



(a) expanding window

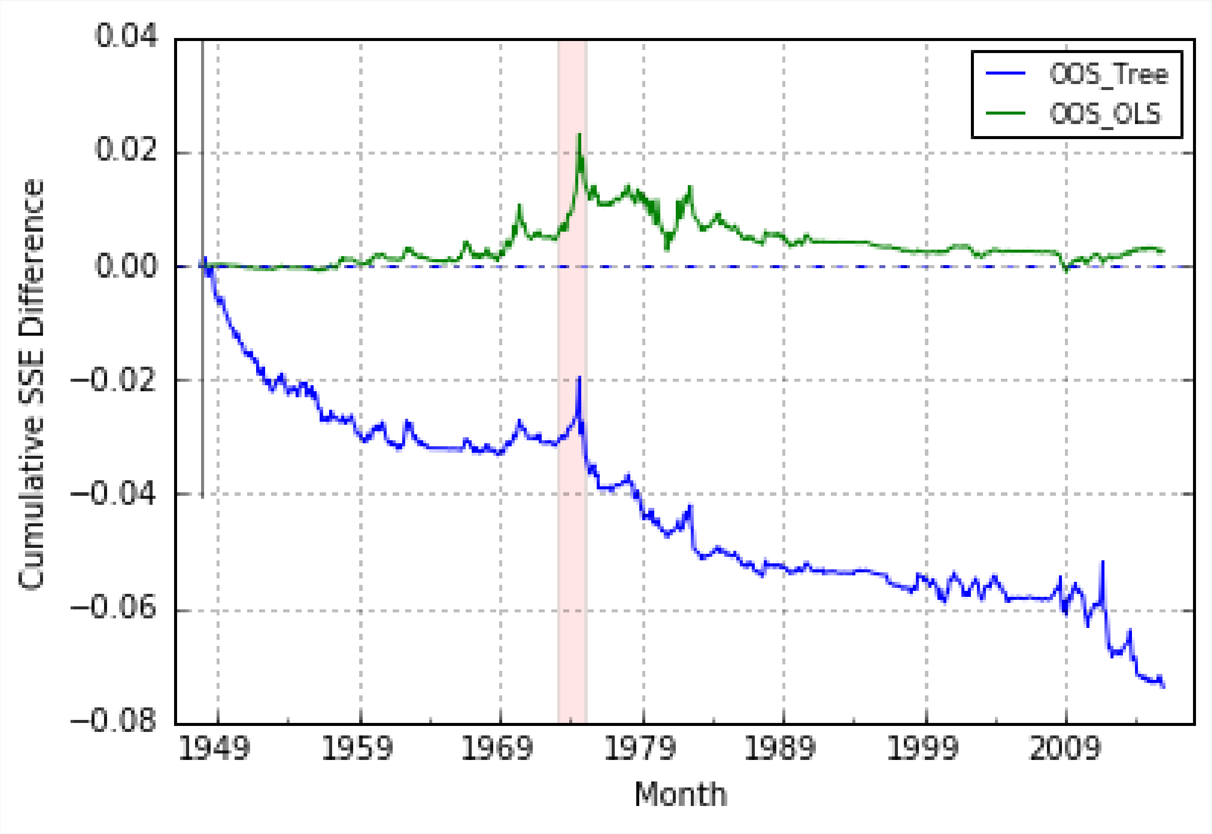
(b) rolling window

Figure: Monthly performance of tms

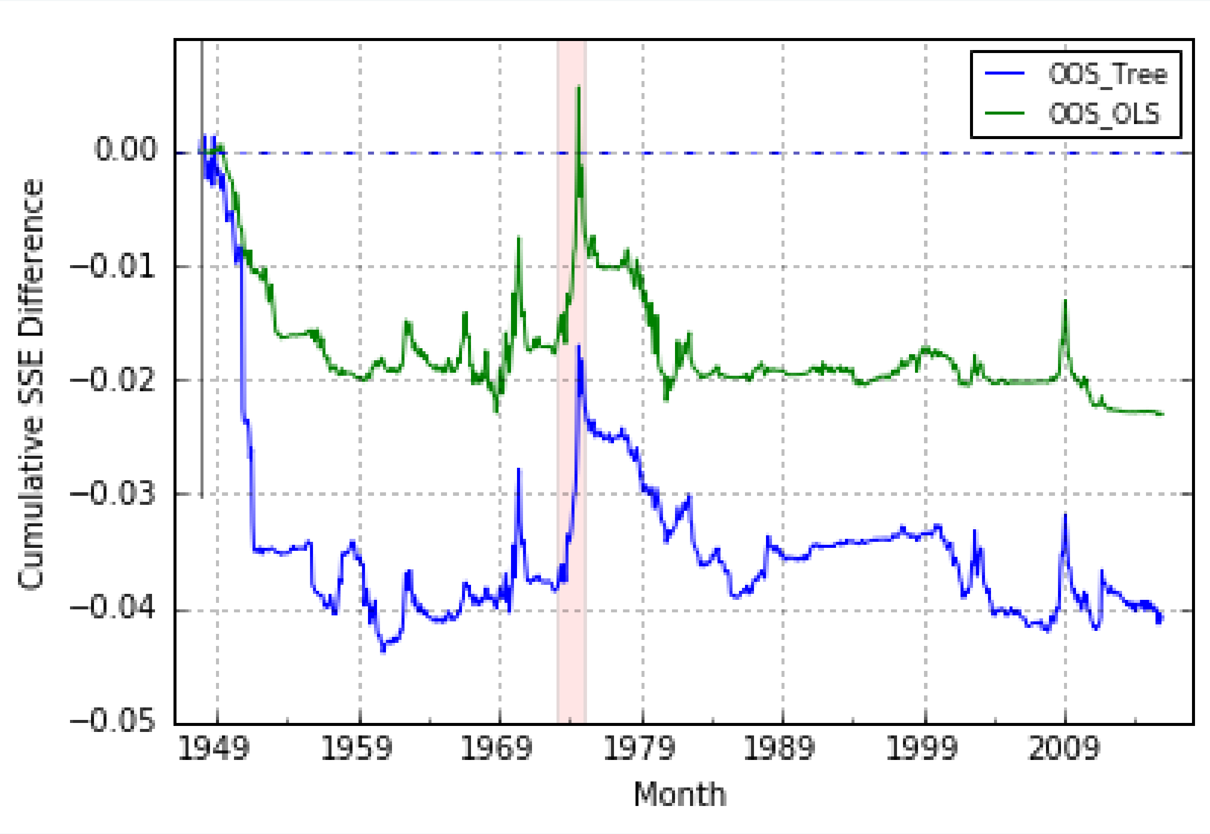
Empirical Results
Decision Tree with single predictor

- The Decision Tree consistently underperforms OLS for all 14 single variable models, with either rolling windows or expanding windows.
- We show the results of Decision Tree for **tbl**, **ep**, and **ltr** (more results can be found in the Appendix to this slideshow).

Empirical Results
Tree with single predictor



(a) expanding window

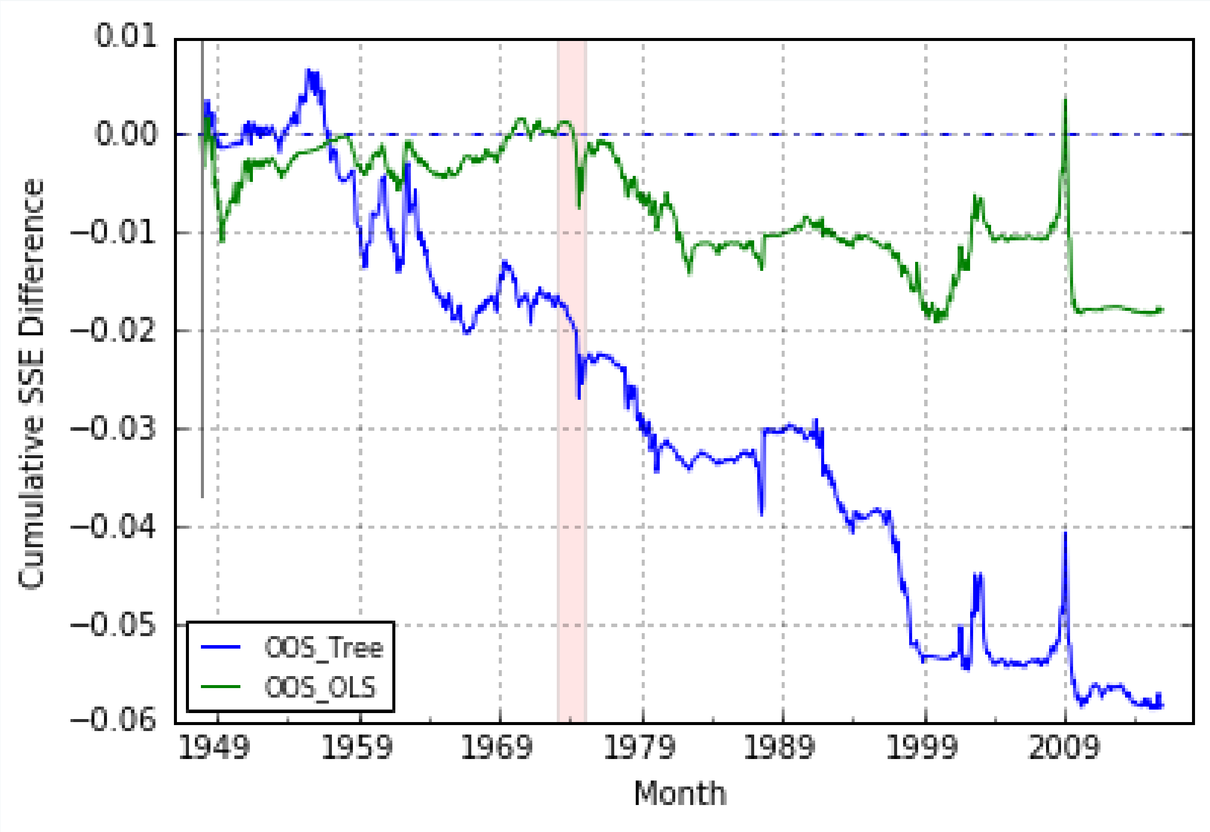


(b) rolling window

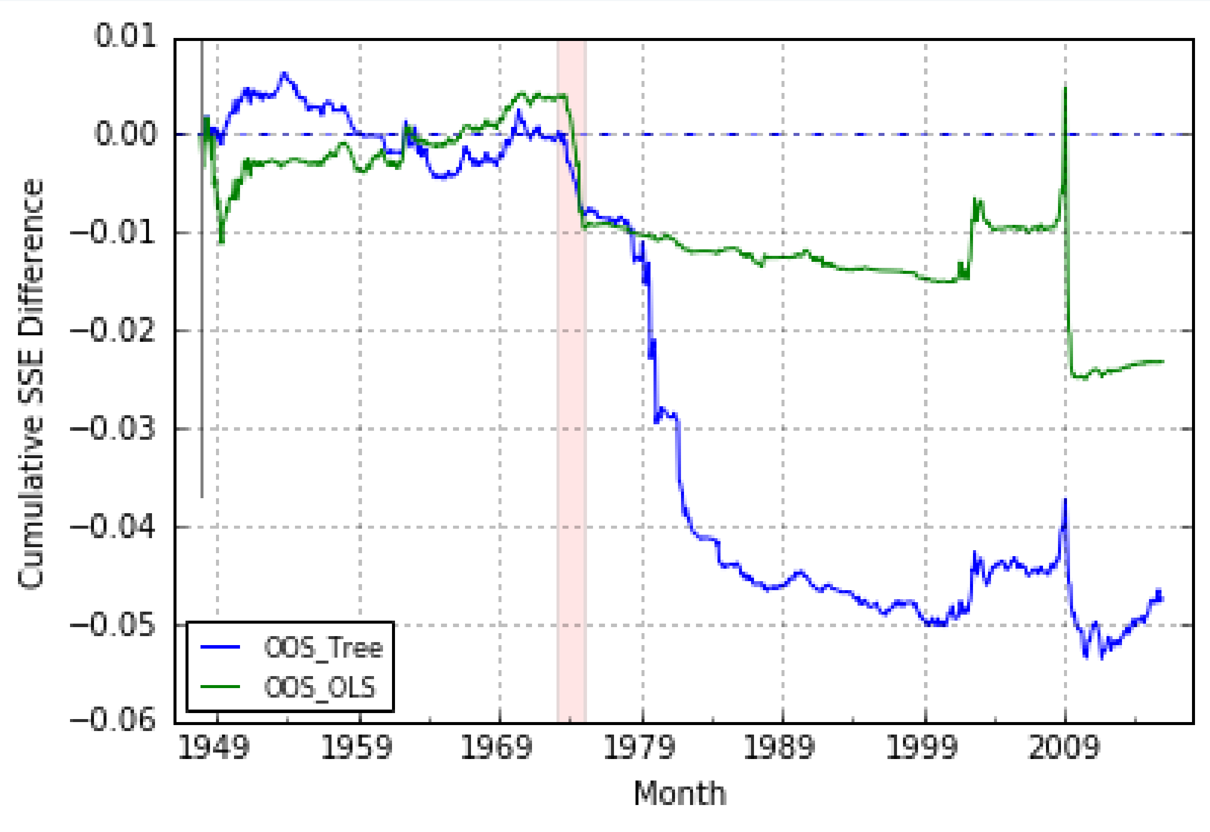
Figure: Monthly performance of tbl

Empirical Results

Tree with single predictor



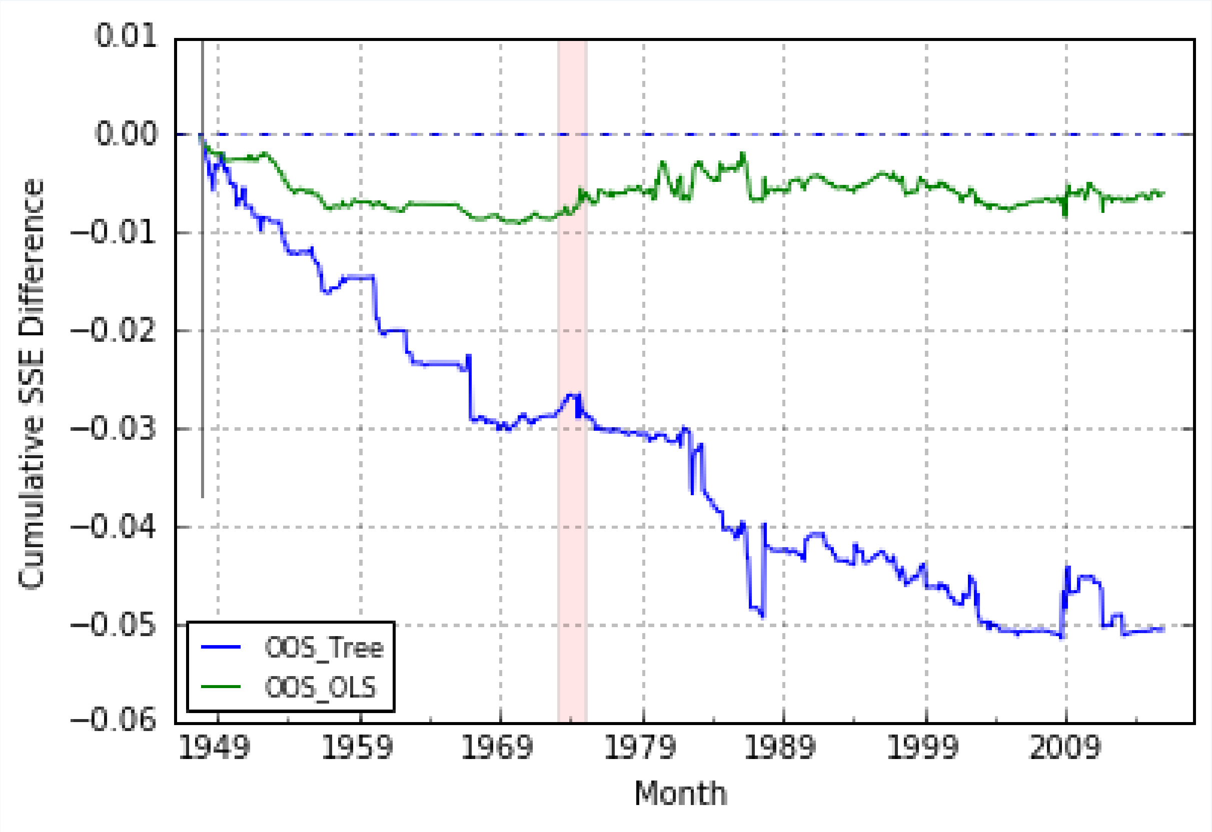
(a) expanding window



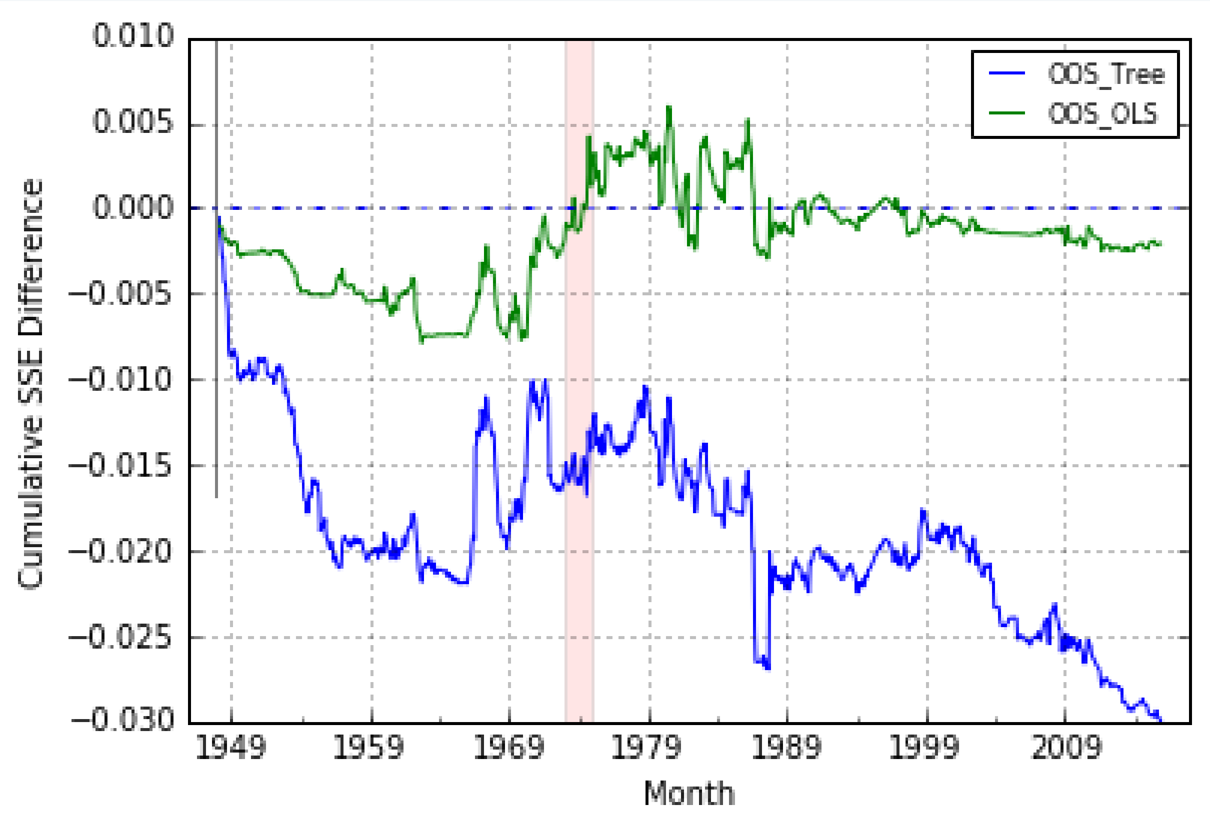
(b) rolling window

Figure: Monthly performance of ep

Empirical Results
Tree with single predictor



(a) expanding window



(b) rolling window

Figure: Monthly performance of ltr

Empirical Results
Multiple Predictor Models

- In the second part of our empirical results, we compare the OOS performance of OLS, LASSO, Decision Tree, and Random Forest using a kitchen sink model with all 14 predictor variables.
- A 240-month rolling window is used and the first prediction starts from January 1948.
- The main results are as follows:
  - Random Forest and LASSO outperform OLS, and LASSO provides the best OOS predictions.
  - Decision Tree performs worse than Random Forest. Since Random Forest consists of a large number of separately grown trees, it usually provides better prediction performance than Decision Tree.
  - The performance of Random Forest can be further improved with LASSO selected features.

Empirical Results
LASSO with Multiple Predictors

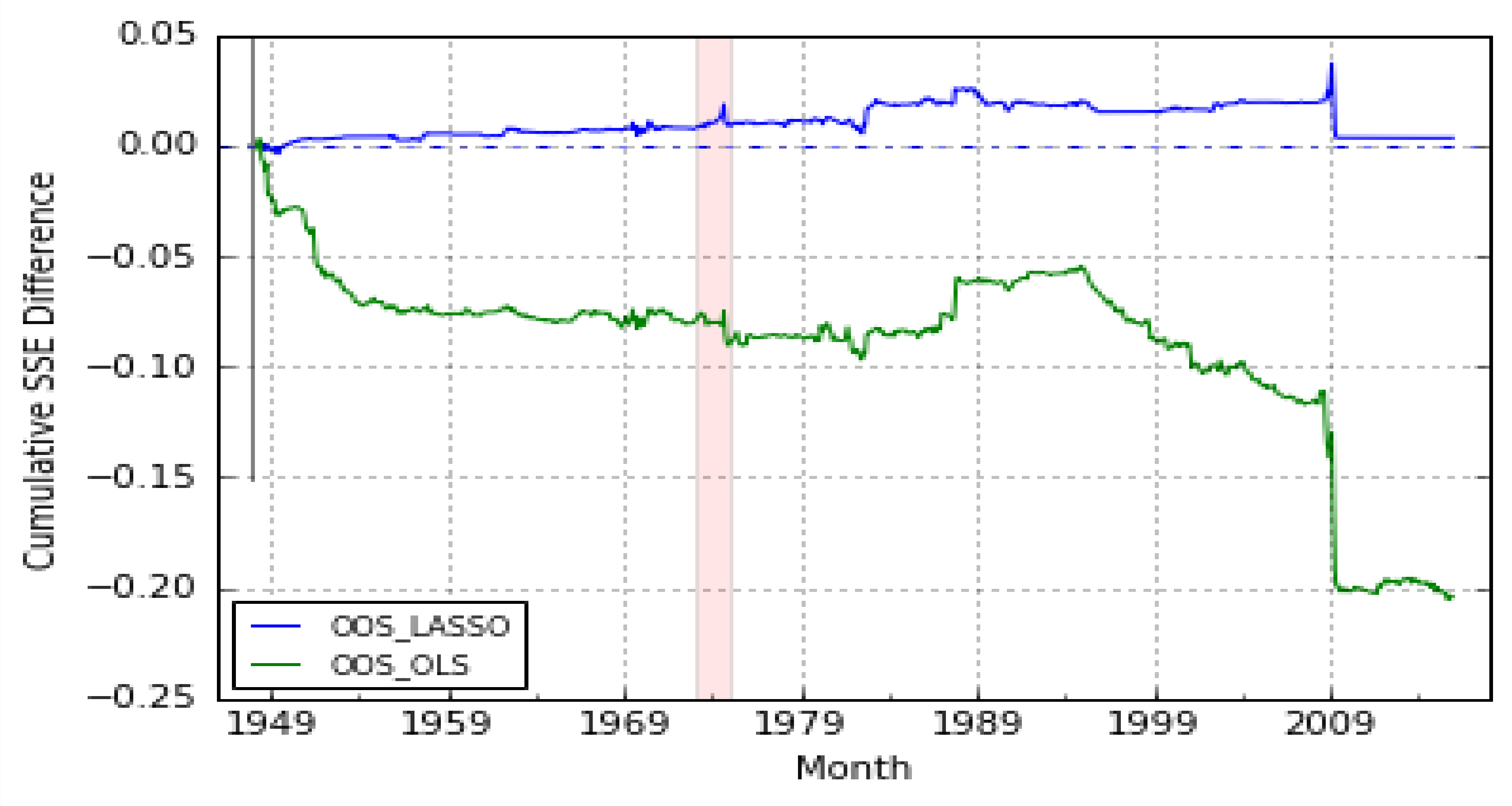


Figure: Monthly performance of LASSO with multiple predictors

Empirical Results
Decision Tree with Multiple Predictors

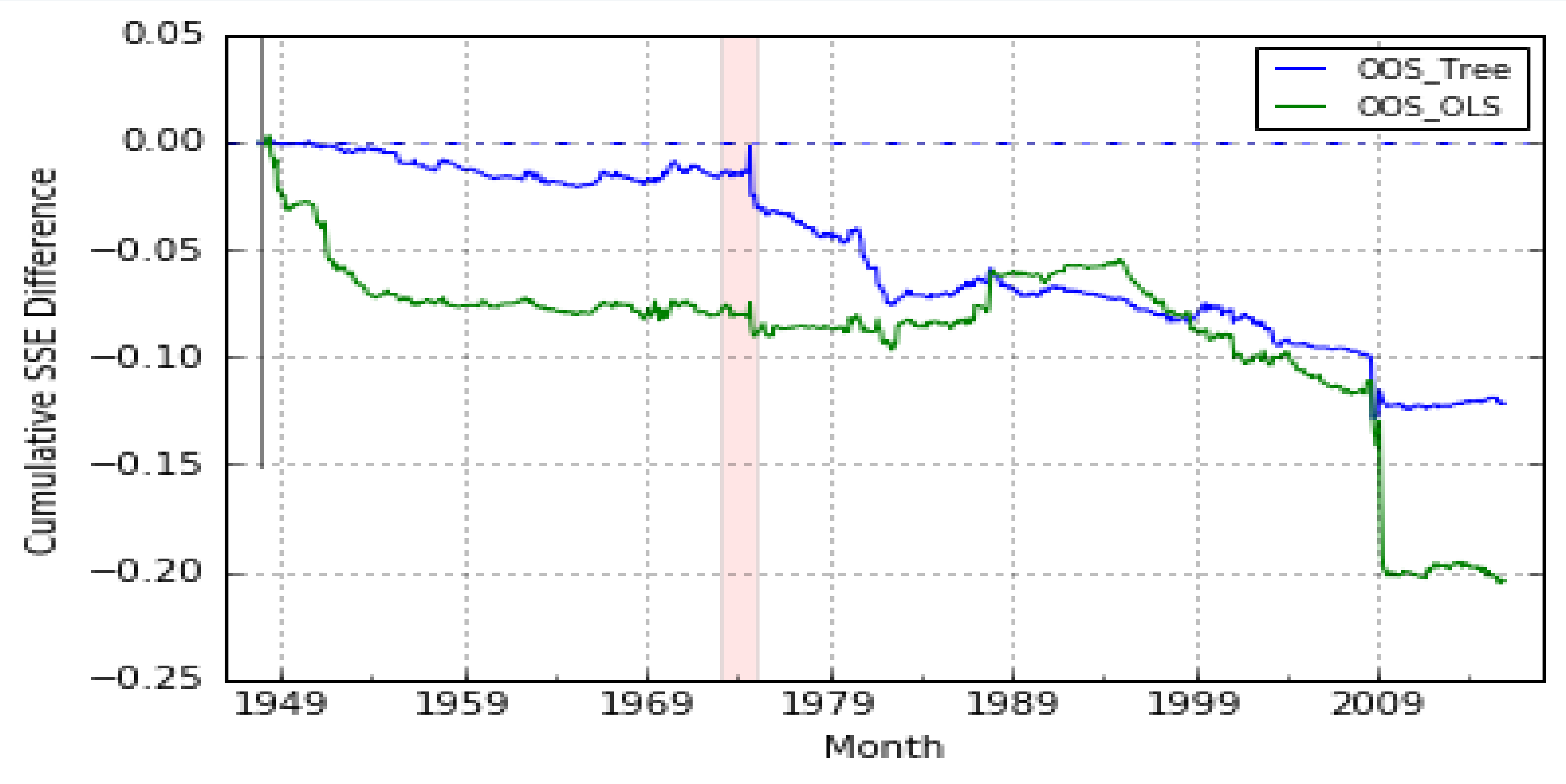


Figure: Monthly performance of Decision Tree with multiple predictors

Empirical Results

Random Forest with Multiple Predictors

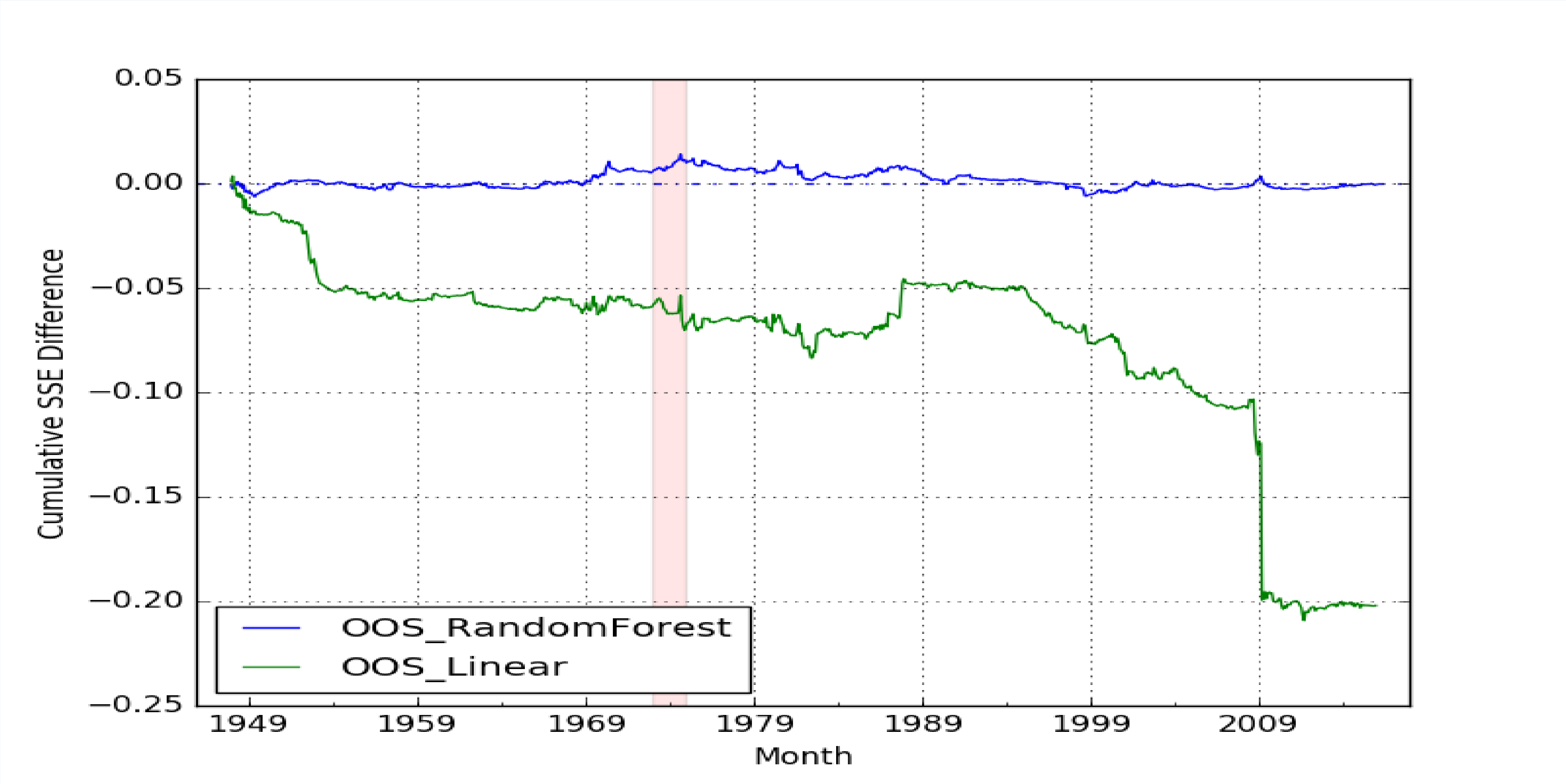


Figure: Monthly performance of Random Forest with multiple predictors

Empirical Results
Random Forest with LASSO selected features

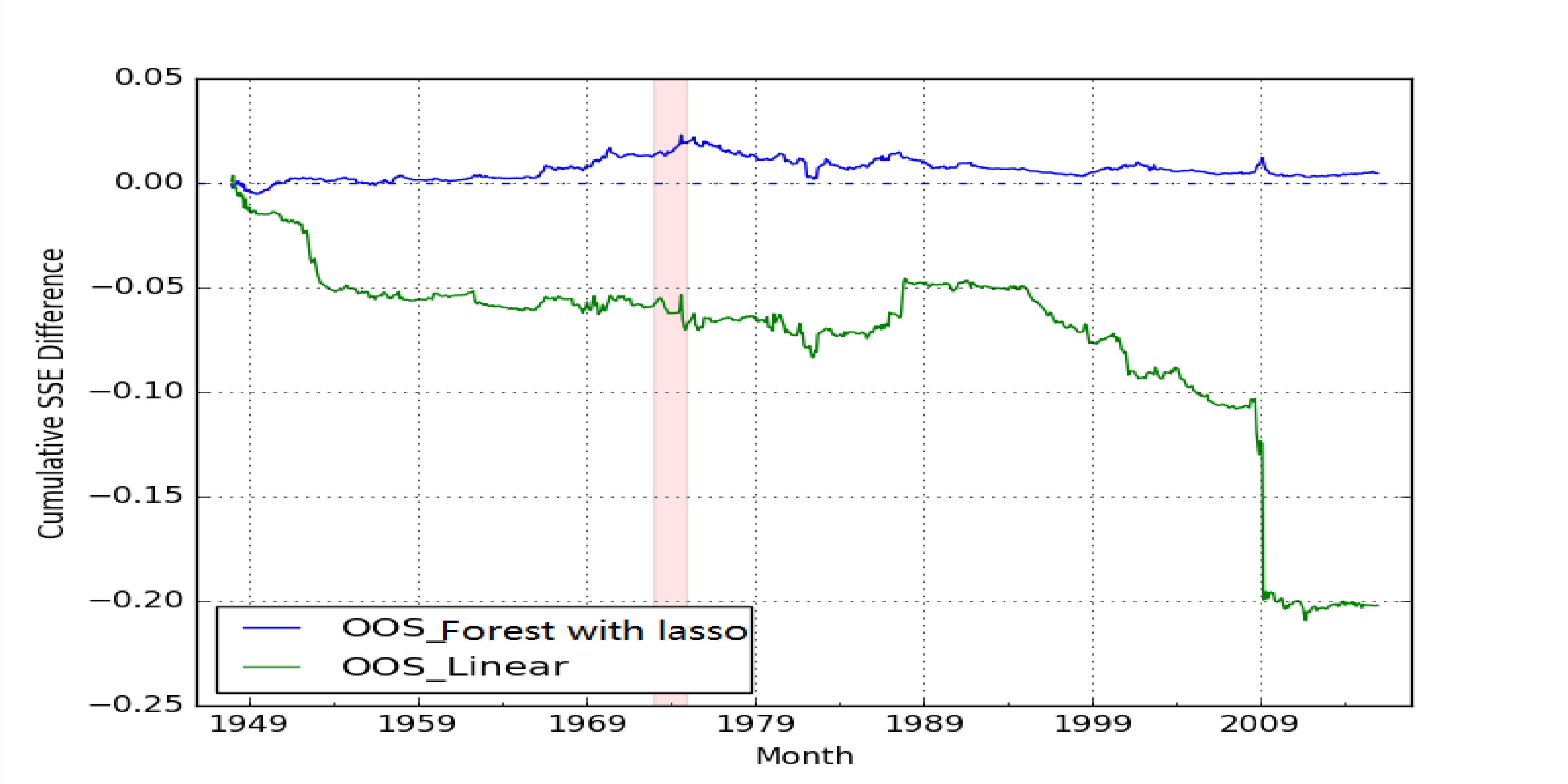


Figure: Monthly performance of Random Forest with LASSO selected features

Imperial College
Business School

Imperial means
Intelligent Business