ANALYSIS OF BAYESIAN SYMBOLIC REGRESSION APPLIED TO CRUDE OIL PRICE

Krzysztof Drachal*

Faculty of Economic Sciences,
University of Warsaw,
Warszawa, Poland

Abstract:

Forecasting crude oil spot prices even in moderate time horizon, like, 1-month, is not an easy task. Herein, the Bayesian Symbolic Regression (BSR) is applied to forecast the WTI spot prices. This novel method applies Bayesian symbolic trees methods to the Symbolic Regression. This type of econometric model can be especially useful when variable (feature) selection is necessary, as well as, model uncertainty emerges. Indeed, the literature claims that there can be several important crude oil price predictors. The great advantage of Symbolic Regression is its potential to "discover" the suitable functional form of a forecasting model.

In particular, world oil production, OECD petroleum consumption, U.S. stocks, MSCI World index, Chinese stock market index, VXO index, U.S. short-term interest rate, the Kilian global economic activity index and U.S. exchange rate were taken as explanatory variables. The period between 1989 and 2021 was analysed. Monthly data were taken.

Several models were taken as benchmarks. In particular, Dynamic Model Averaging (DMA), LASSO, RIDGE, the least-angle regressions, time-varying parameters regression, ARIMA and the no-change (NAÏVE) methods. Forecast accuracy was measured by Root Mean Square Error (RMSE) and other commonly used measures. Besides, forecasts were examined with the Diebold-Mariano test and Model Confidence Set testing procedure.

A strong evidence was found in favour of DMA and ARIMA as superior models (in a sense of forecast accuracy). However, BSR forecasts were found at least not less accurate than those from many competing (benchmark) models.

Keywords:

Bayesian Econometrics, Forecasting, Genetic Programming, Model Uncertainty, Symbolic Regression.

INTRODUCTION

Forecasting crude oil spot price is not an easy task. The problems addressed in this research are two folds. First, there exist numerous potentially important oil price drivers. Up to 1980s researchers were mostly focusing on supply and demand factors. Later, the impact of exchange rates was noticed. In 1990s much more has been observed about interactions with financial markets, for instance, stock market indices, market stress indices, etc. During 2000s and especially during the global financial crisis much attention has been brought to speculative pressures. However, nowadays it is common to consider numerous factors to be important oil price drivers. Besides those mentioned already, several uncertainty

Correspondence:
Krzysztof Drachal

e-mail:
kdrachal@wne.uw.edu.pl

DOI: 10.15308/Sinteza-2022-3-13

indices has been also constructed, etc. As a result there exists uncertainty about the variable selection (feature selection) when an econometric model, like, for example, a multilinear regression, is constructed [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12].

The second problem is that numerous researchers observed that the impact of the given driver on the oil price can be different in different time periods. Moreover, the set of the most important oil price drivers can vary in time. As a result, it is desired for the modelling scheme to be able to capture these two features: time-varying parameters and time-varying structure of the model itself [13] [14].

Besides, non-linear effects and complicated oil market structure are also linked with the functional form specification of a model. For example, whether some non-linear terms should be include in a regression model or not [15] [16].

Symbolic Regression is a method in which in the initial stage a set of operators is created. Then, mimicking the evolutionary processes of cross-over, mutation and selection, the suitable functional form of a model is discovered [17] [18]. Indeed, the standard approach to Symbolic Regression is with Genetic Algorithms [19]. However, recently, it was proposed to replace Genetic Algorithms in Symbolic Regression with Bayesian Symbolic Trees.

First, the evolutionary discovering of a suitable functional structure is then replaced by prior-posterior inference. Secondly, this methods seems to be computationally more efficient [20].

Herein, 1-month ahead crude oil spot prices are forecasted with the novel method of the Bayesian Symbolic Regression (BSR).

2. DATA

Monthly data between Jan 1989 and Apr 2021 were analysed. In particular, the time-series taken for the analysis are reported in Table 1. The motivation for the particular selection of explanatory variables is similar as in [21]. In particular, the transformation codes mean: 0 – no transformation: $Y_t \to Y_t$; 1 – 12-month logarithmic difference: $Y_t \to \log Y_t - \log Y_{t-12}$; 2 – logarithmic difference: $Y_t \to \log Y_t - \log Y_{t-1}$. Three different oil prices were taken as response variables to check for robustness of the outcomes. Before inserting into the models, the variables were also standardized [22]. Means and standard deviations for this transformations were computed on the basis of first 100 observations. Later, first 100 observations were taken as the in-sample period, and the next observations as the out-of-sample period.

Abbreviation	Description	Source	Transformation
WTI	Cushing, OK WTI spot price FOB (dollars per barrel)	[23]	2
Brent	Europe Brent spot price FOB (dollars per barrel)	[23]	2
Dubai	Crude oil, Dubai (dollars per barrel)	[23]	2
Prod_glob	World production of crude oil including lease condensate (Mb/d)	[24]	1
Cons_OECD	OECD refined petroleum products consumption (Mb/d)	[24]	1
Stocks	U.S. ending stocks excluding SPR of crude oil and petroleum products (thousand barrels)	[24]	1
MSCI_World	MSCI World (developed markets, standard, large + mid cap, dollars)	[25]	2
CHI	Chinese stock markets (Hang Seng and Shanghai Composite glued at Dec 1990)	[26]	2
VXO	VXO (month-end closing values)	[27]	0
R_short	U.S 3-month treasury bill: secondary market rate, percent, not seasonally adjusted (TB3MS)	[28]	0
Ec_act	Kilian index	[29]	0
FX	Real Narrow Effective Exchange Rate for United States, index 2010=100, not seasonally adjusted (RNUSBIS)	[28]	2

Table 1 – Time-series description.

3. METHODOLOGY

The models estimated in this research are listed in Table 2. If not stated otherwise, the models were estimated recursively, i.e., since the the in-sample period the forecast for time $t+1$ were computed with all the data available up to the time t [30]. The detailed description of the applied models can be found in the original papers cited.

In the Bayesian Symbolic Regression (BSR) the functional form is assumed to be a linear composition of (potentially non-linear) component functions. In other words, let $x_{1,t},\ldots,x_{n,t}$ be the explanatory time-series (after transformations). Then, it is assumed that the response time-series is modelled as

$$y_t=\beta_0+\beta_1*f_1\,(x_{1,1,t-1},\ldots,x_{1,i,t-1})+\cdots+\beta_K*f_K(x_{K,1,t-1},\ldots,x_{K,i,t-1})$$

with $x_{i,j,t}$ being some of explanatory variables out of $N = 9$ available ones, present in the i-th component, i.e., f_i, with $j=\{1,\ldots,N\}$ and $i =\{1,\ldots,K\}$. The number of components, K, must be kept fixed and set up in the first, initial stage of the BSR. Coefficients β_i are estimated by Ordinary Least Squares [20].

The following operators were used: $neg(x) = -x$ and standard addition +. This minimal set of operators let us focus on variable selection issue, not on aspects with non-linearities or variable transformation [31]. In the Metropolis-Hastings algorithm in the BSR models $M = 50$ iterations were done.

The original BSR considers the outcome derived in the last iteration. For the model averaging schemes the outcomes from all M iterations can be applied. In particular, let $y_1,\ldots,y_{50}$ be the forecasts obtained from 50 iterations. Then, let $w_1,\ldots,w_{50}$ be the weights ascribed to each of these forecasts. The weighted average forecast is defined then by $w_1*y_1+\cdots+w_{50}*y_{50}$ [32] [33] [34].

For the Symbolic Regression with Genetic Programming the population size was taken as 100 and generations were taken equal to 10. The cross-over probability was set up at 0.95 and subtree, hoist and point mutations probabilities were set up at 0.01. Root Mean Square Error (RMSE) was taken as the metric [35].

The cut-off limit for the Dynamic Model Averaging (DMA) and the Bayesian Model Averaging (BMA) with the dynamic Occam's window [36] [37] was set up at 0.5 and number of models in the combination scheme was limited to 1000 [30].

Variance in the state space equation was updated with Exponentially Weighted Moving Average method with the parameter $\kappa = 0.97$ [38]. In case of time-varying parameters regression two models were estimated. The first one with the forgetting factor $\lambda = 1$ (TVP), which corresponds to the BMA scheme. The second one with $\lambda=0.99$, which corresponds to the standard recommended forgetting values and the usual implementation of the DMA method.

The λ parameter (another kind of a parameter, specific for penalized regressions) in the LASSO and the least-angle regression (LARS) method was selected by the t-fold cross-validation using Mean Square Error (MAE) measure (with t denoting the time period). In the elastic net the following mixing parameters were considered: 0.1, 0.2, …, 0.9 [39] [40].

The forecast accuracy was measured with Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Scaled Error (MASE) [41]. It was measured for raw time-series (after backward transformations of the outcomes derived from the models). Two competing forecasts were compared with the Diebold-Mariano test [42] with the modification of Harvey et al. [43] and the Giacomini and Rossi fluctuation test [44] [45]. Three values of the parameter µ (corresponding to the part of the sample used in the rolling testing procedure) of the Giacomini and Rossi fluctuation test were checked: $\mu = \{ 0.3,0.4,0.5 \}$. They represent approximately 7, 10 and 12 years windows. If more forecasts were considered at the same time, they were compared with the Model Confidence Set (MCS) testing procedure with Squared Errors (SE) loss functions and TR statistic [46] [47].

All computations were done in R and Python [48] [49] [50] [51] [52].

Abbreviation	Description	Source
BSR rec	Bayesian Symbolic Regression (BSR) estimated recursively	[20] [53]
BSR av MSE rec	BSR model averaging with weights inversely proportional to Mean Square Error (MSE) estimated recursively	[20]
BSR av EW rec	BSR model averaging with equal weights estimated recursively	[20]
BSR av LL rec	BSR model averaging with weights inversely proportional to log-likelihood estimated recursively	[20]
GP rec	Symbolic Regression with Genetic Programming estimated recursively	[35]
GP av MSE rec	Symbolic Regression with Genetic Programming with weights inversely proportional to MSE estimated recursively	[35]
GP av EW rec	Symbolic Regression with Genetic Programming with equal weights estimated recursively	[35]
BSR fix	BSR with fixed parameters estimated over the in-sample period	[20] [53]
BSR av MSE fix	BSR model averaging with weights inversely proportional to MSE with fixed parameters estimated over the in-sample period	[20]
BSR av EW fix	BSR model averaging with equal weights with fixed parameters estimated over the in-sample period	[20]
BSR av LL fix	BSR model averaging with weights inversely proportional to log-likelihood with fixed parameters estimated over the in-sample period	[20]
GP fix	Symbolic Regression with Genetic Programming with fixed parameters estimated over the in-sample period	[35]
GP av MSE fix	Symbolic Regression with Genetic Programming with weights inversely proportional to MSE with fixed parameters estimated over the in-sample period	[35]
GP av EW fix	Symbolic Regression with Genetic Programming with equal weights with fixed parameters estimated over the in-sample period	[35]
Abbreviation	**Description**	**Source**
DMA	Dynamic Model Averaging with dynamic Occam's window	[36] [37] [30]
BMA	Bayesian Model Averaging with dynamic Occam's window	[36] [37] [30]
DMA 1V	Dynamic Model Averaging over one-variable models	[36] [30]
DMS 1V	Dynamic Model Selection over one-variable models	[36] [30]
BMA 1V	Bayesian Model Averaging over one-variable models	[36] [30]
BMS 1V	Bayesian Model Selection over one-variable models	[36] [30]
LASSO	LASSO regression estimated recursively	[39]
RIDGE	RIDGE regression estimated recursively	[39]
EN	Elastic net estimated recursively	[39]
B-LASSO	Bayesian LASSO regression estimated recursively	[54]
B-RIDGRE	Bayesian RIGDE regression estimated recursively	[54]
LARS	Least-angle regression estimated recursively	[40]
TVP	Time-Varying Parameters regression ($\lambda = 1$)	[36] [30]
TVP f	Time-Varying Parameters regression ($\lambda = 0.99$)	[36] [30]
ARIMA	Auto ARIMA	[55]
HA	Historical Average (recursive)	[56]
NAÏVE	No-change forecast	[56]

Table 2 – Models used in the research.

4. RESULTS

Table 3 reports forecast accuracy measures for the in-sample period, which was used to select the number of components K. In case of the WTI price RMSE was minimised for $K = 6$, but two other measures were minimised with $K = 5$. As a result, $K = 5$ was taken for the further analysis. In case of the Brent and the Dubai prices $K = 9$ was preferred.

The out-of-sample forecast accuracy is reported in Table 4. In case of the WTI price all measures were minimised by the DMA model. The BSR models tend to generate smaller errors than the corresponding GP models and the NAÏVE method. However, quite a few models were able to generate smaller errors than the BSR models. Similar conclusions can be derived for the Brent and the Dubai prices. Anyways, the BSR errors were smaller than those from the NAÏVE method. Secondly, it seems that applying model averaging schemes slightly improved the forecast accuracies.

Table 5 reports the outcomes of the Diebold-Mariano test. The DMA model was chosen as "the best" one. In case of the DMA model the alternative hypothesis of the test was that the DMA forecasts are more accurate than those from the competing tested model. In case of the BSR rec model the alternative hypothesis was that the BSR forecasts were less accurate than those from the competing tested model. The null hypothesis was that both forecasts have the same accuracy. Assuming 5% significance level, for the WTI price forecasts generated by the DMA model were significantly more accurate than all other forecasts except those from the ARIMA model. On the other hand, the forecasts from the BSR rec model can be said to be significantly less accurate only comparing with the forecasts generated from the DMA model. They cannot be said to be significantly less accurate than the forecasts from the other models. The same conclusions can be derived for the Brent price. In case of the Dubai prices, additionally, the BSR rec forecasts were significantly less accurate than those from the ARIMA model.

Table 6 reports the outcomes from the Diebold-Mariano test, in which forecasts from the recursively estimated version of a given model were compared with the corresponding fixed estimations. The alternative hypothesis was that the recursive version generated more accurate forecasts than the fixed version of a model. In case of the WTI price it can be said that, indeed, for BSR av MSE, BSR av LL and GP models recursive estimations improved forecast accuracy comparing with the fixed estimations. The same conclusions are valid for the Brent and the Dubai prices.

The MCS procedure generated the superior set of models in which the DMA and the ARIMA models remained in case of the WTI price. The corresponding p-value was 0.4534. The same sets of models were generated by the MCS procedure in case of the Brent and the Dubai prices. The corresponding p-values were 0.1466 for the Brent price and 0.1372 for the Dubai price.

The outcomes from the Giacomini and Rossi fluctuation test are presented in Figure 1. Only the outcomes for the WTI price and one value of μ are reported. However, for other oil prices or values of μ the outcomes were quite similar. Therefore, to keep this text concise only this selected outcome is reported.

	WTI			Brent			Dubai		
K	RMSE	MAE	MASE	RMSE	MAE	MASE	RMSE	MAE	MASE
1	2.4876	1.3295	1.1349	2.4739	1.3536	1.1295	2.1942	1.1858	1.1092
2	2.4809	1.3255	1.1316	2.4189	1.3292	1.1092	2.1788	1.2162	1.1376
3	2.4674	1.3244	1.1306	2.3824	1.3554	1.1310	2.1818	1.2135	1.1351
4	2.4420	1.3593	1.1604	2.4002	1.3062	1.0900	2.1931	1.1871	1.1104
5	2.4092	**1.3067**	**1.1155**	2.3946	1.3066	1.0903	2.1859	1.1964	1.1191
6	**2.3949**	1.3117	1.1198	2.4120	1.3669	1.1406	2.1668	1.1960	1.1187
7	2.4496	1.3074	1.1161	2.3358	1.3316	1.1112	2.1865	1.2153	1.1368
8	2.4207	1.3497	1.1522	2.4055	1.3833	1.1543	2.1595	1.2094	1.1313
9	2.4190	1.3462	1.1492	**2.3244**	**1.2863**	**1.0733**	**2.0969**	**1.1768**	**1.1007**

Table 3 – Forecast accuracies (in-sample).

	WTI			Brent			Dubai		
Model	RMSE	MAE	MASE	RMSE	MAE	MASE	RMSE	MAE	MASE
BSR rec	5.0792	3.6857	0.9961	5.1259	3.7679	0.9839	4.8835	3.3938	0.9724
BSR av MSE rec	5.1073	3.6441	0.9848	5.1672	3.7598	0.9818	4.8248	3.3822	0.9690
BSR av EW rec	5.1016	3.6307	0.9812	5.1582	3.7490	0.9790	4.8322	3.3900	0.9713
BSR av LL rec	5.1022	3.6368	0.9828	5.1606	3.7431	0.9774	4.8413	3.3877	0.9706
GP rec	5.4779	**3.8469**	**1.0396**	5.6416	4.0173	1.0490	5.3440	3.7039	1.0612
GP av MSE rec	**5.3892**	3.7878	1.0236	5.5263	3.9548	1.0327	5.2311	3.6452	1.0444
GP av EW rec	5.3946	3.7799	1.0215	5.5369	3.9450	1.0302	5.2370	3.6313	1.0404
BSR fix	6.1367	4.2968	1.1612	9.2911	6.5745	1.7168	5.5289	4.0240	1.1530
BSR av MSE fix	6.5785	4.7687	1.2887	**7.2963**	**5.3265**	**1.3909**	**6.5456**	**4.7037**	**1.3477**
BSR av EW fix	6.4214	4.6812	1.2651	**7.2407**	**5.2980**	**1.3835**	**6.4014**	**4.5997**	**1.3179**
BSR av LL fix	6.5152	4.7128	1.2736	**6.9637**	**5.0966**	**1.3309**	**6.3896**	**4.5697**	**1.3093**
GP fix	5.3620	3.9174	1.0587	**5.4922**	**4.0507**	**1.0578**	**5.1852**	**3.7224**	**1.0665**
Model	RMSE	MAE	MASE	RMSE	MAE	MASE	RMSE	MAE	MASE
DMA	**4.9166**	**3.5932**	**0.9711**	**4.9258**	**3.6853**	**0.9624**	**4.5489**	**3.2399**	**0.9283**
BMA	5.1548	3.6720	0.9923	5.1569	3.7849	0.9884	4.8004	3.3708	0.9658
DMA 1V	5.1657	3.6960	0.9988	5.1978	3.7946	0.9909	4.8517	3.3393	0.9568
DMS 1V	5.0884	3.6901	0.9972	5.1461	3.7903	0.9898	4.8482	3.3641	0.9639
BMA 1V	5.2508	**3.7007**	**1.0001**	5.2682	3.8158	0.9964	4.9505	3.4093	0.9768
BMS 1V	**5.2799**	3.7278	1.0074	5.2792	3.8371	1.0020	4.9581	3.4332	0.9837
LASSO	5.1600	3.6683	0.9913	5.1897	3.7746	0.9857	4.8655	3.4006	0.9743
RIDGE	5.1451	3.6522	0.9870	5.1583	3.7473	0.9785	4.8571	3.3842	0.9696
EN	5.1510	3.6638	0.9901	**5.1846**	**3.7702**	**0.9845**	**4.8785**	**3.4035**	**0.9751**
B-LASSO	5.2384	3.6677	0.9912	**5.2712**	**3.7894**	**0.9895**	**4.9649**	**3.4153**	**0.9786**
B-RIDGRE	5.1657	3.6475	0.9857	**5.2548**	**3.7959**	**0.9912**	**4.9187**	**3.3923**	**0.9720**
LARS	5.1502	3.6911	0.9975	**5.1288**	**3.7992**	**0.9921**	**4.8053**	**3.4520**	**0.9890**
TVP	5.1557	3.7284	1.0076	**5.1519**	**3.8355**	**1.0016**	**4.7752**	**3.4534**	**0.9895**
TVP f	5.0883	3.7089	1.0023	**5.1169**	**3.8350**	**1.0014**	**4.7424**	**3.4034**	**0.9751**
ARIMA	5.0009	3.6229	0.9791	**5.1405**	**3.7417**	**0.9771**	**4.6827**	**3.3023**	**0.9462**
HA	33.2715	24.5524	6.6352	**37.4096**	**27.4032**	**7.1559**	**36.7332**	**26.9924**	**7.7338**
NAIVE	5.2934	3.7003	1.0000	**5.3930**	**3.8295**	**1.0000**	**5.0807**	**3.4902**	**1.0000**

Table 4 – Forecast accuracies (out-of-sample).

It can be seen that definitely HA method generated significantly less accurate forecasts than the DMA method over almost all analysed period. Between 2008 and 2014 also GP models generated significantly less accurate forecasts. However, for the other models it seems that the significance of generating more accurate forecasts by the DMA model was mostly temporary.

This feature is the most clearly seen for periods around 2011 and 2014. The outcomes from the Giacomini and Rossi fluctuation test weaken the evidence provided by the Diebold-Mariano test already reported previously in this text.

	WTI		Brent		Dubai	
BSR rec	**0.0332**		**0.0289**		**0.0083**	
BSR av MSE rec	**0.0231**	0.7057	**0.0115**	0.7972	**0.0075**	0.1359
BSR av EW rec	**0.0256**	0.6673	**0.0134**	0.7202	**0.0054**	0.1722
BSR av LL rec	**0.0234**	0.6859	**0.0150**	0.7318	**0.0071**	0.2007
GP rec	**0.0055**	0.9890	**0.0013**	0.9988	**0.0007**	0.9968
GP av MSE rec	**0.0097**	**0.9792**	**0.0025**	0.9978	**0.0012**	0.9927
GP av EW rec	**0.0111**	0.9761	**0.0029**	0.9972	**0.0016**	0.9907
BSR fix	**0.0001**	0.9999	**0.0000**	1.0000	**0.0000**	0.9978
BSR av MSE fix	**0.0000**	0.9999	**0.0000**	1.0000	**0.0000**	0.9998
BSR av EW fix	**0.0000**	0.9999	**0.0000**	1.0000	**0.0000**	0.9997
BSR av LL fix	**0.0000**	0.9999	**0.0000**	1.0000	**0.0000**	0.9998
GP fix	**0.0043**	0.9896	**0.0005**	0.9992	**0.0002**	0.9895
GP av MSE fix	**0.0047**	0.9884	**0.0005**	0.9991	**0.0003**	0.9883
GP av EW fix	**0.0079**	0.9802	**0.0011**	0.9983	**0.0006**	0.9796
DMA		**0.0332**		**0.0289**		**0.0083**
BMA	**0.0256**	0.7901	**0.0396**	0.6146	**0.0066**	0.1821
DMA 1V	**0.0137**	0.8418	**0.0082**	0.8116	**0.0107**	0.3722
DMS 1V	**0.0341**	0.5469	**0.0117**	0.5813	**0.0077**	0.3828
BMA 1V	**0.0035**	0.9800	**0.0050**	0.9442	**0.0026**	0.8008
BMS 1V	**0.0026**	0.9816	**0.0096**	0.9076	**0.0012**	0.7580
LASSO	**0.0096**	0.8375	**0.0106**	0.8115	**0.0062**	0.3795
RIDGE	**0.0125**	0.8403	**0.0120**	0.7223	**0.0080**	0.3186
EN	**0.0128**	0.8271	**0.0113**	0.8082	**0.0052**	0.4625
B-LASSO	**0.0101**	0.9569	**0.0073**	0.9785	**0.0057**	0.8650
B-RIDGRE	**0.0211**	0.8251	**0.0069**	0.9410	**0.0117**	0.6692
LARS	**0.0047**	0.8135	**0.0094**	0.5144	**0.0040**	0.1323
TVP	**0.0024**	0.8102	**0.0061**	0.6039	**0.0047**	0.1357
TVP f	**0.0015**	0.5325	**0.0036**	0.4723	**0.0042**	0.1687
ARIMA	0.2327	0.2060	0.0554	0.5598	0.1012	**0.0414**
HA	**0.0000**	1.0000	**0.0000**	1.0000	**0.0000**	1.0000
NAIVE	**0.0199**	0.9565	**0.0063**	0.9938	**0.0035**	0.9602

Table 5 – The Diebold-Mariano test outcomes (p-values).

5. CONCLUSIONS

Herein, the novel Bayesian Symbolic Regression (BSR) method was applied to forecasting 1-month ahead the WTI, the Brent and the Dubai oil spot prices. Also, the usual Symbolic Regression, i.e., based on the Genetic Programming was applied. Besides, several models dealing with variable uncertainty were also estimated. Common and standard benchmark models, like, the no-change (NAÏVE) or ARIMA models were also estimated. First of all, some small improvement in forecast accuracies can be found from model averaging schemes (i.e., applying forecast combination as the final prediction, instead of model selection approach,

	WTI	Brent	Dubai
BSR	0.2943	0.2028	0.8641
BSR av MSE	0.6662	0.6601	0.1100
BSR av EW	**0.0092**	**0.0022**	**0.0020**
BSR av LL	**0.0001**	**0.0000**	**0.0010**
GP	**0.0153**	**0.0018**	**0.0157**
GP av MSE	0.5803	0.5887	0.6178
GP av EW	0.6879	0.7158	0.7386

Table 6 - The Diebold-Mariano test outcomes (p-values): recursive vs fixed estimations.

in which the final prediction is taken from exactly one component model). Secondly, it could be found that, statistically significantly, the Dynamic Model Averaging and the ARIMA generated more accurate forecasts than all other models. If the common Diebold-Mariano test (considering the whole out-of-sample period at once) supports this conclusion, the Giacomini-Rossi fluctuation test (which distincts sub-periods in the analysed period) provides less evidence towards such a conclusion. However, there is also no significant evidence to state that the BSR models generated less accurate forecasts than the competing benchmark models. Finally, in moderate number of cases recursive estimations significantly improved forecast accuracy over fixed parameters estimation.

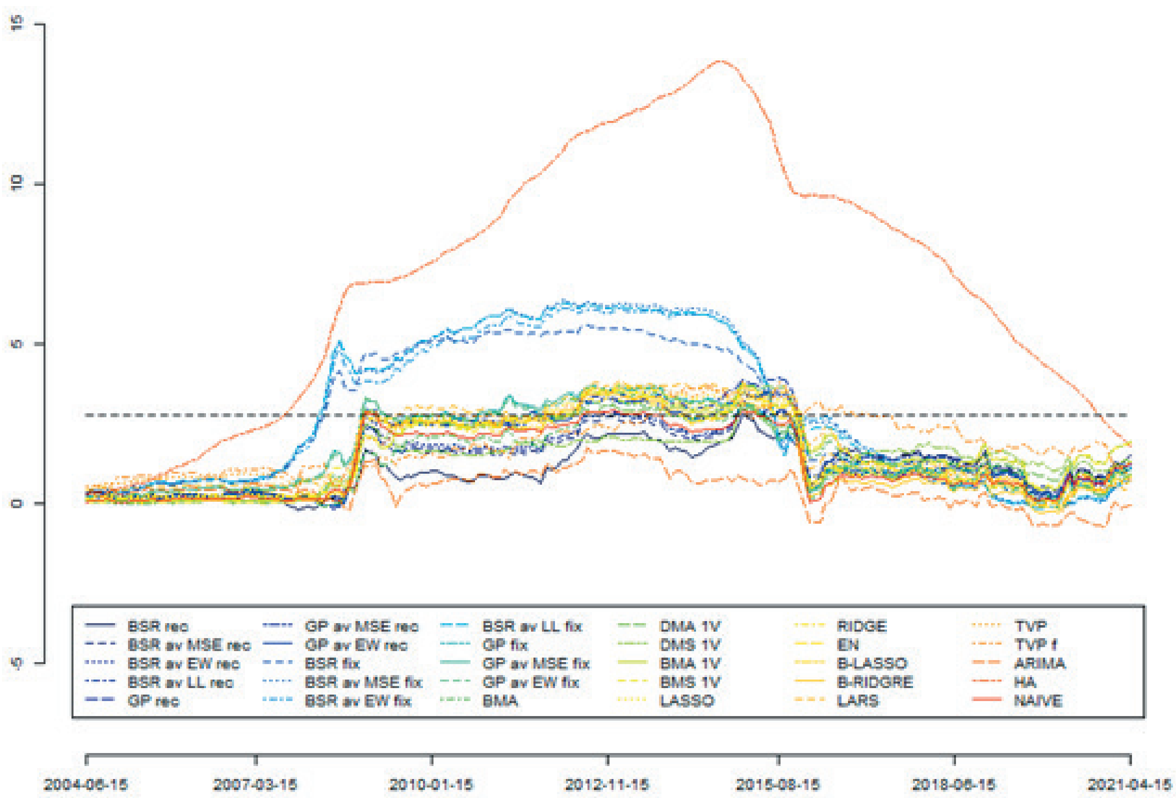

Figure 1 – The outcomes from the Giacomini and Rossi fluctuation test for the WTI price with $\mu = 0.3$.

6. ACKNOWLEDGEMENTS

Research funded by the grant of the National Science Centre, Poland, under the contract number DEC-2018/31/B/HS4/02021.

7. REFERENCES

[1] Y. Wang, L. Liu and C. Wu, "Forecasting the real prices of crude oil using forecast combinations," *Energy Economics,* vol. 66, pp. 337-348, 2017.

[2] S. A. Basher, A. A. Haug and P. Sadorsky, "Oil prices, exchange rates and emerging stock markets," *Energy Economics,* vol. 34, pp. 227-240, 2012.

[3] O. Arslan-Ayaydin and I. Khagleeva, "The dynamics of crude oil spot and futures market," in *Energy Economics and Financial Markets*, Berlin, Springer-Verlag, 2013, pp. 159-173.

[4] J. Manickavasagam, "Drivers of global crude oil price: A review," *IMI Konnect,* vol. 9, pp. 69-75, 2020.

[5] J. Beckmann, R. L. Czudaj and V. Arora, "The relationship between oil prices and exchange rates: Revisiting theory and evidence," *Energy Economics,* vol. 88, p. 104772, 2020.

[6] H. Miao, S. Ramchander, T. Wang and D. Yang, "Influential factors in crude oil price forecasting," *Energy Economics*, vol. 68, pp. 77-88, 2017.

[7] T. Perifanis and A. Dagoumas, "Crude oil price determinants and multi-sectoral effects: A review," Energy Sources, *Part B: Economics, Planning, and Policy*, p. In Press, 2021.

[8] C. W. Su, M. Qin, R. Tao, N. C. Moldovan and O. R. Lobont, "Factors driving oil price - From the perspective of United States," *Energy* , vol. 197, p. 117219, 2020.

[9] W. H. Alredany, "A regression analysis of determinants affecting crude oil price," *International Journal of Energy Economics and Policy*, vol. 8, pp. 110-119, 2018.

[10] Y. C. Chen, K. S. Rogoff and B. Rossi, "Can exchange rates forecast commodity prices?," *The Quarterly Journal of Economics*, vol. 125, pp. 1145-1194, 2010.

[11] L. Juvenal and I. Petrella, "Speculation in the oil market," *Journal of Applied Econometrics*, vol. 30, pp. 621-649, 2014.

[12] Z. Liu, Z. Ding, T. Lv, J. S. Wu and W. Qiang, "Financial factors affecting oil price change and oil-stock interactions: A review and future perspectives," *Natural Hazards*, vol. 95, pp. 207-225, 2019.

[13] M. E. H. Arouri, T. H. Dinh and D. K. Nguyen, "Time-varying predictability in crude-oil markets: The case of GCC countries," *Energy Policy*, vol. 38, pp. 4371-4380, 2010.

[14] N. B. Behmiri and J. R. Pires Manso, "Crude oil price forecasting techniques: A comprehensive review of literature," *CAIA Alternative Investment Analyst Review*, vol. 2, pp. 30-48, 2013.

[15] J. D. Hamilton, "Nonlinearities and the macroeconomic effects of oil prices," *Macroeconomic Dynamic*, vol. 15, pp. 472-497, 2011.

[16] L. Kilian and R. J. Vigfusson, "Nonlinearities in the oil price-output relationship," *Macroeconomics Dynamics*, vol. 15, pp. 337-363, 2011.

[17] I. Dimoulkas, L. Herre, D. Khastieva, E. Nycander, M. Amelin and P. Mazidi, "A hybrid model based on symbolic regression and neural networks for electricity load forecasting," in *15th International Conference on the European Energy Market, New York, NY, Institute of Electrical and Electronics Engineers*, 2018, pp. 1-5.

[18] J. Zegklitz and P. Posik, "Benchmarking state-of-the-art symbolic regression algorithms," *Genetic Programming and Evolvable Machines*, vol. 22, pp. 5-33, 2021.

[19] J. Koza, Genetic Programming, Cambridge, MA: MIT Press, 1998.

[20] Y. Jin, W. Fu, J. Kang, J. Guo and J. Guo, "Bayesian symbolic regression," 2019. [Online]. Available: https://arxiv.org/abs/1910.08892.

[21] K. Drachal, "Forecasting spot oil price in a dynamic model averaging framework - Have the determinants changed over time?," *Energy Economics*, vol. 60, pp. 35-46, 2016.

[22] G. Koop and D. Korobilis, "Large time-varying parameter VARs," *Journal of Econometrics*, vol. 177, pp. 185-198, 2013.

[23] The World Bank, "Commodities markets," 2021. [Online]. Available: https://www.worldbank.org/en/research/commodity-markets.

[24] EIA, "U.S. Energy Information Administration," 2021. [Online]. Available: https://www.eia.gov .

[25] MSCI, "End of day index data search," 2021. [Online]. Available: https://www.msci.com/end-of-day-data-search .

[26] Stooq, "Quotes," 2021. [Online]. Available: https://stooq.com .

[27] CBOE, "VIX historical price data," 2021. [Online]. Available: https://www.cboe.com/tradable_products/vix/vix_historical_data .

[28] FRED, "Economic data," 2021. [Online]. Available: https://fred.stlouisfed.org .

[29] L. Kilian, "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market," *American Economic Review*, vol. 99, pp. 1053-1069, 2009.

[30] K. Drachal, "Dynamic Model Averaging in economics and finance with fDMA: A package for R," *Signals*, vol. 1, pp. 47-99, 2020.

[31] M. Nicolau and A. Agapitos, "Choosing function sets with better generalisation performance for symbolic regression models," *Genetic Programming and Evolvable Machines*, vol. 22, pp. 73-100, 2021.

[32] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference*, Berlin: Springer, 2002.

[33] M. F. J. Steel, "Model averaging and its use in economics," *Journal of Economic Literature*, pp. 644-719, 2020.

[34] J. H. Stock and M. W. Watson, "Combination forecasts of output growth in a seven-country data set," *Journal of Forecasting,* vol. 23, pp. 405-430, 2004.

[35] T. Stephens, "Genetic Programming in Python, with a scikit-learn inspired API: gplearn," 2021. [Online]. Available: https://github.com/trevorstephens/gplearn .

[36] A. E. Raftery, M. Karny and P. Ettler, "Online prediction under model uncertainty via Dynamic Model Averaging: Application to a cold rolling mill," vol. 52, pp. 52-66, 2010.

[37] L. Onorante and A. E. Raftery, "Dynamic model averaging in large model spaces using dynamic Occam's window," *European Economic Review*, vol. 81, pp. 2-14, 2016.

[38] G. Koop and D. Korobilis, "Forecasting inflation using Dynamic Model Averaging," *International Economic Review,* vol. 53, pp. 867-886, 2012.

[39] J. Friedman, T. Hastie and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, pp. 1-22, 2010.

[40] T. Hastie and B. Efron, "lars: least angle regression, lasso and forward stagewise," 2013. [Online]. Available: https://CRAN.R-project.org/package=lars .

[41] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, pp. 679-688, 2006.

[42] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, vol. 13, pp. 253-263, 1995.

[43] D. Harvey, S. Leybourne and P. Newbold, "Testing the equality of prediction mean squared errors," *International Journal of Forecasting*, vol. 13, pp. 281-291, 1997.

[44] R. Giacomini and B. Rossi, "Forecast comparisons in unstable environments," *Journal of Applied Econometrics*, vol. 25, pp. 595-620, 2010.

[45] A. Jordan and F. Krueger, "murphydiagram: Murphy diagrams for forecast comparisons," 2019. [Online]. Available: https://CRAN.R-project.org/package=murphydiagram .

[46] P. R. Hansen, A. Lunde and J. M. Nason, "The model confidence set," *Econometrica*, vol. 79, pp. 453-497, 2011.

[47] M. Bernardi and L. Catania, "The model confidence set package for R," *International Journal of Computational Economics and Econometrics*, vol. 8, pp. 144-158, 2018.

[48] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, 2018. [Online]. Available: https://www.R-project.org .

[49] G. Van Rossum and F. L. Drake Jr, Python Reference Manual, Amsterdam: Centrum voor Wiskunde en Informatica, 1995.

[50] C. R. Harris, K. J. Millman, S. J. Van Der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. Van Kerkwijk, M. Brett, A. Haldane, J. Fernandez Del Rio, M. Wiebe, P. Peterson, P. Gerard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, pp. 357-362, 2020.

[51] The pandas Development Team, "pandas-dev/pandas: Pandas," 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3509134 .

[52] W. McKinney, "Data structures for statistical computing in Python," in *Proceedings of the 9th Python in Science Conference 445*, Austin, TX , SciPy, 2010, pp. 56-61.

[53] Y. Jin, "A Bayesian MCMC based symbolic regression algorithm," 2021. [Online]. Available: https://github.com/ying531/MCMC-SymReg .

[54] R. B. Gramacy, "monomvn: Estimation for MVN and Student-t data with monotone missingness," 2019. [Online]. Available: https://CRAN.R-project.org/package=monomvn .

[55] R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting: The forecast package for R," *Journal of Statistical Software*, vol. 26, pp. 1-22, 2008.

[56] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, Melbourne: OTexts, 2021.