A Work Project, presented as part of the requirements for the Award of a Master Degree in Finance from the NOVA – School of Business and Economics.

A FACTOR AUGMENTED VECTOR AUTOREGRESSIVE MODEL AND A STACKED DE-NOISING AUTO-ENCODERS FORECAST COMBINATION TO PREDICT THE PRICE OF OIL.

Parravicini Giovanni, 32290

A Project carried out on the Master in Finance Program, under the supervision of:

Professor Franco Francesco

Professor Pedro Valls Pereira

January 2019

ABSTRACT

The following dissertation aims to show the benefits of a forecast combination between an econometric and a deep learning approach. On one side, a Factor Augmented Vector Autoregressive Model (FAVAR) with naming variables identification following Stock and Watson (2016)[1]; on the other side, a Stacked De-noising Auto-Encoder with Bagging (SDAE-B) following Zhao, Li and Yu (2017)[2] are implemented. From January 2010 to September 2018 Two-hundred-eighty-one monthly series are used to predict the price of the West Texas Intermediate (WTI). The model performance is analysed by Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and Directional Accuracy (DA). The combination benefits from both SDAE-B's high accuracy and FAVAR's interpretation features through impulse response functions (IRFs) and forecast error variance decomposition (FEVD).

---

[1] Stock, J.H. and Watson, M.W., 2016. Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics* (Vol. 2, pp. 415-525). Elsevier.
[2] Zhao, Y., Li, J. and Yu, L., 2017. A deep learning ensemble approach for crude oil price forecasting. *Energy Economics*, *66*, pp.9-16.

CHAPTER 1

Chapter 1: Introduction

The influential statistician George E. P. Box warned: “All models are wrong, but some are useful". The author believes that in forecasting exercises perfection doesn’t exist, generally speaking, due to complexity, in our world. Although empirical evidence proven human behaviours to be flawed, or in other words irrational, Economics courageously tries to explain them. Such process is all but easy, and Economists, although improving, are far from a unanimous model that describe accurately the laws of supply and demand disruptions through time.

Forecasting is not a recent science. Its roots stem from our innate ability that one exploits unconsciously when thinking about the future. Luckily, the science has evolved enough to take a certain distance from who claim to predict the future by reading the hand’s palm. The idea of modelling nature and social’s behaviours comes later with new advances: Mathematics and more recently Statistics and Economics. In this sense, improvements, as for describing the law of nature as deterministic, have to wait more than a millennium to appear. Adam Smith (1776)[3] defined “Political Economics” as “an inquiry into the nature and causes of the wealth of nations” Before it was renamed “Economics” by Alfred Marshall in the 19th century.

Technology has a clear feed-forward impact in the innovation process. The more the technological advancement, the higher the rate of subsequent innovations. Big steps have been taken in the past 100 years since Economists tried to describe our society by given linear equations. An example of such advances, strictly connected to both the models I will present in this paper, are Big Data or as Economists tends to define it: “Data rich environments”. As mentioned before, technology causes technology growth, and indeed a clear exponential pattern can be traced out on the amount of data available every year as pointed out in the recent report by McKinsey. The quality and quantity of

[3] Smith, A., 1817. *An Inquiry into the Nature and Causes of the Wealth of Nations* (Vol. 2).

information have opened room for new techniques and computational power that previously were though inaccessible due to hardware or software constraints.

The paper combines two innovative approaches that make use of Big Data in the compelling exercise of forecasting the oil price. On one side the author proposes an Econometric approach in a Data-rich environment based on State Space modelling literature, precisely a Factor Augmented Vector Autoregressive model (FAVAR) that follows Stock and Watson (2016)[1]. This model has been widely used in Macroeconomics since its introduction by Bernanke, Boivin and Eliasz (2005)[4] mainly for the important advantages in structural analysis; on the other hand, the author evaluates a novel Deep Learning ensemble approach called Stacked De-noising Auto-encoders with Bagging (SDAE-B) proposed for the first time by Zhao, Li, Yu (2017)[2] , who were able to show the model's superior out of sample forecasting accuracy compared to other frequently used Big Data approaches. This paper follows the procedures of the two mentioned papers unless otherwise stated.

The eventual combination would produce a model that exploits both structural analysis interpretation properties, addressing policy makers' concerns, and exceptional forecasting accuracy deriving from the pattern recognition noise-filtering of the auto-encoders. Both models, in a sense, complement each other. The FAVAR reaches a good forecasting accuracy, but it's not the main power of the model. Using the Econometric approach, one can expect to identify the structural shocks, obtain the impulse response functions (IRFs) and the forecast error variance decomposition (FEVD). On the other hand, the Big Data approach, produces high accuracy forecasting but, as its "black box" nickname suggests, it doesn't allow the econometrician to understand the underlying structural dynamics.

Although one Econometrician may prefer a model instead of another, because of personal opinion and beliefs - after all forecasting is more art than a precise science - predictive model combinations

---

[4] Bernanke, B.S., Boivin, J. and Eliasz, P., 2005. Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly journal of economics*, *120*(1), pp.387-422.

have been proven successfully in the literature (Diebold and Pauly, 1987[5]; Makridakis, 1989[6]; De Gooijer and Hyndman, 2006[7]; Pesaran and Pick, 2011[8]). The author uses a simple mean forecast combination. Bayesian averaging is also taken into consideration, but shown only if it suggests statistical significant improvements.

The oil price has been covered extensively recently by the media due to geopolitical tensions and price instability; but its volatility has been central to macro-economists since the 70s when the OPEC and the Middle East wars revolutionized the industry landscape. As frequently underlined in the literature, although the 80s were a relatively calm period, excluding the Iraqi attacks to Kuwait, since the twentieth century began, the oil price literally ranged from $30/barrel to $140/barrel causing not infrequent socio-economic consequences for policy makers. For an in depth historical digression see Baumeister and Kilian (2016)[9], who exhaustively analyse the past 40 years of oil price data. The past unmanageable volatility gave birth to an increasing number of documents and papers attempting to reproduce supply and demand's dynamics. After the recent amount of scientific literature production is taken into consideration, one may argue that an unanimous consensus is far from being reached. Nevertheless, important results have been achieved, for instance, the fact that supply explains significantly less variance of oil prices than demand at quarterly and annual data. This is somewhat counterintuitive given the focus of media on global supply.

Although the aim of the following dissertation isn't a deep dive into the energy market, some words regarding the importance of oil price are needed. Oil represents, followed by coffee, natural gas and gold, the most traded commodity in the financial markets (OECD, 2018). Crude Oil products assume different names according to its sweet-bitter or light-heavy composition. In US the production has been recently disrupted by the introduction of "Shale Oil" which revolutionized the timing of

---

[5] Diebold, F.X. and Pauly, P., 1987. Structural change and the combination of forecasts. *Journal of Forecasting*, *6*(1), pp.21-40.
[6] Makridakis, S., 1989. Why combining works?. International Journal of Forecasting, 5(4), pp.601-603.
[7] De Gooijer, J.G. and Hyndman, R.J., 2006. 25 years of time series forecasting. *International journal of forecasting*, *22*(3), pp.443-473.
[8] Pesaran, M.H. and Pick, A., 2011. Forecast combination across estimation windows. *Journal of Business & Economic Statistics*, *29*(2), pp.307-318.
[9] Baumeister, C. and Kilian, L., 2016. Forty years of oil price fluctuations: Why the price of oil may still surprise us. *Journal of Economic Perspectives*, *30*(1), pp.139-60.

extraction. No matter what the commodity is called or where it is produced, it plays an important role for politic relationships and economies since the industrial revolution (Zou, Zhao, Zhang and Ziong, 2016)[10]. Even if developed economies are looking into renewable energies to meet the climate goals established by the global summit held in Paris in 2017, for many oil producer developing countries the switch may last longer due to their economic dependency. As a matter of fact there is a long list of countries which massively depends on Oil revenue: Kuwait, Libya, Saudi Arabia, Iraq, Angola, Oman, Azerbaijan, Venezuela, Chad, Brunei to name a few (World Bank, 2018). One may expect an increase in diversification from oil price exposure given the past decades instability, instead Ross (1999)[11], by reviewing the literature and analysing the recent data, reveals a heterogeneity of measures across regions, underlying for instance a failure to act from North and sub-Saharan Africa oil producer countries versus non-oil producers. Venezuela's dramatic condition, as of mid-2018, adds one more example to the black-list of the oil-dependent countries that by failing to act and innovate, projected the country in currency corrections and hyperinflationary environments, ultimately culminating in bankruptcies.

The Directional Accuracy (DA), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are compared in order to access forecasting accuracy for both models. These measures reflect the most popular performance measurement in the recent literature from Diebold and Mariano (2002)[12], Mostafa and El Masry (2016)[13], Yu et al (2016)[14] and Zhao, Li, Yu (2017)[2].

---

[10] Zou, C., Zhao, Q., Zhang, G. and Xiong, B., 2016. Energy revolution: From a fossil energy era to a new energy era. *Natural Gas Industry B*, *3*(1), pp.1-11.

[11] Ross, M.L., 1999. The political economy of the resource curse. *World politics*, *51*(2), pp.297-322.

[12] Diebold, F.X. and Mariano, R.S., 2002. Comparing predictive accuracy. *Journal of Business & economic statistics*, *20*(1), pp.134-144.

[13] Mostafa, M.M. and El-Masry, A.A., 2016. Oil price forecasting using gene expression programming and artificial neural networks. *Economic Modelling*, *54*, pp.40-53.

[14] Yu, L., Dai, W. and Tang, L., 2016. A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting. *Engineering Applications of Artificial Intelligence*, *47*, pp.110-121.

Chapter 2

Chapter 2: Dataset

Despite following previous literature, this paper brings new variations and innovations. Motivated by the empirical evidence on low supply importance for oil price's variation (Juvenal and Petrella, 2011[15], 2015[16]), the model overweight real demand features by including variables concerning the top twenty countries in terms of gross domestic product price purchasing parity (GDP PPP). Additionally, the dataset is oriented toward the US Economy. The reason is that the WTI is produced and consumed more in US than it is exported abroad, hence the price is mostly exposed to domestic fluctuations. Nevertheless, price co-movement is common between oil&gas commodities and therefore the motivation for introducing the countries that would impact global real demand the most. Moreover, after careful literature research and comparison, the dataset is extended to about 281 monthly variables, compared to the 139 quarterly used in Stock and Watson (2016)[1] and the 198 monthly used in Zhao, Li, Yu (2017)[2]. The series are temporarily trimmed on January 2010 and extended to September 2018. Overall, the variables aggregate into 13 dimensions: "Industrial Production", "Employment", "Orders, Inventories and Sales", "Housing activities", "General Prices", "Income", "Productivity and Earnings", "Rates", "Money and Credit", "Exchange Rates", "Activity", "Assets Prices" and finally "Oil related variables". Moreover, when possible, these 13 categories are further split between USA and Global. When such division is available, "USA" refers to variables related to the US economy only, whereas "Global" refers to variables related to the top 20 global economies ranked by gross domestic product adjusted for price purchasing parity. One may argue that more variables don't necessarily improve performance. The statement is valid and nor the FAVAR neither the SDAE-B are exempt from this rule, and indeed both models address the issue: the former by dimensionality reduction and the latter by feature selection algorithm.

---

[15] Juvenal, L. and Petrella, I., 2011. *Speculation in the oil market, Federal Reserve Bank of St* (No. 2001). Louis, Working Paper.
[16] Juvenal, L. and Petrella, I., 2015. Speculation in the oil market. *Journal of Applied Econometrics*, *30*(4), pp.621-649.

Initially, around 380 variables are identified and downloaded from Bloomberg and FRED database according to the literature mentioned in this paper. The identification of the thirteen groups of variables follows more Stock and Watson (2016)[1] implementation rather than Zhao, Li, Yu (2017)[2]. In this step, thirty-six variables are removed due to inconsistencies and missing data; while seventeen variables, where the missing data accounted for not over 10% of the overall number of observation, are linearly interpolated according to the growth of the series. The time coverage of the series goes from January 2010 to September 2018 for a total of 105 data points.

In order to capture real demand, the nominal series are deflated by the CPI core inflation in decimal form re-indexed at the beginning of the observation period as in Stock and Watson (2016)[1]. As the paper aims to reflect real demand, the object then becomes to remove any part of the variable's change that is attributable to price movements, arriving at a real, or inflation adjusted, indicator.

The series are furthermore converted to be covariance-stationary and such transformation is tested by the popular Advanced Dickey-Fuller (ADF) and Phillip-Perron (PP) tests (Dickey and Fuller, 1979[17], Phillip and Perron, 1988[18]). The logic behind the integration is set to a threshold confidence level of 0.05. Specifically, if the ADF's or the PP's p-value are larger than 0.05, a transformation is applied. The test is repeated, and if one of the two tests' critical values are still larger than the threshold, an additional transformation is carried out. This process is computed by further differentiation until all series are covariance stationary. No series required more than two differentiations.

To conclude the pre-processing of the information, as in Stock and Watson (2016)[1] and Zhao, Li, Yu (2017)[2], the dataset is standardized. This step is necessary for both models and in particular for the FAVAR, where one of the first steps is a principal component analysis (PCA), which is the non-parametric approach alternative, suggested instead of the computationally inefficient Gibbs Sampling. There is a debate in the literature on whether one method should be preferred instead of the other, but given that none of the two methods has been proven being superior, the two-step

[17] Dickey, D.A. and Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, *74*(366a), pp.427-431.

[18] Phillips, P.C. and Perron, P., 1988. Testing for a unit root in time series regression. *Biometrika*, *75*(2), pp.335-346.

approach, the PCA, is preferred due to its simplicity (Stock and Watson 2016[1]). In the standardization step the series is demeaned and divided by the standard deviation, guaranteeing a range of variation that on average lies between minus three and plus three. Therefore, each series has mean equal zero and standard deviation equal to one.

The series subject to these transformations are shown in the appendix. The choice of the series has been made to allow both models to bring value to the analysis. Most variables represent the driving forces of oil price and were taken combining multiple datasets including Zagaglia (2010)[19], Naser (2016)[20] and Stock and Watson (2016)[1]. It is worth mentioning that the series for US economy had a greater weight than the other country's economies. This is due to the fact that the West Texas Intermediate is produced and consumed in US and more subject to US economy swings. Nevertheless, a new introduction is carried on within the features. In particular the top twenty countries for GDP PPP are selected among the whole population of countries. The reason, as mentioned early in the paper, is that demand has historically played a more important role than supply in driving oil prices in the long run (Juvenal and Petrella, 2011[15], 2015[16], Kilian and Lütkepohl, 2017[21] and Stock and Watson, 2016[1]). Of these twenty countries, whenever available, the following variables, according to the fourteen dimensions are added: industrial production, employment, general prices, interest rates, exchange rates, asset indicators and oil related variables.

Chapter 2: Factor Augmented Vector Autoregressive Model

The Factor Augmented Vector Auto-regressive (FAVAR) model was proposed the first time by Bernanke, Boivin and Eliaz (2005)[4]. Their scope was to disentangle the monetary shocks. A recent version can be found in Namini (2018)[22]. After this successful paper, the model has been extended to

[19] Zagaglia, P., 2010. Macroeconomic factors and oil futures prices: a data-rich model. *Energy Economics*, *32*(2), pp.409-417.
[20] Naser, H., 2016. Estimating and forecasting the real prices of crude oil: A data rich model using a dynamic model averaging (DMA) approach. *Energy Economics*, *56*, pp.75-87.
[21] Kilian, L. and Lütkepohl, H., 2017. *Structural vector autoregressive analysis*. Cambridge University Press.
[22] Siami-Namini, S., 2018. The Effect of Monetary Policy Shocks on the Real Economy: A FAVAR Approach. Res J Econ 2: 1. *of*, *9*, p.2.

study policy uncertainty, oil prices, pass through inflation effect and fiscal policy (Belke, Osowski, 2017[23], Prüser and Schlösser, 2017[24]; Zagaglia, 2010[19], Lombardi 2012[25], Aastveit, 2013[26], Ratti and Vespignani, 2016[27] and Stock and Watson, 2016[1]; Ribon 2011[28], Conflitti and Luciani, 2017[29]; Roulleau-Pasdeloup, 2011[30]). The amount of literature has grown so massively that more than one literature reviews and in-depth identification procedures have been recently presented: Barhoumi, Darné and Ferrara (2013)[31], Stock and Watson (2016)[1], and Kilian and Lütkepohl (2017)[21]. This paper focus on extensions related to oil prices. Generally speaking the FAVAR is a Dynamic Factor Model (DFM) of which one or more factors are observable. Although still growing, since its creation in Geweke (1977)[32] extensive researches have been made on DFM after Marcellino et al (2000)[33] and Stock and Watson (2002)[34] demystified its characteristics. The model is mainly used in Economics and has grown in popularity because of the increasing amount of data. As pointed out by Stock and Watson (2016)[1], the model performs better in data rich environments than simple Vector Auto-regressions (VAR), and the peculiarity, is that most of the useful feature for structural analysis of VARs can be extended to FAVARs.

Initially DFMs were used for monitoring economic activity. Nowadays they are also used for now-casting and forecasting. These three functions reflect the main ability of the model: summarizing large amount of data in few factors. These factors can be estimated using both a parametric (for instance Gibbs sampling) and a non-parametric two steps approach (Principal Component Analysis).

---

[23] Belke, A. and Osowski, T., 2017. *International effects of euro area versus US policy uncertainty: A FAVAR approach* (No. 689). Ruhr Economic Papers.

[24] Prüser, J. and Schlösser, A., 2017. *The effects of economic policy uncertainty on European economies: Evidence from a TVP-FAVAR* (No. 708). Ruhr Economic Papers.

[25] Lombardi, M.J., Osbat, C. and Schnatz, B., 2012. Global commodity cycles and linkages: a FAVAR approach. *Empirical Economics*, *43*(2), pp.651-670.

[26] Aastveit, K.A., Natvik, G.J.J. and Sola, S., 2013. Economic uncertainty and the effectiveness of monetary policy.

[27] Ratti, R.A. and Vespignani, J.L., 2016. Oil prices and global factor macroeconomic variables. *Energy Economics*, *59*, pp.198-212.

[28] Ribon, S., 2011. Augmented, V.A.R. Research Department Bank of Israel.

[29] Conflitti, C. and Luciani, M., 2017. Oil price pass-through into core inflation.

[30] Roulleau-Pasdeloup, J. and Doz, C., 2011. *The dynamic effects of fiscal policy: a FAVAR approach* (No. dumas-00650820).

[31] Barhoumi, K., Darné, O. and Ferrara, L., 2013. Testing the number of factors: An empirical assessment for a forecasting purpose. *Oxford Bulletin of Economics and Statistics*, *75*(1), pp.64-79.

[32] Geweke, J., 1977. The dynamic factor analysis of economic time series. *Latent variables in socio-economic models*.

[33] Marcellino, M., Stock, J.H. and Watson, M.W., 2000. A dynamic factor analysis of the EMU. *manuscript, http://www. igier. uni-bocconi. it/whos. php*.

[34] Stock, J.H. and Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, *97*(460), pp.1167-1179.

The Principal Component Analysis (or PCA) estimates the factors more efficiently than other methods (in terms of computational power required) obtaining non-inferior results and therefore is more frequent in the literature (Marcellino, 2017)[35]. The implementation is described later on, but for the moment, it is enough to think about the factors as the orthogonal unit eigenvectors that minimize the squared distances of the multi-dimensional matrix representing the dataset.

The process of structural analysis works in FAVARs as well as in VARs, although some identification supplements are required. Kilian and Lütkepohl (2017)[21], Gallegati et al (2016)[36], Stock and Watson (2016)[1] debate extensively regarding these properties. In particular, depending on the objective of the study and the identification procedure, two or three additional restrictions are needed. The above-mentioned papers describe exhaustively the variations for all the methodologies. Another important step is the selection of the number of factors. For indexing purpose, Stock and Watson (2016)[1], shows that it is better to choose an additional factor, than incurring in the so called omitted variable bias. For forecasting and now-casting, also due to overfitting issues, the choice is more tedious and can be conducted graphically through the scree plot or analytically through the information criteria. Even if one may use the same information criteria depicted in VARs' lag selection process, Bai and Ng (2002)[37] and Amenguel and Watson (2007)[38] developed an ad hoc test that is frequently used in the literature suggested by Stock and Watson (2016)[1].

Stock and Watson (2016)[1] proposes a 207 variables quarterly data from 1984Q1 to 2014Q4 of which only 139 variables are used to estimate the factors. All the variables are transformed to be integrated of order zero and de-trended. After analysing the IC criteria, the number of factors is set to 8 because, as they say, "it is important that the factor innovations span the space of the structural shocks and the higher factors capture variation". Although the series are available from 1959Q1, due to the parameter instability found in the data the model is produced as mentioned before since 1984Q1. Approaches

---

[35] Marcellino, M., 2017. An Introduction to Factor Modelling.
[36] Gallegati, M., Ramsey, J.B. and Semmler, W., 2016. AE-FSI. *Dynamic Modeling, Empirical Macroeconomics, and Finance: Essays in Honor of Willi Semmler*, p.195.
[37] Bai, J. and Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), pp.191-221.
[38] Amengual, D. and Watson, M.W., 2007. Consistent estimation of the number of dynamic factors in a large N and T panel. *Journal of Business & Economic Statistics*, *25*(1), pp.91-96.

such as in Mumtaz, Zabczyk and Ellis (2011)[39], Eickmeier, Lemke and Marcellino (2015)[40] solves the issue by building a time varying FAVAR (TV-FAVAR).

The dynamic factor models (DFMs) are expressed either in static or dynamic form. Hereafter the static form is used. The state space model representation of a static (stacked) FAVAR is as follow:

$$X_t = \Lambda F_t + e_t \tag{1}$$

$$F_t = \Phi(\mathrm{L}) F_t + G\eta_t \tag{2}$$

Where the observed equation (1) equals the $Nx1$ vector $X_t$of known time series with the sum of the common component ΛF_t and the idiosyncratic disturbances e_t. If this latter element is uncorrelated with the factors innovations at all leads and lags then the DFM is called exact DFM. Generally speaking this is a strong assumption and e_t is modelled as an autoregressive process as in equation (3). Each v_{it}in this case is i.i.d. and the model is called non-exact DFM.

$$e_{it} = \delta_i(L)e_{it-1} + v_{it} \tag{3}$$

The common component is the dot product of the matrix $\Lambda = (\lambda_0, \lambda_1 \ldots, \lambda_p)$ where $\lambda_q$is Nxq and the matrix $F_t = (f_t', f_{t-q}', \ldots, f_{t-p}')'$ where F_t is a $rx1$ of static factors. Equation (2) is the vector autoregressive model expressed in canonical form by stacking the r static factors. The number of static factors r, is by construction typically greater than the number of dynamic factors q, and can be assessed by a combination of a-priori knowledge, visual inspection of a scree plot, and the use of information criteria (IC). The scree plot in particular shows the variance explained by factors, while the IC, such as Bai and Ng (2002)[41] and Amenguel and Watson (2007)[42], provide a penalty function that measures the cost/benefit introduction of an additional factor. Finally, $G = [I_q \ 0_{qx(r-q)}]'$ and η_t, as it was v_t, are assumed to be Gaussian.

[39] Mumtaz, H., Zabczyk, P. and Ellis, C., 2011. What lies beneath? A time-varying FAVAR model for the UK transmission mechanism.
[40] Eickmeier, S., Lemke, W. and Marcellino, M., 2015. Classical time varying factor-augmented vector auto-regressive models—estimation, forecasting and structural analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *178*(3), pp.493-533.
[41] Bai, J. and Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), pp.191-221.
[42] Amengual, D. and Watson, M.W., 2007. Consistent estimation of the number of dynamic factors in a large N and T panel. *Journal of Business & Economic Statistics*, *25*(1), pp.91-96.

Upon standardization of the data, the two-step factor estimation begins with the first set of restrictions on the observation equation (1). As opposed to the Principal Component Normalization used in Bernanke, Boivin and Eliaz (2005)[4], the normalization used in this dissertation, called Named factor normalization (NFN), follows Stock and Watson (2016)[1]. The NFN is used because in line with the scope of this dissertation, that is, to exploit the structural analysis features of the FAVAR. Moreover, this normalization, allows for contemporaneous correlation of the principal components and take the form:

$$\Lambda^{NF} = \begin{bmatrix} I_r \\ \Lambda^{NF}_{r+1:n} \end{bmatrix} \tag{4}$$

Since the VAR form is untouched and unrestricted, two more identification procedures need to be done on the observation equation (1) in order to identify the structural shocks and broadcast them from equation (2) to equation (1) on the individual series of the array X_t. The identification procedures are not all needed for forecasting, but are mainly computed to make sure that the estimated factors are identified and their space spanned is identified. For forecasting purposes, the normalization is enough to identify the space spanned by the factors while the factors themselves are left unidentified. Without these techniques, one would not able to reconcile the Impulse Response Functions and Forecast Error Variance Decomposition with the identified shocks.

After these theoretical assumptions are converted into mathematical restrictions on the observation equation, a principal component (PCA) algorithm runs on all (or on an arbitrary number) of the standardized series inside the array X_t. The PCA is a linear dimensionality reduction that uses a Singular Value Decomposition (SVD) of the data to project it to a lower dimensional space. This technique brings the 281[th] dimensions into some pre-determined number of factors that corresponds to the principal components of the spectral decomposition ranked by variance explained of the dataset from which the factors are extracted.

Once the factors are identified and estimated, by reconstructing equation (2), a selected variable can be forecast in two different ways. The most frequent method is to project the factors and then recall the individual desired variable from the dataset X_t as follow:

$$E[X_{it} | X_t, F_t X_{t-1}, F_{t-1}, \dots] = \alpha_i^F(L)F_t + \delta_i(L)X_{it} \tag{5}$$

Where $\alpha_i^F = \Lambda_i \Phi(L) - \delta_i(L)\Lambda_i$ and F_t equation (5) represents the matrix of the current values of the stacked vector of factors. Given a selected variable to forecast, X_{it}, an alternative approach is to extract the factors from a dataset excluding the variable of interest. This method allows to directly forecast the variable with the factors in a VAR process without recalling it from the observation equation:

$$Y_t = c + \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \dots \Pi_p Y_{t-p} + \varepsilon_t \tag{6}$$

Where $Y_t = (X_i, F_1 \dots F_r)'$, $\Pi_p$are the estimated variable coefficients and ε_t the idiosyncratic shocks. For this VAR process, the h-step ahead forecast is as follow:

$$Y_{T+h|T} = c + \Pi_1 Y_{T+h-1|T} + \dots + \Pi_p Y_{t+h-p|T} \tag{7}$$

One important caveat of the second procedure to forecast is that the element of the matrix Y_t must be orthogonal. In order for the vectors to be orthogonal, the dot product should be approximately zero. Such adjustment can be implemented by the modified Gram Schmidt process (Björck, Å., 1967)[43] as follow:

$$\begin{array}{cccc} u_k^{(1)} = & v_k & - & proj_{u_1}(v_k) \\ \vdots & \vdots & & \vdots \\ u_k^{(k-1)} = & u_k^{(k-2)} & - & proj_{u_{k-1}}(u_k^{(k-2)}) \end{array} \tag{8}$$

where $v_1 \dots v_k$ are a definite, linearly independent set of vectors in space and $u_1 \dots u_k$ are the generated orthogonal set that spans the same k-dimensional subspace. The above process, is the modified version of the Gram-Schmidt process, that generates u not only orthogonal to $u_k^{(k-1)}$, but also against any errors introduced in computation of $u_k^{(k-1)}$.

[43] Björck, Å., 1967. Solving linear least squares problems by Gram-Schmidt orthogonalization. *BIT Numerical Mathematics*, *7*(1), pp.1-21.

Chapter 2: Stacked De-noising Auto-encoders with Bagging

Although the SDAE-B proposed by Zhao, Li, Yu (2017)[2] is fairly recent, it follows a longer trial of Machine and Deep Learning attempts to forecast energy prices (more on this later). Stacked De-noising Auto-encoders with Bagging is the full name of the hybrid used on 198 monthly series to forecast the West Texas Intermediate (WTI) price. A hybrid is a combination of two statistical techniques, in this case, of a deep learning approach (SDAE) and an ensemble learning approach (Bagging). The paper's conclusions, that motivated this research, confirmed statistical effectiveness of this model over other Big Data comparable approaches.

The model comes from the need to describe the price of oil dependency to multi-factors and not only to supply and demand or even purely past values. Namely, the model is exposed to 198 variables grouped in five categories: supply and demand, substitution effect from other source of energy (natural gas, coal, renewable energies…), financial markets, economic growth, technology and irregular events (Zhao, Li, Yu, 2017)[2].

Among the pure machine learning algorithms that tried similar exercises in the literature we find: genetic algorithms (Kaboudan, 2001)[44], neural networks (Moshiri and Foroutan, 2006)[45], support vector machine (Xie et al 2006)[46], semi supervised learning (Greenwood-Nimmo et al, 2013)[47], gene expression programming (Mostafa and el Masry, 2016)[13]. Among the hybrid machine learning algorithms in the literature we find: NARX (Godarzi et al, 2014)[48], ANFIS (Ghaffari and Zare, 2009)[49], combination of NN and GA (Chiroma et al, 2015)[50], IBL (Gabralla et al, 2013)[51], ensemble

[44] Kaboudan, M.A., 2001. Compumetric forecasting of crude oil prices. In *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on* (Vol. 1, pp. 283-287). IEEE.
[45] Moshiri, S. and Foroutan, F., 2006. Forecasting nonlinear crude oil futures prices. *The Energy Journal*, pp.81-95.
[46] Xie, W., Yu, L., Xu, S. and Wang, S., 2006, May. A new method for crude oil price forecasting based on support vector machines. In *International Conference on Computational Science* (pp. 444-451). Springer, Berlin, Heidelberg.
[47] Greenwood-Nimmo, M. and Shin, Y., 2013. Taxation and the asymmetric adjustment of selected retail energy prices in the UK. *Economics Letters*, *121*(3), pp.411-416.
[48] Godarzi, A.A., Amiri, R.M., Talaei, A. and Jamasb, T., 2014. Predicting oil price movements: A dynamic Artificial Neural Network approach. *Energy Policy*, *68*, pp.371-382.
[49] Ghaffari, A. and Zare, S., 2009. A novel algorithm for prediction of crude oil price variation based on soft computing. *Energy Economics*, *31*(4), pp.531-536.
[50] Chiroma, H., Abdulkareem, S. and Herawan, T., 2015. Evolutionary Neural Network model for West Texas Intermediate crude oil price prediction. *Applied Energy*, *142*, pp.266-273.
[51] Gabralla, L.A., Jammazi, R. and Abraham, A., 2013, August. Oil price prediction using ensemble machine learning. In *Computing, Electrical and Electronics Engineering (ICCEEE), 2013 International Conference on* (pp. 674-679). IEEE.

models that first decompose oil price series into components and then combine the forecast by NNs (Xiong et al, 2013[52], Yu et al, 2008[53], Yu et al, 2016[14]).

There are two essential components of a hybrid structure: on one hand the model is fitted on a training sample and the result is used to forecast a test sample. On the other hand, there is an additional technique used for enhancing the forecasting ability of the entire model. Regarding the latter three popular tools are used in the literature above: Bagging, Boosting and Stacking. In particular Bagging decreases the model's variance; Boosting decreases the model's bias; and Stacking increases the predictive force of the classifier.

To understand better the stacked de-noising auto-encoders with Bagging the model is broken down into its components. An auto-encoder is a one hidden layer neural network where its input and output size are equal. The deterministic function connecting the inputs to the output is described in the implementation paragraph. De-noising is the action of cleaning partially corrupted input through AE and, as emphasised in Vincent at al (2010)[54], is important for the extraction of useful features when minimizing the average reconstruction error in the loss function. Multiple levels of DAE are stacked one on another to improve the information reconstruction ability of the classifier (SDAE). In this step, the parameters are tuned by popular algorithms like the gradient descent. Finally, the ensemble component is combined. Bagging, or bootstrapping aggregation (Breiman, 1996)[55] is a powerful tool frequently used in the literature for forecasting that take an average over predictions from all the trained base models.

The main intuition behind using Auto-encoder is that the network learns the latent variables from the raw features while retaining the capability to produce the raw input back from the latent features. Hence, it predicts back the raw features, the input. This is in contrast with the popular neural network

---

[52] Xiong, T., Bao, Y. and Hu, Z., 2013. Beyond one-step-ahead forecasting: evaluation of alternative multi-step-ahead forecasting models for crude oil prices. *Energy Economics*, *40*, pp.405-415.

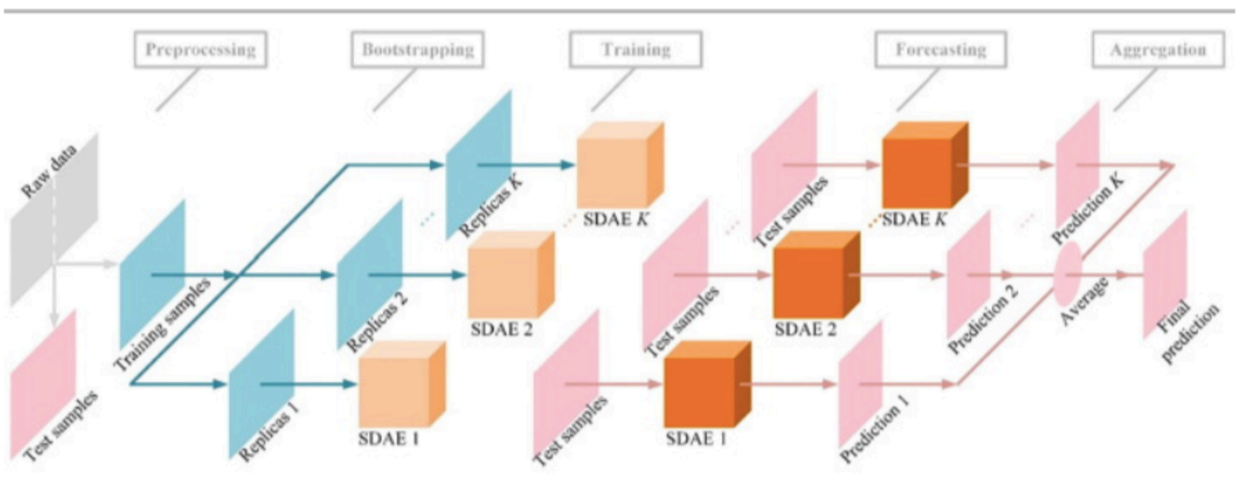
[53] Yu, L., Wang, S. and Lai, K.K., 2008. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics*, *30*(5), pp.2623-2635.

[54] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y. and Manzagol, P.A., 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, *11*(Dec), pp.3371-3408.

[55] Breiman, L., 1996. Bagging predictors. *Machine learning*, *24*(2), pp.123-140.

(ANN) that tries to predict the labels, whereas in our case the model predicts the input only at the output. The loss is based on the performance between raw input and predicted input (at the output layer). By minimizing the loss function the raw input is reconstructed. To this process, if a noise is added to the raw input, the process takes the name of "De-noising". De-noising is sometimes substituted in the literature by "Sparsity" Auto-encoders as in Moussavi-Khalkhali Jamshidi (2016)[56]. Once the core model is structured, Stacking the process is a powerful tool used to increase the predictive power of the model. The idea, as the word suggests, is that the reconstructed input, at the output layer, is passed to a superior Auto-encoder and processed. Each Auto-encoder is structured as the previous, but the corrupted input within each level is different, therefore no level is the same. On top of the last $K + 1$ Auto-encoder, a simple algorithm is added to connect the output to the supervised cost function. For detailed graphical representation Vincent et al (2010)[51] is the standard in the literature. To better understand the mathematical process behind the model, a graphical representation taken from Zhao, Li, Yu (2017)[2] is depicted below:

Figure 1 SDAE-B. Source: Y. Zhao, J. Li, L. Yu (2017)



In practice, the sample is divided into training and test with an 80-20% ratio as in Yu et al, (2008)[50]. To be consistent with the FAVAR, the model is applied over the same sample. In particular the dataset

[56] Moussavi-Khalkhali, A. and Jamshidi, M., 2016, December. Constructing a Deep Regression Model Utilizing Cascaded Sparse Autoencoders and Stochastic Gradient Descent. In Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on (pp. 559-564). IEEE.

is stationary according to ADF and PP tests; It is standardized and the variables are ordered by names following the Stock and Watson (2016)[1] naming variable identification. This latter manipulation doesn't alter the performance of the deep-learning model by any means. The West Texas Intermediate is extracted from the dataset and chosen as the target variable.

For simplicity, a one hidden layer NN is built as main structure of the AE where the input equals the output. The input X maps to output Y following the determinist function:

$$y = f_\theta(x) = \emptyset_f(Wx + b) \tag{9}$$

Where the parameters are W and b, respectively a dxd weight matrix and a bias vector.

The output Y is sub-sequentially mapped to vector z, following the equation:

$$z = g_{\theta'}(y) = \emptyset_g(W'y + b') \tag{10}$$

where the parameters W' and b' are the corresponding dxd weight matrix and bias vector as before.

Each parameter is optimized to minimize the average reconstruction error by following the equation:

$$\theta^*, \theta'^* = \arg min \frac{1}{n} \sum_{i=1}^{n} L(x^i, z^i) = \arg min \frac{1}{n} \sum_{i=1}^{n} L(x^i, g_\theta(f_\theta(x^i))) \tag{11}$$

Where the loss function L would be the traditional squared error function $L(x, z) = |x - z|^2$.

Finally, Bootstrapping aggregation is implemented (Bagging). A set of K training samples is fed to K SDAE models generating K predictions that are aggregate by averaging at the end of the process.

Chapter 3

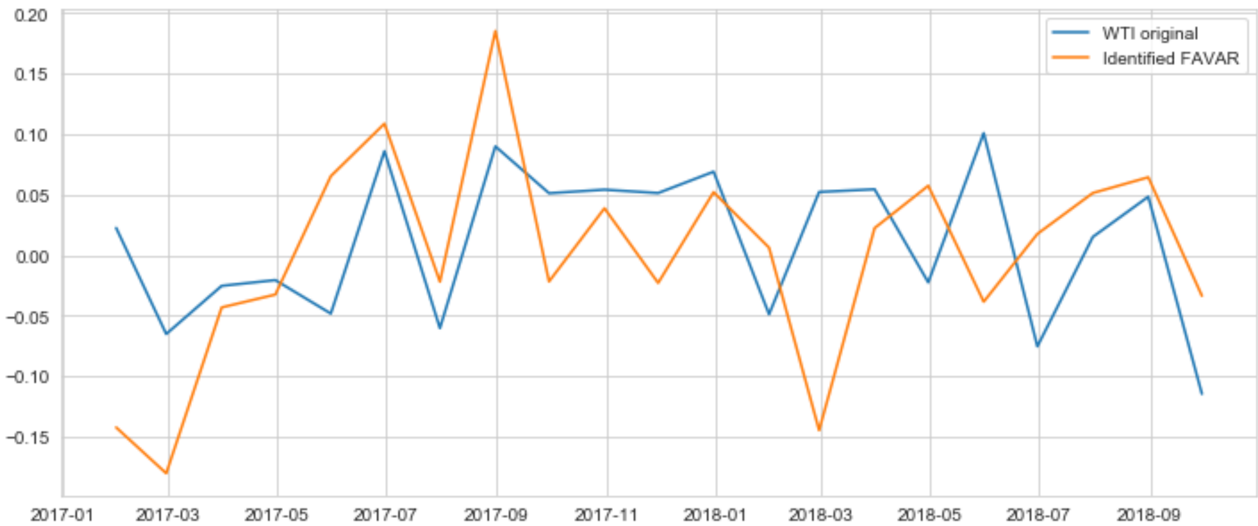
Chapter 3: FAVAR and SDAE-B implementation

The FAVAR is implemented by applying the naming factor normalization. In the Identified FAVAR, five unobserved factors are estimated on 69 selected variables according to Bai and Ng (2002)[57] and Amenguel and Watson (2007)[58] following Stock and Watson (2016)[1] implementation and identification restrictions. The modified Gram-Schmidt algorithm is carried on in order to

[57] Bai, J. and Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), pp.191-221.
[58] Amengual, D. and Watson, M.W., 2007. Consistent estimation of the number of dynamic factors in a large N and T panel. *Journal of Business & Economic Statistics*, *25*(1), pp.91-96.

orthogonalise the factors to the West Texas Intermediate by rotating their space around the latter variable (hence, without changing its elements). After this transformation, the factors are once again controlled for stationarity according to the popular Advanced Dickey-Fuller (ADF) and Phillip-Perron (PP) tests (Dickey and Fuller, 1979[59], Phillip and Perron, 1988[60]). The construction of the Identified FAVAR continued with the implementation of equation (2) as a VAR on 6 variables and eight lags. The number of lags has been assigned by Akaike Information Criterion. The performance of the fitted model on the test sample is shown in Figure 2.

Figure 2 Identified FAVAR forecast performance visualization

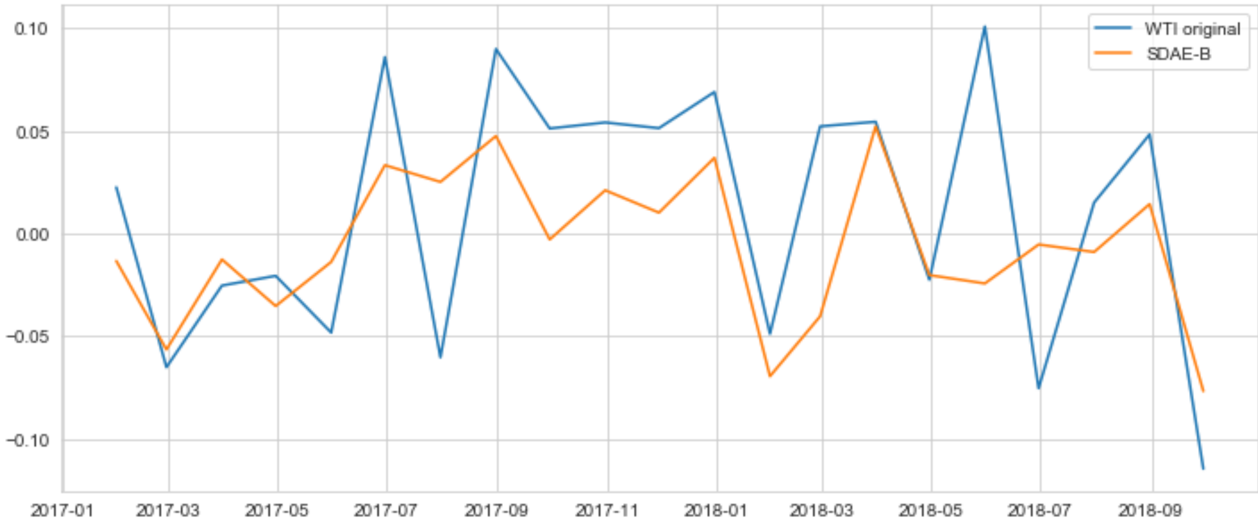


The implementation of the popular Deep Neural Network (DNN) requires two steps, as it is a hybrid combination itself between two different methods: the Stacked De-noising Auto-encoders and the Bagging. 80% of the observations from 2010-01-31 to 2016-12-31 are extracted as training sample. As the bootstrapping aggregation process requires, during the first phase k sets of replicas of the training sample are reproduced. On each k set a SDAE is fitted. The SDAE is literally obtained by stacking several De-noising Auto-encoders. The hidden layer of the first DAE at layer one becomes the input of the DAE at the next layer and so on for k times. The first layer DAE gets as input the input of the SDAE, and the hidden layer of the last DAE represents the output. All this process is

[59] Dickey, D.A. and Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, *74*(366a), pp.427-431.
[60] Phillips, P.C. and Perron, P., 1988. Testing for a unit root in time series regression. *Biometrika*, *75*(2), pp.335-346.

reproduced in Matlab2018 using the function SDAEB() available in the Machine Learning package. The forecast performance of the algorithm on the test sample is reproduced in Figure 3 below.

Figure 3 SDAE-B forecast performance visualization

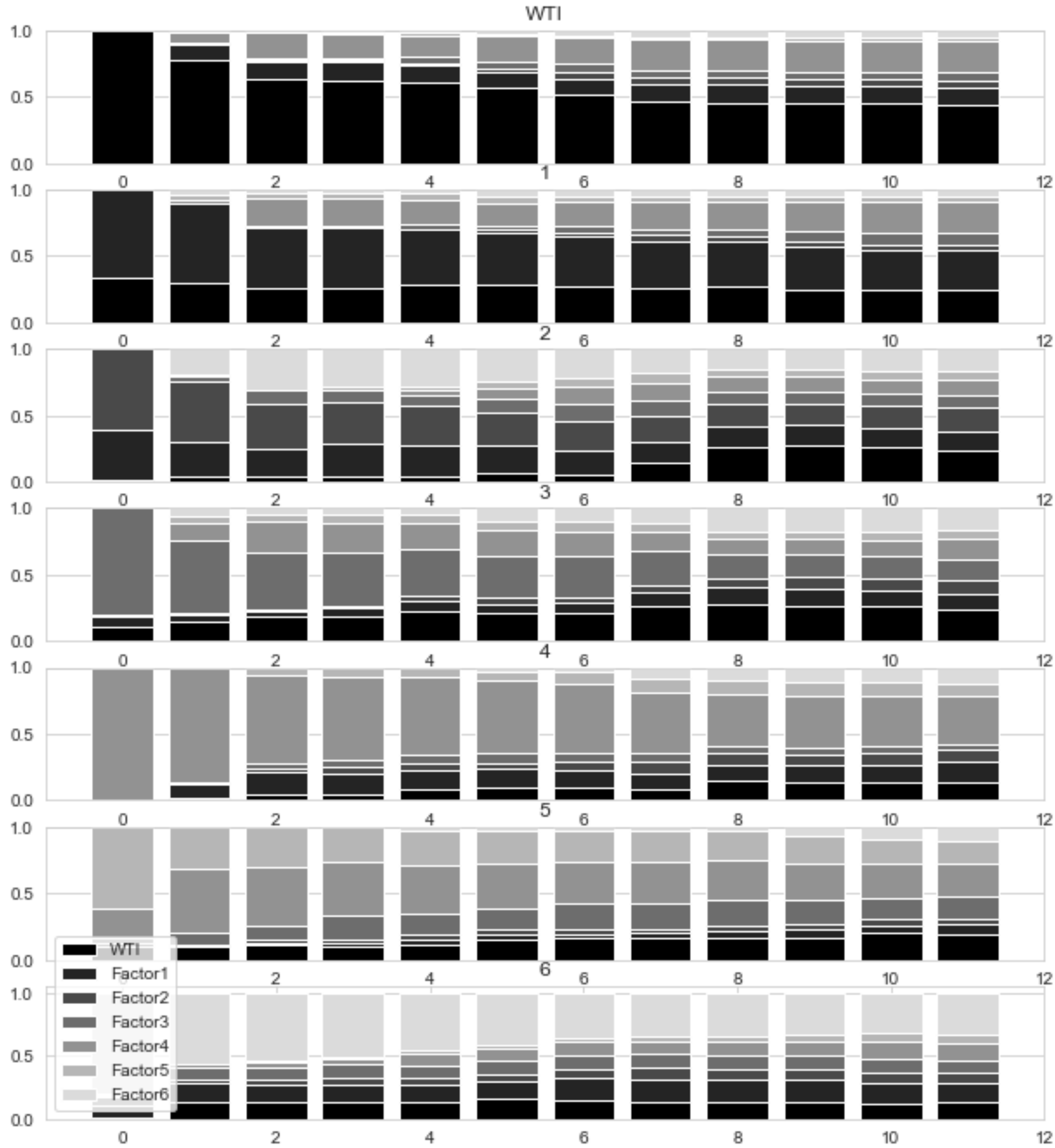


Without digressing too far from the scope of the dissertation, as we are discussing the implementation phase of the models, in Figure 4, is reported the FEVD from the Identified FAVAR first equation relative to the WTI. For policy making such representation may be relevant as a tool of structural analysis because shows the contribute of the WTI series to the overall variance of the model at 2, 4, 6, 8, 10 and 12 months. In other words, it shows the amount of information that the WTI contributes to the other variables in the auto-regression. Exogenous shocks to the WTI explains small forecast error variance of all factors till 12 months. Given that the factors are extracted from a dataset representing mostly the US economy, the results in Figure 4 are relevant when addressing policy issues.

Chapter 3: Performance Evaluation

The Directional Accuracy (DA), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are compared in order to access forecasting accuracy for both models. These measures reflect the most popular performance measurement in the recent literature from Diebold and

Figure 4 Forecast Error Variance Decomposition of the Identified FAVAR



Mariano (2002)[61], Mostafa and El Masry (2016)[62], Yu et al (2016)[63] and Zhao, Li, Yu (2017)[2]. The measurement equations are proposed in equation (12), (13) and (14). where $a_t = 1$ if $(y_{t+1} - y_t)(\hat{y}_{t+1} - y_t) \geq 0$ or $a_t = 0$ otherwise and N is the size of the prediction. The forecast performance between the Identified FAVAR and the SDAE-B are summarized in the table 2. The table shows, as

---

[61] Diebold, F.X. and Mariano, R.S., 2002. Comparing predictive accuracy. *Journal of Business & economic statistics*, *20*(1), pp.134-144.

[62] Mostafa, M.M. and El-Masry, A.A., 2016. Oil price forecasting using gene expression programming and artificial neural networks. *Economic Modelling*, *54*, pp.40-53.

[63] Yu, L., Dai, W. and Tang, L., 2016. A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting. *Engineering Applications of Artificial Intelligence*, *47*, pp.110-121.

expected, the superior accuracy of the SDAE-B compared to the Identified FAVAR. The DA is equal for both models, although this should be tested on a greater sample size. As a matter of fact, the sample is only 20% of the 105 observations.

$$DA = 1/N \sum_{t=1}^{N} a(t) * 100pct \tag{12}$$

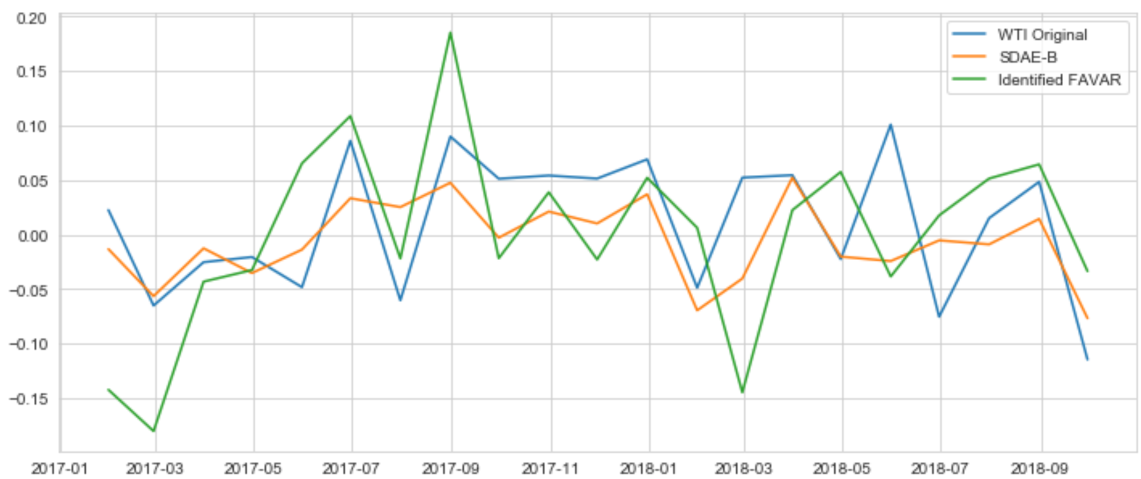
$$MAPE = 1/N \sum_{t=1}^{N} \left| \frac{y_t - \hat{Y}_t}{y_t} \right| \tag{13}$$

$$RMSE = \sqrt{1/N \sum_{t=1}^{N} (y_t - \hat{y}_t)^2} \tag{14}$$

Table 1 Forecast performance of Identified FAVAR and SDAE-B

	RMSE	DA	MAPE
Identified FAVAR	0.087	0.714	1.581
SDAE-B	0.051	0.714	0.971

Figure 5 Identified FAVAR vs SDAE-B forecast performance visualization



Chapter 3: Forecast combination

The next exercise is to combine the two techniques. In doing so, Bates and Granger (1969)[64] is the standard discussion in the literature. Fancy combination has not yet proven superior results compared

[64] Bates, J. M. and Granger, C. W. (1969). The combination of forecasts. Or, pages 451–468.

to simple averages in out of sample forecasting. Two combination methods are proposed. The first represents the naïve simple average forecast and the second is a weighted forecast based on the Root Mean Squared Error (RMSE). Although the first method gives 50% weight to both techniques, the second method ensure more significance to the SDAE-B assigning around 65% of the weight. The visual result is shown in figure 6 and the performances are measured and summarized in table 3.

Figure 6 Simple Average and RMSE Weighted Average forecast combination

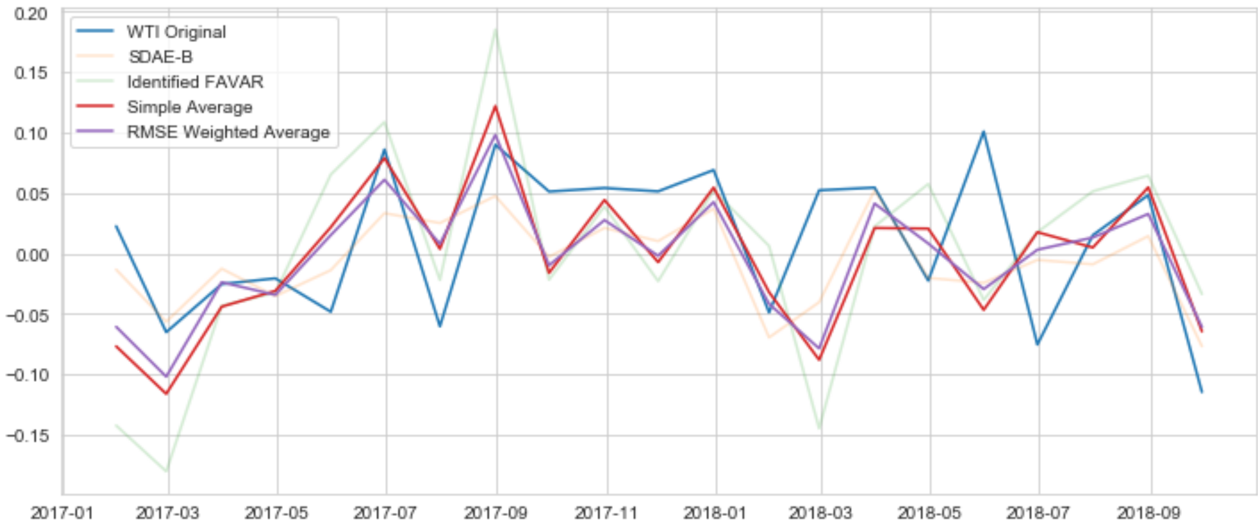


Figure 6 shows the original series of the West Texas Intermediate (blue line) and the two models in light orange and light green. The Simple Average and the RMSE Weighted Average combination are in red and violet respectively. The benefit of the model combinations pop up immediately even with the simplest of the averages. This once again confirms Bates and Granger (1969)[64] who suggest that such result could imply that combining more than just two techniques may also be beneficial.

Table 2 Simple Average and RMSE Weighted Average forecast combination

	RMSE	DA	MAPE
Identified FAVAR	0.087	0.714	1.581
SDAE-B	0.051	0.714	0.971
Simple Average	0.064	0.714	1.035
RMSE Weighted Avg	0.058	0.714	0.877

Table 3 confirms what the visualization suggests and further shows the performance quantitatively. The MAPE from the RMSE Weighted Average combination is even superior to the best model. The DA as expected is constant, given that both original models achieve the same score.

Chapter 4

Chapter 4: Conclusions

The following dissertation shows the benefit of a forecast combination between an econometric and a deep neural network approach. The two models implemented are a Factor Augmented Vector Autoregressive Model and a Stacked De-noising Auto-encoder with Bagging, the fit performance on the test sample is represented in Figure 2 and 3 respectively. Both techniques are fairly recent in the literature and represent important achievements in forecasting. On one side, the FAVAR exploits structural analysis features such as Impulse Response Functions and Forecast Error Variance Decomposition; on the other side the SDAE-B is able to forecast accurately. One example of structural analysis visualization is depicted in Figure 4. Both models achieve a Directional Accuracy of 71.4%. The SDAE-B is superior in terms of Root Mean Squared Error and Mean Absolute Percentage Error compared to the FAVAR. The two forecast combinations proposed are the Simple Average and the RMSE Weighted Average where the model with the highest RMSE is penalized. In this latter, 65% of the relative forecast importance is assigned to the SDAE-B. The weighted combination shown in Figure 6 is the best combination among the two, achieving a RMSE of 0.058, DA of 0.714 and a MAPE superior also to the SDAE-B at about 0.877. Further challenges and interesting discussions may arise by comparing the forecast performance of an unidentified versus an identified FAVAR or even analysing different identification and their implications for forecasting accuracy. Clearly, extending the sample size would only be beneficial in confirming the results obtained in this document; finally, as proposed by Bates and Granger (1969)[64], it would be interesting to combine more than two models and try alternatives weights.

BIBLIOGRAPHY

Aastveit, K.A., Natvik, G.J.J. and Sola, S., 2013. Economic uncertainty and the effectiveness of monetary policy.

Al-Mudhaf, A. and Goodwin, T.H., 1993. Oil shocks and oil stocks: evidence from the 1970s. *Applied Economics*, *25*(2), pp.181-190.

Alexeev, M. and Conrad, R., 2009. The elusive curse of oil. *The Review of Economics and Statistics*, *91*(3), pp.586-598.

Amengual, D. and Watson, M.W., 2007. Consistent estimation of the number of dynamic factors in a large N and T panel. *Journal of Business & Economic Statistics*, *25*(1), pp.91-96.

Bai, J. and Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), pp.191-221.

Barhoumi, K., Darné, O. and Ferrara, L., 2013. Testing the number of factors: An empirical assessment for a forecasting purpose. *Oxford Bulletin of Economics and Statistics*,*75*(1), pp.64-79.

Bates, J. M. and Granger, C. W. (1969). The combination of forecasts. Or, pages 451–468.

Baumeister, C. and Kilian, L., 2016. Forty years of oil price fluctuations: Why the price of oil may still surprise us. *Journal of Economic Perspectives*, *30*(1), pp.139-60.

Belke, A. and Osowski, T., 2017. *International effects of euro area versus US policy uncertainty: A FAVAR approach* (No. 689). Ruhr Economic Papers.

Bernanke, B.S., Boivin, J. and Eliasz, P., 2005. Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly journal of economics*, *120*(1), pp.387-422.

Björck, Å., 1967. Solving linear least squares problems by Gram-Schmidt orthogonalization. *BIT Numerical Mathematics*, *7*(1), pp.1-21.

Breiman, L., 1996. Bagging predictors. *Machine learning*, *24*(2), pp.123-140.

Chiroma, H., Abdulkareem, S. and Herawan, T., 2015. Evolutionary Neural Network model for West Texas Intermediate crude oil price prediction. *Applied Energy*, *142*, pp.266-273.

Conflitti, C. and Luciani, M., 2017. Oil price pass-through into core inflation.

Costantini, M. and Pappalardo, C., 2008. *Combination of forecast methods using encompassing tests: An algorithm-based procedure* (No. 228). Reihe Ökonomie/Economics Series, Institut für Höhere Studien (IHS).

De Gooijer, J.G. and Hyndman, R.J., 2006. 25 years of time series forecasting. *International journal of forecasting*, *22*(3), pp.443-473.

Diebold, F.X. and Mariano, R.S., 2002. Comparing predictive accuracy. *Journal of Business & economic statistics*, *20*(1), pp.134-144.

Diebold, F.X. and Pauly, P., 1987. Structural change and the combination of forecasts. *Journal of Forecasting*, *6*(1), pp.21-40.

Dickey, D.A. and Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, *74*(366a), pp.427-431.

Eickmeier, S., Lemke, W. and Marcellino, M., 2015. Classical time varying factor-augmented vector auto-regressive models—estimation, forecasting and structural analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *178*(3), pp.493-533.

Gabralla, L.A., Jammazi, R. and Abraham, A., 2013, August. Oil price prediction using ensemble machine learning. In *Computing, Electrical and Electronics Engineering (ICCEEE), 2013 International Conference on* (pp. 674-679). IEEE.

Gallegati, M., Ramsey, J.B. and Semmler, W., 2016. AE-FSI. *Dynamic Modeling, Empirical Macroeconomics, and Finance: Essays in Honor of Willi Semmler*, p.195.

Geweke, J., 1977. The dynamic factor analysis of economic time series. *Latent variables in socio-economic models*.

Ghaffari, A. and Zare, S., 2009. A novel algorithm for prediction of crude oil price variation based on soft computing. *Energy Economics*, *31*(4), pp.531-536.

Godarzi, A.A., Amiri, R.M., Talaei, A. and Jamasb, T., 2014. Predicting oil price movements: A dynamic Artificial Neural Network approach. *Energy Policy*, *68*, pp.371-382.

Greenwood-Nimmo, M. and Shin, Y., 2013. Taxation and the asymmetric adjustment of selected retail energy prices in the UK. *Economics Letters*, *121*(3), pp.411-416.

Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B. and Sethupathy, G., 2016. The age of analytics: Competing in a data-driven world. *McKinsey Global Institute*, *4*.

Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B. and Sethupathy, G., 2016. The age of analytics: Competing in a data-driven world. *McKinsey Global Institute*, *4*.

http://www.oecd.org/publications/international-trade-by-commodity-statistics-22195076.htm

https://www.worldbank.org

Juvenal, L. and Petrella, I., 2011. *Speculation in the oil market, Federal Reserve Bank of St* (No. 2001). Louis, Working Paper.

Juvenal, L. and Petrella, I., 2015. Speculation in the oil market. *Journal of Applied Econometrics*, *30*(4), pp.621-649.

k, L., 2013. Testing the number of factors: An empirical assessment for a forecasting purpose. *Oxford Bulletin of Economics and Statistics*, *75*(1), pp.64-79.

Kaboudan, M.A., 2001. Compumetric forecasting of crude oil prices. In *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on* (Vol. 1, pp. 283-287). IEEE.

Kilian, L. and Lütkepohl, H., 2017. *Structural vector autoregressive analysis*. Cambridge University Press.

Lombardi, M.J., Osbat, C. and Schnatz, B., 2012. Global commodity cycles and linkages: a FAVAR approach. *Empirical Economics*, *43*(2), pp.651-670.

Makridakis, S., 1989. Why combining works?. International Journal of Forecasting, 5(4), pp.601-603.

Marcellino, M., 2017. An Introduction to Factor Modelling.

Marcellino, M., Stock, J.H. and Watson, M.W., 2000. A dynamic factor analysis of the EMU. *manuscript, http://www. igier. uni-bocconi. it/whos. php*.

Mirmirani, S. and Cheng Li, H., 2004. A comparison of VAR and neural networks with genetic algorithm in forecasting price of oil. In *Applications of Artificial Intelligence in Finance and Economics* (pp. 203-223). Emerald Group Publishing Limited.

Moshiri, S. and Foroutan, F., 2006. Forecasting nonlinear crude oil futures prices. *The Energy Journal*, pp.81-95.

Mostafa, M.M. and El-Masry, A.A., 2016. Oil price forecasting using gene expression programming and artificial neural networks. *Economic Modelling*, *54*, pp.40-53.

Moussavi-Khalkhali, A. and Jamshidi, M., 2016, December. Constructing a Deep Regression Model Utilizing Cascaded Sparse Autoencoders and Stochastic Gradient Descent. In Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on (pp. 559-564). IEEE.

Mumtaz, H., Zabczyk, P. and Ellis, C., 2011. What lies beneath? A time-varying FAVAR model for the UK transmission mechanism.

Naser, H., 2016. Estimating and forecasting the real prices of crude oil: A data rich model using a dynamic model averaging (DMA) approach. *Energy Economics*, *56*, pp.75-87.

Pesaran, M.H. and Pick, A., 2011. Forecast combination across estimation windows. *Journal of Business & Economic Statistics*, *29*(2), pp.307-318.

Phillips, P.C. and Perron, P., 1988. Testing for a unit root in time series regression. *Biometrika*, *75*(2), pp.335-346.

Prüser, J. and Schlösser, A., 2017. *The effects of economic policy uncertainty on European economies: Evidence from a TVP-FAVAR* (No. 708). Ruhr Economic Papers.

Ratti, R.A. and Vespignani, J.L., 2016. Oil prices and global factor macroeconomic variables. *Energy Economics*, *59*, pp.198-212.

Ribon, S., Augmented, V.A.R. 2011. Research Department Bank of Israel.

Ross, M.L., 1999. The political economy of the resource curse. *World politics*, *51*(2), pp.297-322.

Roulleau-Pasdeloup, J. and Doz, C., 2011. *The dynamic effects of fiscal policy: a FAVAR approach* (No. dumas-00650820).

Shin, H., Hou, T., Park, K., Park, C.K. and Choi, S., 2013. Prediction of movement direction in crude oil prices based on semi-supervised learning. *Decision Support Systems*, *55*(1), pp.348-358.

Siami-Namini, S., 2018. The Effect of Monetary Policy Shocks on the Real Economy: A FAVAR Approach. Res J Econ 2: 1. *of*, *9*, p.2.

Smith, A., 1817. *An Inquiry into the Nature and Causes of the Wealth of Nations* (Vol. 2). Рипол Классик.

SOOD, V. and BAPNA, I., Factors Affecting Price Discovery In Crude Oil: A Literature Review. *ISSN 2026-691X EDITORIAL BOARD*.

Stock, J.H. and Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, *97*(460), pp.1167-1179.

Stock, J.H. and Watson, M.W., 2016. Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics* (Vol. 2, pp. 415-525). Elsevier.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y. and Manzagol, P.A., 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, *11*(Dec), pp.3371-3408.

Xie, W., Yu, L., Xu, S. and Wang, S., 2006, May. A new method for crude oil price forecasting based on support vector machines. In *International Conference on Computational Science* (pp. 444-451). Springer, Berlin, Heidelberg.

Xiong, T., Bao, Y. and Hu, Z., 2013. Beyond one-step-ahead forecasting: evaluation of alternative multi-step-ahead forecasting models for crude oil prices. *Energy Economics*, *40*, pp.405-415.

Yu, L., Wang, S. and Lai, K.K., 2008. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics*, *30*(5), pp.2623-2635.

Yu, L., Zhao, Y. and Tang, L., 2014. A compressed sensing based AI learning paradigm for crude oil price forecasting. *Energy Economics*, *46*, pp.236-245.

Yu, L., Dai, W. and Tang, L., 2016. A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting. *Engineering Applications of Artificial Intelligence*, *47*, pp.110-121.

Zagaglia, P., 2010. Macroeconomic factors and oil futures prices: a data-rich model. *Energy Economics*, *32*(2), pp.409-417.

Zhao, Y., Li, J. and Yu, L., 2017. A deep learning ensemble approach for crude oil price forecasting. *Energy Economics*, *66*, pp.9-16.

Zou, C., Zhao, Q., Zhang, G. and Xiong, B., 2016. Energy revolution: From a fossil energy era to a new energy era. *Natural Gas Industry B*, *3*(1), pp.1-11.

APPENDIX

Overall, the variables aggregate into 13 dimensions: "Industrial Production", "Employment", "Orders, Inventories and Sales", "Housing activities", "General Prices", "Income", "Productivity and Earnings", "Rates", "Money and Credit", "Exchange Rates", "Activity", "Assets Prices" and finally "Oil related variables". Moreover, when possible, these 13 categories are further split between USA and Global. When such division is available, "USA" refers to variables related to the US economy only, whereas "Global" refers to variables related to the top 20 global economies ranked by gross domestic product adjusted for price purchasing parity.

TABLE APPENDIX: DATA SERIES	
Name	Description
1a Industrial Production, USA	
IPDurGooMat	Industrial Production: Durable Materials SA
IPNDuGooMat	Industrial Production: nondurable Materials
IPDurConsGoo	Industrial Production: Durable Consumer Goods
IPAuto	IP: Automotive products
IPNDurConsGod	Industrial Production: Nondurable Consumer Goods
IPBusEquip	Industrial Production: Business Equipment
CapUtiTot	US Capacity Utilization Special Aggregates Excluding High Tech Total Industry
1b Industrial Production, Global	
IP_CHN	IMF China Industrial Production
IP_IND	Production in Total Manufacturing for India
IP_JPN	Production in Total Manufacturing for Japan
IP_GER	Production in Total Manufacturing for Germany
IP_RUS	Production in Total Manufacturing for Russian Federation
IP_IDN	Production in Total Manufacturing for Indonesia
IP_BRA	Production in Total Manufacturing for Brazil
IP_GBN	UK Industrial Production
IP_FRA	Production in Total Manufacturing for France
IP_MXN	Production of Total Construction in Mexico
IP_ITA	Production in Total Manufacturing for Italy
IP_TUR	Production in Total Manufacturing for Turkey
IP_KOR	Production in Total Manufacturing for Korea
IP_ESP	Production in Total Manufacturing for Spain
IP_CAN	IMF Canada Industrial Production SA
IP_THA	Production in Total Manufacturing for Thailand
2a. Employment, USA	
NFP_MLO	US Employees on Nonfarm Payrolls Natural Resources & Mining SA
NFP_CST	US Employees On Nonfarm Payrolls By Industry Construction SA
NFP_DUR	Employees on Nonfarm payroll by Industry Durable Goods SA

NFP_NDU	Employees on Nonfarm payroll by Industry Nondurable Goods SA
NFP_WSL	US Employees On Nonfarm Payrolls By Industry Wholesale Trade SA
NFP_RET	US Employees On Nonfarm Payrolls By Industry Retail Trade Total SA
NFP_TRW	Employees on Nonfarm payroll by Industry Transportation & Warehousing SA
NFP_INF	US Employees on Nonfarm Payrolls Information SA
NFP_FIN	US Employees On Nonfarm Payrolls By Industry Finance Insurance Real Estate SA
NFP_EDH	US Employees on Nonfarm Payrolls Education & Health Services SA
NFP_LEH	US Employees on Nonfarm Payrolls Leisure & Hospitality SA
NFP_GOS	US Employees on Nonfarm Payrolls State Govt SA
NFP_GOL	US Employees on Nonfarm Payrolls Local Govt SA
NLF_CNIP	US Civilian Noninstitutional Population Total NSA
NLF_CLF	US Employment Civilian Labor Force Total in Labor Force SA Household Survey
NLF_DJB	Not in LF and don`t want a job now
NLF_WJB	US Civilians Not in Labor Force But Currently Want a Job NSA
NLF_WMA	US Unemployed & Discouraged & Margin as # Labor Force & Margin NSA
NLF_WDI	US Number of Discouraged Workers NSA
UN_ERS	US Employment Part Time for Economic Reasons SA
UN_NERS	US Employment Part Time Workers Noneconomic Reasons SA
UN_NTMR	Number Of Unemployed Not On Temporary Layoff SA
UN_JBL	US Unemployment Job Leavers Total SA
UN_5WE	US Unemployment Duration Less Than 5 Weeks SA
UN_15W	US Unemployment Duration Over 15 Weeks SA
UN_27W	US Unemployment Duration 27 Weeks and Over SA
OVR_MNF	US Average Weekly Hours All Employees Manufacturing Overtime Hours SA
OVR_DUR	US Average Weekly Hours All Employees Durables Gds Overtime Hours SA
UNMP_USA	U-3 US Unemployment Rate Total in Labor Force Seasonally Adjusted
2b. Employment, Global	
UNMP_JPN	Japan Unemployment Rate SA
UNMP_DEU	Germany Total Civilian Unemployment
UNMP_RUS	Unemployment Rate Russia
UNMP_BRA	Unemployment Rate Brazil
UNMP_GBR	Unemployment Rate UK
UNMP_FRA	Unemployment Rate France
UNMP_MEX	Unemployment Rate Mexico
UNMP_ITA	Unemployment Rate Italy
UNMP_TUR	Unemployment Rate Turkey
UNMP_KOR	Unemployment Rate South Korea
UNMP_ESP	Unemployment Rate Spain
UNMP_CAN	Unemployment Rate Canada
UNMP_THA	Unemployment Rate Thailand
UNMP_AUS	Unemployment Rate Australia
3. Orders, Inventories and Sales, USA	
PMI_NOR	ISM Manufacturing Report on Business New Orders SA
PMI_PRD	ISM Manufacturing Report on Business Production SA
PMI_EMP	ISM Manufacturing Report on Business Employment SA
PMI_DEL	ISM Manufacturing Report on Business Supplier Deliveries SA
PMI_BIN	ISM Manufacturing Report on Business Inventories NSA

PMI_CIN	ISM Manufacturing Report on Business Customers' Inventories NSA
PMI_PRC	ISM Manufacturing Report on Business Prices Index NSA deflated
PMI_BCK	ISM Manufacturing Report on Business Backlog of Orders NSA
PMI_BX	ISM Manufacturing Report on Business Export Orders SA
PMI_BM	ISM Manufacturing Report on Business Imports SA
SL_TOT	Adjusted Retail & Food Services Sales Total SA
SL_XMV	Adjusted Retail Sales Less Motor Vehicle and Parts Dealers SA
SL_XMG	Adjusted Retail Sales Less Autos and Gas Stations SA
SL_SB	Adjusted Retail Sales Food Services and Drinking Places SA
SL_ECO	E-COMMERCE SALES QUARTERLY
4. Housing Activity, USA	
HstartsNE	US New Privately Owned Housing Units Started Northeast 1 Unit Structure SAAR
HstartsMW	US New Privately Owned Housing Units Started Midwest 1 Unit Structure SAAR
HstartsS	US New Privately Owned Housing Units Started South 1 Unit Structure SAAR
HstartsW	US New Privately Owned Housing Units Started West 1 Unit Structure SAAR
PermitsN	US New Privately Owned Housing Auth by Building Permits Northeast 1 Unit SAAR
PermitsMW	US New Privately Owned Housing Auth by Building Permits Midwest 1 Unit SAAR
PermitsS	US New Privately Owned Housing Auth by Building Permits South 1 Unit SAAR
PermitsW	US New Privately Owned Housing Auth by Building Permits West 1 Units SAAR
5a. General Prices, USA	
PCE_DEF	PCE DEF Index
PPI_FGF	PPI final demand for foods
PPI_ENE	PPI final demand for energy
PPI_LOG	PPI logging
PPI_MNF	PPI total mining utilities and manufacturing
PPI_CON	PPI construction
PPI_TRD	PPI trade
PPI_TRS	PPI transportation and warehousing
CPSFFOOD	Food
CPUPENER	Energy
CPUPENCM	Energy Commodities
CPSHFOCB	Fuel Oil and Other Fuels
CPIQFUOS	Fuel oil
CPIQPKFS	Propane, Kerosene, and Firewood
CPSTMTFL	Motor fuel
CPSTGAS	Gasoline All Types
CPIQGURS	Unleaded Regular Gasoline
CPIQGUMS	Gasoline, Unleaded Midgrade
CPIQGUPS	Gasoline, Unleaded Premium
CPIQOMFS	Other Motor Fuels
CPSHGE	Energy Services
CPIQELS	Electricity
CPIQUPGS	Utility (Piped) Gas Service
CPUPAXFE	All Items Less Food and Energy Rebased at 31/01/10
CPCATOT	Commodities
CPSSTOT	Services
CPCADUR	Durables

CPUPNOND	Nondurables
CPSHTOT	Housing
CPUETOT	Education and Communication
CPSRTOT	Recreation
CPSFTOT	Food and beverages
CPIQALFS	Apparel Less Footwear
CPSHFU	Fuels and Utilities
CPUMTOT	Medical care
CPSTTOT	Transportation
CPIQUPTS	Utilities and Public Transportation
CPSHHHFO	Household Furnishings and Operations
CPUOTOT	Other goods and services
CPI_USA	Consumer price index usa
5b. General Prices, Global	
CPI_CHN	China CPI Total at Constant Price 1978=100
CPI_IND	consumer price index india
CPI_JPN	consumer price index japan
CPI_DEU	consumer price index germany
CPI_RUS	consumer price index russia
CPI_IDN	consumer price index indonesia
CPI_BRA	FGV Brazil IGP-M CPI IPC-M
CPI_GBN	consumer price index uk
CPI_FRA	consumer price index france
CPI_MXN	consumer price index mexico
CPI_ITA	consumer price index ita
CPI_TUR	consumer price index turkey
CPI_KOR	consumer price index south korea
CPI_ESP	consumer price index spain
CPI_CAN	consumer price index canada
CPI_THA	consumer price index thailand
CPI_SAU	consumer price index australia
6. Income, USA	
PI_PI	US Personal Income SAAR Deflated
PI_WSD	Personal Income Wage & Salary Disbursements SAAR Deflated
PI_DIV	Personal Income Personal Dividend Income SA Deflated
PI_INT	Personal Income Personal Interest Income SA Deflated
PI_TPP	Personal Income Transfer Payments to Persons SA Deflated
DISP_INC	US Disposable Personal Income Nominal Dollars SAAR Deflated
PCE$	Personal Consumption Expenditures (current $) Deflated
PCE_DUR	US Personal Consumption Expenditures Durable Goods Nominal Dollars SAAR Deflated
PCE_NDU	US Personal Consumption Expenditures Non Durable Goods Nominal Dollars SAAR Deflated
PCE_SRV	US Personal Consumption Expenditures Services Nominal Dollars SAAR Deflated
SAVING	Personal Savings Deflated
SAV_RATE	US Personal Saving as a % of Disposable Personal Income
DUR%PCE	US Durable Goods Spending as a % PCE Current Dollars SAAR
NDU%PCE	US Nondurable Good Spending as a % PCE Current Dollars SAAR
SRV%PCE	US Service Spending as a % PCE Current Dollars SAAR

7. Productivity and Earnings, USA	
ERP_PRV	Average hourly earnings Deflated
ERP_MLO	Mining and Logging Deflated
ERP_CST	Construction Deflated
ERP_DUR	Manufacturing Durable goods Deflated
ERP_NDU	Manufacturing Non durable goods Deflated
ERP_WSL	Wholesale Trade Deflated
ERP_RET	Retail Trade Deflated
ERP_TRW	Transportation and warehousing Deflated
ERP_INF	Information Deflated
ERP_FIN	Financial Activities Deflated
ERP_PRB	Professional and Business Services Deflated
ERP_EDH	Education and Health Services Deflated
ERP_TTU	Trade, Transportation, and Utilities Deflated
8. Rates, USA	
FFR	US Federal Funds Effective Rate
TB3	IMF US Treasury Bill 3 Month Rate
BAA_G10	US Corporate BAA 10 Year Spread
MRG_G10	US Bloomberg Fannie to Govt Spread 10 Year
TB24_TB6	Treasury Spreads 2 Year - 6 month
G10_TB3	Treasury Spreads 10 Year - 3 month
TED_SPD	Ted Spread
9. Money and Credit, USA	
M1	Federal Reserve United States Money Supply M1 SA Deflated
M2	Federal Reserve United States Money Supply M2 SA Deflated
LOA_LEA	US Commercial Bank Assets Loans & Leases Commercial & Industrial SA Deflated
REV_CRE	Revolving Consumer Credit Owned and Securitized SA Flow Deflated
COS_CRE	Total Consumer Credit Owned and Securitized SA Flow Deflated
COS_AUT	Federal Reserve Consumer Credit Commercial Bank Rate 48 Month New Car Deflated
10. Exchange Rates, Global	
EURUSD	EURUSD Spot Exchange Rate - Price of 1 EUR in USD
USDCNH	USDCNH Spot Exchange Rate - Price of 1 USD in CNH
USDINR	USDINR Spot Exchange Rate - Price of 1 USD in INR
USDJPY	USDJPY Spot Exchange Rate - Price of 1 USD in JPY
USDZAR	USDZAR Spot Exchange Rate - Price of 1 USD in ZAR
USDIDR	USDIDR Spot Exchange Rate - Price of 1 USD in IDR
USDBRL	USDBRL Spot Exchange Rate - Price of 1 USD in BRL
GBPUSD	GBPUSD Spot Exchange Rate - Price of 1 GBP in USD
USDMXN	USDMXN Spot Exchange Rate - Price of 1 USD in MXN
USDTRY	USDTRY Spot Exchange Rate - Price of 1 USD in TRY
USDKRW	USDKRW Spot Exchange Rate - Price of 1 USD in KRW
USDCAD	USDCAD Spot Exchange Rate - Price of 1 USD in CAD
USDTHB	USDTHB Spot Exchange Rate - Price of 1 USD in THB
AUDUSD	AUDUSD Spot Exchange Rate - Price of 1 AUD in USD
USDSAR	USDSAR Spot Exchange Rate - Price of 1 USD in SAR
USDCHF	USDCHF Spot Exchange Rate - Price of 1 USD in CHF
11. Activity, Global	

STE_65	IISI World Total Steel Production Data-Currently 65 countries
STE_EU	World Steel Association Crude Steel Production Data/European Union
KILIAN	Kilian Global Economic Activity Index
WEATHER	Average (°F) in Alabama
12. Asset Prices, Global	
SP100	S&P Global 100 Index
MSCIW	MSCI World Index
DJIA	Dow Jones Industrial Average
NIKKEI	Nikkei 225
FTSE100	FTSE 100 Index
SHANGHAIEXC	Shanghai Stock Exchange Composite Index
SENSEX	S&P BSE SENSEX Index
HANGSENG	Hong Kong Hang Seng Index
IBOVESPA	Ibovespa Brasil Sao Paulo Stock Exchange Index
TSXEXC	S&P/TSX Composite Index
CAC40	CAC 40 Index
DAX	Deutsche Boerse AG German Stock Index DAX
FTSEMIB	FTSE MIB Index
MOEX	MOEX Russia Index
KOSPI	Korea Stock Exchange KOSPI Index
IBEX35	IBEX 35 Index
SETEXC	Stock Exchange of Thailand SET Index
AUSEXC	Australian Stock Exchange All Ordinaries Index
VOL	Chicago Board Options Exchange OEX Volatility Index
SHILLER20	S&P CoreLogic Case-Shiller 20-City Composite Home Price SA Index Deflated
SHILLER10	S&P CoreLogic Case-Shiller 10-City Composite Home Price SA Index Deflated
GOLD	IMF UK USD per Ounce of Gold End of Period Deflated
13. Oil related variables, Global	
GASOLINE	Gasoline All Types
EXP_OPEC	Exports by Selected Countries SA OPEC Deflated
IMP_CIF_OPEC	General Imports of Crude Oil OPEC Total CIF Value in thousand of Dlr NSA Deflated
IMP_CUS_OPEC	General Imports of Crude Oil OPEC Total Customs Value in thousand of Dollars NSA Deflated
IMP_TOT_OPEC	General Imports of Crude Oil OPEC Total Qty in Thousands of Barrels NSA
IMP_TOT_NOPE	General Imports of Crude Oil from Non-OPEC CIF Value in thousand of Dlr NSA Deflated
IMP_CUS_NOPE	General Imports of Crude Oil from Non-OPEC Customs Value in thousand of Dlr NSA Deflated
IMP_TOT_NOPE1	General Imports of Crude Oil from Non-OPEC Qty in Thousands of Barrels NSA
SUPPLY_NOPE	International Crude Oil and Liquid Fuels Supply Non OPEC
NAT_GAS	Natural gas Deflated
WTI	West Texas Intermediate Spot
BRENT	Brent Deflated
WTIBRENT_SPRDF	Bloomberg Fair Value Price/NYMEX WTI Futures minus ICE Brent Futures Month 1
WTIBRENT_SPRDS	Bloomberg Fair Value Price/WTI-Brent Crude Oil Spread Monthly
DOES_OPEC	DOE Monthly OPEC Total Crude Oil (excluding condensates) Supply
DOEC_OPEC_TOT	DOE Monthly OPEC Total Crude Oil Production Capacity
DOEC_OPEC_EXC	DOE Monthly OPEC Total Surplus Crude Oil Production Capacity
BLB_Y_OPEC	Bloomberg Total OPEC Crude Oil Production Output Data
EI_OPEC_PROD	Energy Intelligence Group OPEC Crude Oil Production Data

EI_ROW_Q	Energy Intelligence Group Oil Products Non-OECD Rest of World Demand Data
EI_OECD_Q	Energy Intelligence Group OECD Oil Products Demand Data
EI_OIL_Q	Energy Intelligence Group Oil Product Demand Data
BH_OPEC_RIG	Baker Hughes OPEC Countries Oil And Gas Rotary Rig Count Data
BH_NOPE_RIG	Baker Hughes Non-OPEC Countries Oil And Gas Rotary Rig Count Data
BH_US_RIG	Baker Hughes U.S. Oil And Gas Rotary Rig Count Data
BH_CAN_RIG	Baker Hughes Canadian Oil And Gas Rotary Rig Count Data
BH_LATAM_RIG	Baker Hughes Latin America Oil And Gas Rotary Rig Count Data
BH_MDE_RIG	Baker Hughes Middle East Oil And Gas Rotary Rig Count Data
BH_FEAST_RIG	Baker Hughes Far East Oil And Gas Rotary Rig Count Data
BH_EU_RIG	Baker Hughes Europe Oil And Gas Rotary Rig Count Data
BH_AFR_RIG	Baker Hughes Africa Oil And Gas Rotary Rig Count Data
EIA_AME_STOCK	International Energy Agency OECD Americas Industry Crude Oil Stocks
EIA_EU_STOCK	International Energy Agency OECD Europe Industry Crude Oil Stocks
EIA_PAC_STOCK	International Energy Agency OECD Pacific Industry Crude Oil Stocks
EIA_AME_Y	International Energy Agency Americas Industry Total Product Stocks
EIA_EU_Y	International Energy Agency Europe Industry Total Product Stocks
EIA_PAC_Y	International Energy Agency Pacific Industry Total Product Stocks
DOES_48_Y	DOE Crude Oil Lower 48 States Production Data
GAS_Y	Natural Gas Production Estimates - Lower 48
ST_Y_FRCST	DOE Short Term Outlook Total Crude Oil Production Forecast Monthly
ST_GAS_Y	United States Short Term Energy Outlook Dry Natural Gas Production
ST_COAL_Y	United States Short Term Energy Outlook Coal Production
ST_OIL_P	United States Short Term Energy Outlook Energy Prices Crude Oil Deflated
ST_GAS_P	United States Short Term Energy Outlook Natural Gas Henry Hub Deflated
ST_COAL_P	United States Short Term Energy Outlook Energy Prices Coal Deflated
IMP_PAD1	DOE PSM PADD 1 Imports of Crude Oil
IMP_PAD2	DOE PSM PADD 2 Imports of Crude Oil
IMP_PAD3	DOE PSM PADD 3 Imports of Crude Oil
IMP_PAD4	DOE PSM PADD 4 Imports of Crude Oil
IMP_PAD5	DOE PSM PADD 5 Imports of Crude Oil
EXP_PAD2	DOE US PADD II Exports of Crude Oil
WTI_futures	West Texas Intermediate Futures