



Máster de Ingeniería Industrial

Trabajo de Fin de Máster

*Forecasting Natural Gas Prices*

Author

Paula Albendea López

Director

Elena María Díaz Aguiluz

Madrid

January 2023

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título  
Forecasting Natural Gas Prices  
en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el  
curso académico 2022-2023 es de mi autoría, original e inédito y  
no ha sido presentado con anterioridad a otros efectos. El Proyecto no es plagio  
de otro, ni total ni parcialmente y la información que ha sido tomada  
de otros documentos está debidamente referenciada.



Fdo.: Paula Albendea López

Fecha: 11/01/2023

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO



Fdo.: Elena María Díaz Aguiluz

Fecha: 23/01/2023

# FORECASTING NATURAL GAS PRICES

**Author: Albendea López, Paula.**

Director: Díaz Aguiluz, Elena María

Collaborating Entity: Universidad Pontificia Comillas

## EXECUTIVE SUMMARY

### 1. INTRODUCTION

Natural gas is a hydrocarbon composed of several light gases found in nature. This gas was formed millions of years ago deep beneath the ground from the remains of plants and animals that built up in thick layers on the earth's surface and ocean floors [11].

Nowadays, this gas is used as a fossil energy fuel, and it has become an essential means of energy in today's world. Natural gas is used mainly to generate electricity, to run heating systems, as a fuel for vehicles, and for industrial production. The electric power sector, the industrial sector, the residential sector, the commercial sector, and the transportation sector are the ones that use this fuel the most [13].

Demand for natural gas is increasing due to its price-calorific value ratio. According to the U.S. Energy Information Association (EIA), the United States used about 31.5 quadrillion British thermal units (Btu) in 2020 which makes 34% of U.S. total energy consumption [13].

After a 4% drop in 2020 due to COVID-19 crisis, natural gas demand progressively recovered in 2021 as consumption returned near to its pre-crisis levels in mature markets. An average growth rate of 1.5% of demand per year is expected for the period of 2019-2025. Half of this growth in consumption is

attributable to the Asia Pacific region, due mainly by the raise of gas usage in China and India [31].

On the production side, North America and the Middle East are expected to contribute to half of the net increase in natural gas supply. Russia is the second largest producer and one of the biggest exporters of natural gas. On February 24<sup>th</sup>, 2022, Russia decided to launch an offensive to Ukraine after a long escalation of tension between the two countries. The sanctions imposed on Russia by the European Union and the United States, caused price of gas to rocket in EU territories, where almost 40% of the gas supply comes from Russia [32].



*Figure 1. Natural gas prices in USD/MMBtu from February 24th 2022 to June 27th 2022 [33].*

On top of all the crisis caused by the war in Ukraine, the Freeport LNG terminal in Texas was closed on June 9<sup>th</sup>, 2022, due to a fire. The shutdown was initially scheduled to last three weeks, but the authorities later announced that it would be extended until the end of 2022 [33].

Given all these events, price per megawatt-hour for July contracts in Europe, traded on the Netherlands-based virtual natural gas trading point (TTF), rose to EUR 127.17 (USD 133.49) on June 22<sup>nd</sup>, 2022, up 60% from EUR 79.40

(USD 83.35) on June 8<sup>th</sup>, which was the lowest closing level in the last four months since the start of the Russia-Ukraine war on February 24<sup>th</sup> [33].



*Figure 2. Price in €/MWh for July contracts traded in Dutch TTF Gas Futures (February 21st, 2022 - June 27th, 2022) [34].*

All of these events made the price of natural gas to be one of the topics of interest in the past few months. This paper tries to find a technique that can forecast the price of natural gas in the most accurate way possible.

## 2. METHODOLOGY

The study is divided into 8 chapters. The first chapter is dedicated to studying general information about natural gas, such as its formation, different techniques to extract it from the ground, transform it into an energy fuel and transport it to the end-user. Then, global supply, the biggest producers, storers, and exporters are identified. The same is applied with global demand, studying how this demand is distributed across sectors. Finally, environmental implications of extracting, producing, and burning natural gas are flagged out.

Secondly, a chapter is dedicated to compile the variables that affect natural gas prices and describe how the indicators used in the model are built. These variables were identified in a previous study (*TFG: Modelling the Natural Gas Market*) and more detailed research can be there found.

Before building the ARX model, a benchmark model is constructed to measure the forecasting performance of this ARX model. This benchmark model consists of a first order autoregressive model, in which the price of natural gas for a period ahead is dependent on an independent term (alpha) and the price for natural gas in the current period multiplied by a scaling parameter (beta). The model is noted as AR(1) and defined by the following equation:

$$y_{t+1} = \alpha + \beta_1 y_t + \varepsilon_{t+1} \quad \varepsilon_t \sim (0, \sigma^2)$$

*Equation 1. Benchmark Model [1].*

where  $\varepsilon_{t+1}$  is white noise,  $\alpha$  is a constant, and  $\beta_1$  is the parameter of the model (1).

After estimating the parameters of the model using a regression, the natural gas price forecast for the following month is calculated for each of the months of the period September 2010-January 2022. The forecasting performance of this benchmark model will then be compared to the ARX model.

In the ARX model, the price of natural gas in the following period ( $y_{t+1}$ ) depends on the price of natural gas a period before the prediction ( $y_t$ ), and the year-to-year or month-to-month variation of other variables (*see Chapter 2*) also a month before the prediction ( $\Delta f_{1t} \dots \Delta f_{8t}$ ).

$$y_{t+1} = \alpha + \beta_1 y_t + \beta_2 \Delta f_{1t} + \dots + \beta_8 \Delta f_{7t} + \varepsilon_{t+1} \quad \varepsilon_t \sim (0, \sigma^2)$$

*Equation 2. ARX Model [1].*

The forecasting performance of each model is evaluated by the out-of-sample  $R^2$  statistic, suggested by Campbell and Thompson [4] to compare the  $\hat{y}_{t+1}$  and  $\bar{y}_{t+1}$  forecasts at a 1-month horizon.  $\hat{y}_{t+1}$  is the forecast using the ARX model and  $\bar{y}_{t+1}$  is the forecast based on the benchmark model. This  $R_{OS}^2$  statistic is given by:

$$R_{OS}^2 = 1 - \frac{\sum_{t=t_0-1}^{T-1} (y_{t+1} - \hat{y}_{t+1})^2}{\sum_{t=t_0-1}^{T-1} (y_{t+1} - \bar{y}_{t+1})^2}$$

*Equation 3. R2 out of sample [1].*

where  $y_{t+1}$  is the realized and real value of the price of natural gas,  $t_0$  is the start of the forecasting sample (September 2010) and  $T$  is the end of the sample (December 2021). If  $R_{OS}^2 > 0$ , the predictive model forecasts ( $\hat{y}_{t+1}$ ) outperform the benchmark forecasts ( $\bar{y}_{t+1}$ ).

We then examine whether the results are statistically significant by estimating the MSPE-adjusted statistic developed by Clark and West (2007).

We first define

$$\varepsilon_{t+1} = (y_{t+1} - \bar{y}_{t+1})^2 - [(y_{t+1} - \hat{y}_{t+1})^2 - (\hat{y}_{t+1} - \bar{y}_{t+1})^2]$$

*Equation 4. MSPE-adjusted statistic [1].*

and then regress  $\{\varepsilon_{t+1}\}_{s=t_0}^{T-1}$  on a constant and calculate the t-statistic under the null hypothesis that the constant is zero. The p-value is obtained with a standard normal distribution, for a one-tailed test.

After building a complete ARX model using all the variables mentioned in *TFG: Modelling the Natural Gas Market*, we also build different ARX models including in each of them only some and not all of the nine variables and measure and compare their performance.

Finally, the concept of “machine learning” is described as well as the different techniques within this field to forecast the behavior of any variable. Using Time Series with Long short-term memory (LSTM) Neural Networks, we predict Natural Gas price for the period of September 2010 to January 2022, similarly to what we do with the Autoregressive model, and compare the predictions with the real values obtained at EIA’s website. We use Natural Gas

historical monthly price downloaded from EIA’s website (51) from September 2000 to December 2021. As software frameworks, we use TensorFlow and Keras.

The parameters of the model are summarized in the following table:

• Stochastic gradient descent with backpropagation error	✓
• Test vs. Training Set	20% vs. 80%
• Time-lags	6 months
• Number of hidden layers	4
• Number of neurons per layer	200
• Optimizer	Adam
• Activation Function	ReLU
• Output layer	Dense
• Number of epochs	50
• Loss function	MSE
• Dropout Regularization Technique	✓

*Table 1. Summary of Machine Learning Model Parameters.*

For training the model, a `model.fit()` training loop checks at the end of every epoch whether the loss is no longer decreasing, considering the minimum delta and patience (16 in our model). Once the loss is found no longer decreasing the training terminates. Next, we compute two performance metrics:  $R^2$  (variance score) and Mean Squared Error (RMSE, test score).

The last step in the study is to compare the results obtained with the autoregressive models and the LSTM model.

### 3. RESULTS AND CONCLUSIONS

As mentioned above, after building both the benchmark autoregressive model and the ARX model, the forecasting performance and statistical



significance of each model is evaluated by the out-of-sample  $R^2$  statistic ( $R_{OS}^2$ ) suggested by Campbell and Thompson [4] and the MSPE-adjusted statistic developed by Clark and West (2007), respectively.

In the case of the model that uses eight variables, the  $R_{OS}^2$  is -0.0315, which means that the predictive model forecasts ( $\hat{y}_{t+1}$ ) do not outperform the benchmark forecasts ( $\bar{y}_{t+1}$ ). In light of these results, we build different models using different combinations of variables.

In *Table 2* we summarize the  $R_{OS}^2$  and Clark and West statistics for each model built and rank them according to its forecasting performance compared to the benchmark model and statistical significance. Highlighted in bold, there are the models that both outperform the benchmark model and are statistically significant (90% significance or above).

Rank	REDTI	Drilling Activity	Production	World Industrial Production	Electricity Consump.	Inventories	Monetary Policy	$R_{OS}^2$	Clark & West
<b>Model 1</b>	X		X		X	X		<b>0.0348</b>	<b><math>7.894 \cdot 10^{-5}</math></b>
<b>Model 2</b>	X							<b>0.0389</b>	<b><math>2.827 \cdot 10^{-5}</math></b>
<b>Model 3</b>	X	X	X		X	X		<b>0.0311</b>	<b><math>2.687 \cdot 10^{-4}</math></b>
<b>Model 4</b>	X		X			X		<b>0.0300</b>	<b>0.0010</b>
Model 5	X		X	X	X	X		-0.0076	<b><math>3.147 \cdot 10^{-4}</math></b>
Model 6	X	X	X	X	X	X		-0.0155	<b>0.0020</b>
Model 7	X		X			X	X	-0.0181	<b>0.0168</b>
Model 8	X	X	X	X	X	X	X	-0.0315	<b>0.0512</b>

*Table 2 . Summary of  $R_{OS}^2$  and MPSE obtained for each of the ARX models built.*

Natural gas prices are also predicted from September 2010 to January 2022 using machine learning, as mentioned above. *Figure 3* shows the comparison of the predicted values to de actual Natural Gas Henry Hub spot prices.

In order to compare the performance of the Machine Learning model to the autoregressive models, the R2 (variance score) and Root Mean Squared Error (RMSE, test score) is calculated for the benchmark autoregressive model and the ARX Model 1. Model 1 is used because it is the one with the best performance metrics.

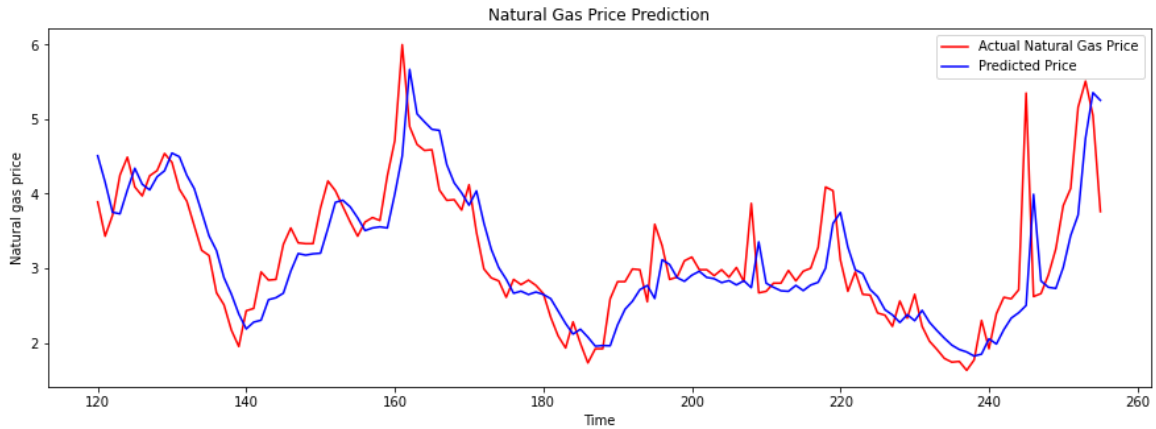


Figure 3. September 2010 – January 2022 Natural Gas Price Prediction using LSTM Model vs. Actual Natural Gas Price

The following values for all three models are obtained:

	Machine Learning	BM AR Model	ARX Model 1
R <sup>2</sup>	0.64 (64%)	0.32 (32%)	0.34 (34%)
RMSE	0.52€	0.51€	0.50€

Table 3. Performance metrics of the three different models

In terms of the variance explained by the model, the machine learning model is way superior to the autoregressive model, explaining approximately 30% more variance than the other two models. However, in terms of RMSE, all models perform a very similar error in average when predicting natural gas, around +/- 0.50€.

Taking these performance metrics, the ARX model that includes, apart from the historical natural gas price, the variations in residential demand due to weather conditions, in natural gas production, in natural gas inventories and in

electricity consumption, performs slightly better than the benchmark model that includes only historical natural gas prices to predict future prices.

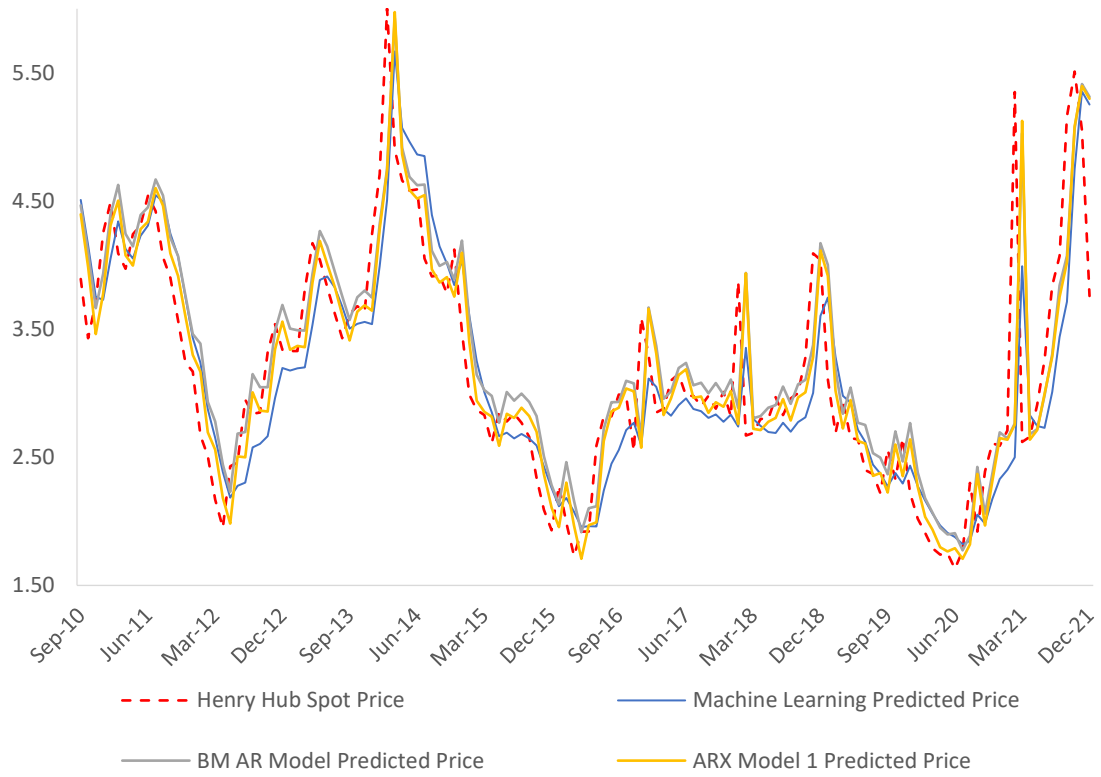


Figure 4. Machine Learning vs AR Model predictions vs Actual Price.

# PREDICCIÓN DEL PRECIO DEL GAS NATURAL

**Autor: Albendea López, Paula.**

Directora: Díaz Aguiluz, Elena María

Entidad Colaboradora: Universidad Pontificia Comillas

## RESUMEN DEL PROYECTO

### 1. INTRODUCCIÓN

El gas natural es un hidrocarburo compuesto de varios gases ligeros que se encuentran en la naturaleza. Este gas se formó hace millones de años en las profundidades de los suelos a partir de los restos de plantas y animales que se acumularon en gruesas capas en la superficie de la tierra y los fondos oceánicos [11].

Hoy en día, este gas se utiliza como combustible fósil, y se ha convertido en un medio esencial de energía. El gas natural se utiliza principalmente para generar electricidad, para hacer funcionar los sistemas de calefacción, como combustible para vehículos y para la producción industrial. El sector eléctrico, el sector industrial, el sector residencial, el sector comercial y el sector del transporte son los que más utilizan este combustible [13].

La demanda de gas natural está aumentando debido a su relación precio-poder calorífico. Según la Asociación de Información de Energía de los Estados Unidos (EIA), los Estados Unidos utilizaron alrededor de 31,5 cuatrillones de unidades térmicas británicas (Btu) en 2020, lo que representa el 34% del consumo total de energía de los Estados Unidos [13].

Después de una caída del 4% en 2020 debido a la crisis de COVID-19, la demanda de gas natural se recuperó progresivamente en 2021 a medida que el

consumo regresó, en los mercados maduros, a un nivel próximo al de antes de la crisis sanitaria. Para el periodo 2019-2025, se espera una tasa de crecimiento promedio de la demanda del 1,5% anual. La mayor parte de este crecimiento es atribuible a la región de Asia Pacífico, debido principalmente al aumento del consumo de gas en China y la India [31].

En el lado de la producción, se espera que América del Norte y Oriente Medio contribuyan a la mitad del aumento neto en el suministro de gas natural. Rusia es el segundo mayor productor y uno de los mayores exportadores de gas natural. El 24 de febrero de 2022, Rusia decidió lanzar una ofensiva a Ucrania después de una larga escalada de tensión entre los dos países. Las sanciones impuestas a Rusia por la Unión Europea y los Estados Unidos provocaron que el precio del gas se disparara en los territorios de la UE, donde casi el 40% del suministro de gas proviene de Rusia [32].



*Figura 1. Precios del gas natural en USD/MMBtu del 24 de febrero de 2022 al 27 de junio de 2022 [33].*

Además de la crisis causada por la guerra en Ucrania, la terminal de Freeport LNG en Texas se cerró el 9 de junio de 2022, debido a un incendio. El cierre estaba programado inicialmente para durar tres semanas, pero, más tarde, las autoridades anunciaron que se extendería hasta finales de 2022 [33].

Dados todos estos eventos, el precio por megavatio-hora para los contratos de julio en Europa, negociados en el punto de comercio virtual de gas natural (TTF) con sede en los Países Bajos, aumentó a EUR 127.17 (USD 133.49) el 22 de junio de 2022, un 60% más que EUR 79.40 (USD 83.35) el 8 de junio, que fue el nivel de cierre más bajo en los últimos cuatro meses desde el inicio de la guerra Rusia-Ucrania el 24 de febrero [33].



*Figura 2. Precio en €/MWh para contratos de julio negociados en futuros de gas TTF holandeses (21 de febrero de 2022 - 27 de junio de 2022) [34].*

Todos estos eventos han hecho que el precio del gas natural haya sido uno de los temas más comentados en los últimos meses. Este documento trata de encontrar una técnica que pueda predecir el precio del gas natural de la manera más precisa posible.

## 2. METODOLOGÍA

El estudio se divide en 8 capítulos. Un primer capítulo está dedicado a estudiar información general sobre el gas natural, como su formación, diferentes técnicas para su extracción, para su transformación en combustible energético y para su transporte hasta el usuario final. Luego, se identifica la oferta global, los mayores productores, almacenistas y exportadores. Lo mismo se realiza con la demanda global, estudiando cómo se distribuye esta demanda entre los diferentes

sectores. Finalmente, se señalan las implicaciones ambientales de la extracción, producción y quema de gas natural.

En segundo lugar, se dedica un capítulo a recopilar las variables que afectan a los precios del gas natural y describir cómo se construyen los indicadores utilizados en el modelo. Estas variables fueron identificadas en un estudio previo (*TFG: Modelling the Natural Gas Market*), donde se puede encontrar una investigación más detallada.

Antes de construir el modelo ARX, se construye un modelo de referencia para evaluar el rendimiento de este modelo ARX. Este modelo de referencia consiste en un modelo autorregresivo de primer orden, en el que el precio del gas natural para el período siguiente depende de un término independiente (alfa) y el precio del gas natural en el período actual multiplicado por un parámetro de escala (beta). El modelo se define como AR(1) y mediante la siguiente ecuación:

$$y_{t+1} = \alpha + \beta_1 y_t + \varepsilon_{t+1} \quad \varepsilon_t \sim (0, \sigma^2)$$

*Ecuación 3. Modelo de referencia [1].*

donde  $\varepsilon_{t+1}$  es ruido blanco,  $\alpha$  es una constante, y  $\beta_1$  es el parámetro del modelo [1].

Después de estimar los parámetros del modelo mediante una regresión, se calcula el pronóstico del precio del gas natural para el mes siguiente para cada uno de los meses del período septiembre 2010-enero 2022. El rendimiento del modelo de referencia se comparará con el modelo ARX.

En el modelo ARX, el precio del gas natural en el siguiente período ( $y_{t+1}$ ) depende del precio del gas natural un período anterior a la predicción ( $y_t$ ), y de la variación interanual o mensual de otras variables (véase el *Capítulo 2*) también un mes antes de la predicción ( $\Delta f_{1t} \dots \Delta f_{8t}$ ).

$$y_{t+1} = \alpha + \beta_1 y_t + \beta_2 \Delta f_{1t} + \dots + \beta_8 \Delta f_{7t} + \varepsilon_{t+1} \quad \varepsilon_t \sim (0, \sigma^2)$$

*Ecuación 2. Modelo ARX [1].*

El rendimiento de cada modelo se evalúa mediante el estadístico R2 fuera de la muestra, sugerida por Campbell y Thompson [4] para comparar los pronósticos en un horizonte de 1 mes.  $\hat{y}_{t+1}$  es el valor predicho del precio del gas natural utilizando el modelo ARX y  $\bar{y}_{t+1}$  es la predicción basada en el modelo de referencia. El estadístico  $R_{OS}^2$  viene dado por:

$$R_{OS}^2 = 1 - \frac{\sum_{t=t_0-1}^{T-1} (y_{t+1} - \hat{y}_{t+1})^2}{\sum_{t=t_0-1}^{T-1} (y_{t+1} - \bar{y}_{t+1})^2}$$

*Ecuación 3. Estadístico R2-OS [1].*

donde es el  $y_{t+1}$  valor registrado y real del precio del gas natural,  $t_0$  es el inicio de la muestra de predicción (septiembre de 2010) y  $T$  es el final de la muestra (diciembre de 2021). Si  $R_{OS}^2 > 0$ , las predicciones del modelo ARX ( $\hat{y}_{t+1}$ ) superan a las predicciones del modelo de referencia ( $\bar{y}_{t+1}$ ).

A continuación, examinamos si los resultados son estadísticamente significativos mediante la estimación del estadístico MSPE ajustado, propuesto por Clark y West (2007).

Primero definimos:

$$\varepsilon_{t+1} = (y_{t+1} - \bar{y}_{t+1})^2 - [(y_{t+1} - \hat{y}_{t+1})^2 - (\hat{y}_{t+1} - \bar{y}_{t+1})^2]$$

*figura4. Estadístico MPSE ajustado [1].*

y luego se calcula la regresión  $\{\varepsilon_{t+1}\}_{s=t_0}^{T-1}$  mediante una constante y se calcula el estadístico t bajo la hipótesis nula de que la constante es cero. El p-value se obtiene con una distribución normal estándar, para una prueba de una cola.

Después de construir el modelo ARX utilizando todas las variables mencionadas en el *TFG: Modelando el Mercado del Gas Natural*, también



construimos diferentes modelos ARX que combinan algunas de estas variables y medimos y comparamos el rendimiento de todos ellos.

Finalmente, se describe el concepto de "machine learning" así como las diferentes técnicas dentro de este campo para pronosticar el comportamiento de cualquier variable. Utilizando Series Temporales con Redes Neuronales de Memoria de Largo a Corto Plazo (LSTM), predecimos el precio del gas natural para el período de septiembre de 2010 a enero de 2022, de manera similar a lo que hacemos con el modelo autorregresivo, y comparamos las predicciones con los valores reales obtenidos en el sitio web de la EIA. Para ello, utilizamos el precio mensual histórico de Gas Natural descargado del sitio web de EIA (51) desde septiembre de 2000 hasta diciembre de 2021. Como marcos de software, utilizamos TensorFlow y Keras.

Los parámetros del modelo se resumen en la siguiente tabla:

• Descenso de gradiente estocástico con error de retropropagación	✓
• Set de test vs. Set de entrenamiento	20% vs. 80%
• Retrasos	6 meses
• Número de capas ocultas	4
• Número de neuronas por capa	200
• Optimizador	Adam
• Función de activación	ReLU
• Capa de salida	Densa
• Número de epochs	50
• Función de pérdida	ECM
• Técnica de regularización (Dropout)	✓

*Tabla 1. Resumen de los parámetros del modelo de aprendizaje automático.*

Para entrenar el modelo, un bucle de entrenamiento `model.fit()` verifica al final de cada *epoch* si el error deja de disminuir, considerando el delta mínimo y la paciencia (16 en nuestro modelo). Una vez que se verifica que el error ya no disminuye, el entrenamiento termina. A continuación, calculamos dos métricas de rendimiento:  $R^2$  (puntuación de varianza) y el Error Cuadrático Medio (RMSE, puntuación de la prueba).

El último paso de este estudio es comparar los resultados obtenidos con los modelos autorregresivos y el modelo LSTM.

### 3. RESULTADOS Y CONCLUSIONES

Como se mencionó anteriormente, después de construir tanto el modelo autorregresivo de referencia como el modelo ARX, el rendimiento de la predicción y la significación estadística de cada modelo se evalúan, respectivamente, mediante el estadístico  $R^2$  fuera de la muestra ( $R_{OS}^2$ ) sugerida por Campbell y Thompson [4] y el estadístico MSPE ajustado, desarrollado por Clark y West (2007).

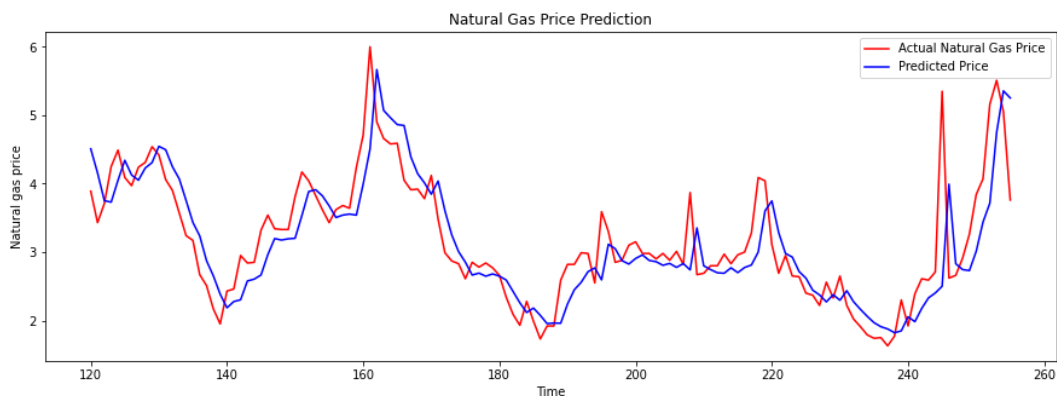
En el caso del modelo que utiliza ocho variables, el  $R_{OS}^2$  es -0.0315, lo que significa que la precisión en la predicción del modelo ARX ( $\hat{y}_{t+1}$ ) no superan a la precisión de las predicciones del modelo de referencia AR(1) ( $\bar{y}_{t+1}$ ). A la luz de estos resultados, construimos diferentes modelos utilizando diferentes combinaciones de variables.

En la *Tabla 2* resumimos los  $R_{OS}^2$  y los estadísticos de Clark y West para cada modelo y clasificamos estos modelos de acuerdo con su rendimiento de predicción en comparación con el modelo de referencia y su significación estadística. Resaltados en negrita, están los modelos que superan al modelo de referencia y son estadísticamente significativos (90% de significación o superior).

Rank	REDTI	Actividad perforación	Producción	Producción Industrial Mundial	Consumo Eléctrico	Inventario	Política Monetaria	$R^2_{OS}$	Clark & West
<b>Modelo 1</b>	X		X		X	X		<b>0.0348</b>	<b><math>7.894 \cdot 10^{-5}</math></b>
<b>Modelo 2</b>	X							<b>0.0389</b>	<b><math>2.827 \cdot 10^{-5}</math></b>
<b>Modelo 3</b>	X	X	X		X	X		<b>0.0311</b>	<b><math>2.687 \cdot 10^{-4}</math></b>
<b>Modelo 4</b>	X		X			X		<b>0.0300</b>	<b>0.0010</b>
Modelo 5	X		X	X	X	X		-0.0076	<b><math>3.147 \cdot 10^{-4}</math></b>
Modelo 6	X	X	X	X	X	X		-0.0155	<b>0.0020</b>
Modelo 7	X		X			X	X	-0.0181	<b>0.0168</b>
Modelo 8	X	X	X	X	X	X	X	-0.0315	<b>0.0512</b>

*Tabla 2.  $R^2$ -OS y estadístico de Clark & West obtenidos para cada uno de los modelos ARX construidos.*

El precio del gas natural también se predice desde septiembre de 2010 hasta enero de 2022 utilizando el aprendizaje automático o “machine learning”, como se mencionó anteriormente. *La Figura 3* muestra la comparativa entre los valores predichos mediante redes neuronales LSTM y los precios reales del Gas Natural en el Henry Hub.



*Figura 3. Comparativa gráfica del precio del gas natural predicho por el modelo de aprendizaje automático y el precio real (septiembre 2010 - enero 2022)*

Para comparar el rendimiento del modelo que usa “machine learning” con los modelos autorregresivos, se calcula el  $R^2$  (puntuación de varianza) y el error cuadrático medio raíz (RMSE, puntuación de prueba) para el modelo

autorregresivo de referencia y el modelo 1 ARX. Se utiliza el modelo 1 porque es el que tiene las mejores métricas de rendimiento.

Se obtienen los siguientes valores para los tres modelos:

	Machine Learning	Modelo Ref AR	Modelo 1 ARX
R <sup>2</sup>	0.64 (64%)	0.32 (32%)	0.34 (34%)
RMSE	0.52€	0.51€	0.50€

Tabla 3. Resumen de los R<sup>2</sup> y RMSE obtenidos para cada uno de los tres modelos.

En términos de la varianza explicada por el modelo, el modelo que usa aprendizaje automático es muy superior a los modelos autorregresivos, explicando aproximadamente un 30% más de varianza que los otros dos modelos. Sin embargo, en términos de RMSE, todos los modelos realizan un error promedio muy similar al predecir el precio del gas natural, alrededor de +/- 0.50 €.

Tomando estas métricas, el modelo ARX que incluye, aparte del precio histórico del gas natural, las variaciones en la demanda residencial debido a las condiciones climáticas, los niveles de producción de gas natural, los niveles de inventario de gas natural y el consumo de electricidad, tiene un comportamiento ligeramente mejor que el modelo de referencia que incluye solo los precios históricos del gas natural para predecir los precios futuros.

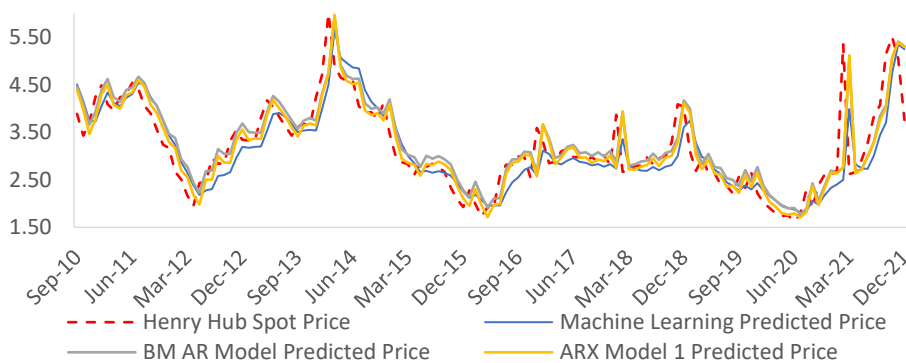


Figura 4. Predicciones usando Aprendizaje Automático vs usando Modelo Autorregresivo vs Precio real registrado en EIA.





Máster de Ingeniería Industrial

Trabajo de Fin de Máster

*Forecasting natural gas prices*

Author

Paula Albendea López

Director

Elena María Díaz Aguiluz

Madrid

January 2023

## **Acknowledgements**

To my project director, Elena María Díaz Aguiluz, for her dedication and closeness and for letting me work at my own pace while helping me in all I needed. I will always be grateful for her advice, patience and help in this project.

To my friends and partner, for always supporting me through the most difficult times and for always pushing me to continue.

To all the professors who, throughout my studies, have taught me everything I know today.

To my family, for the support they have always given me and the patience they have. Especially to my parents for all the efforts they make to always give me the best.

THANK YOU

# Table of Contents

<b>INTRODUCTION.....</b>	<b>28</b>
1.1. Motivation.....	28
1.2. Objectives.....	32
1.3. Methodology.....	34
<b>AN OVERVIEW OF NATURAL GAS.....</b>	<b>41</b>
2.1. Definition .....	41
2.2. Supply .....	43
2.3. Demand.....	47
2.4. Environmental Implications.....	50
<b>VARIABLE SELECTION .....</b>	<b>52</b>
3.1. Indicators of supply .....	53
3.2. Indicators of demand.....	56
3.3. Natural Gas Historical Price.....	63
<b>METHODOLOGY I: AUTOREGRESSIVE MODEL.....</b>	<b>66</b>
4.1. Benchmark Model .....	66
4.2. ARX.....	69
4.3. Forecasting Statistics .....	71
<b>METHODOLOGY II: MACHINE LEARNING.....</b>	<b>74</b>
5.1. The prediction problem .....	76



5.2.Dimensionality reduction: regression with regularization and Support Vector Machine .....	77
5.3. Decision trees .....	79
5.4. Artificial Neural Networks.....	81
5.5. Tensor Flow and Keras .....	86
5.6. Forecasting Natural Gas Prices with Machine Learning .....	88
<b>SUMMARY OF EMPIRICAL RESULTS .....</b>	<b>94</b>
6.1. Natural Gas Price Forecasts with Autoregressive Model .....	94
6.2. Natural Gas Price Forecasts with Machine Learning.....	98
6.1. Results Comparison: AR Model vs. Machine Learning .....	99
<b>CONCLUSIONS.....</b>	<b>102</b>
<b>FUTURE LINES OF INVESTIGATION.....</b>	<b>107</b>
<b>ANNEXES .....</b>	<b>109</b>
Sustainable Development Goals .....	110
MATLAB Code .....	114
Variables Matrix .....	117
Python Code.....	132
Predicted vs. Actual Prices .....	138
<b>BIBLIOGRAPHY .....</b>	<b>142</b>

# List of Figures

Figure 1.1. US 2020 Natural Gas Consumption by sector [26].....	29
Figure 1.2. Natural gas prices in USD/MMBtu from February 24th 2022 to June 27th 2022 [33].....	30
Figure 1.3. Price in €/MWh for July contracts traded in Dutch TTF Gas Futures (February 21st, 2022 - June 27th, 2022) [34].....	31
Figure 2.1. Petroleum and natural gas formation [12]. .....	42
Figure 2.2. Schematic geology of natural gas resources [12].....	42
Figure 2.3. Distribution of proved reserves by geographical area in 2000, 2010 and 2020 in percentages [19].....	45
Figure 2.4. World's natural gas supply by country. Source: own elaboration using data from NS Energy [44].....	46
Figure 2.5. Major natural gas trade movements in 2020 (Bcm) [19].....	47
Figure 2.6. US 2020 Natural Gas Consumption by sector [14].....	48
Figure 3.1. Year-to-year percent change of U.S. Natural Gas Gross Withdrawals. Source: own elaboration with data from EIA [6]. .....	54
Figure 3.2. Percentage change month-to-month of the number of rotary rigs in operation in the U.S. Source: own elaboration with data from EIA [6]. .....	55
Figure 3.3. Percentage change year-to-year of U.S. Natural Gas underground storage volume. Source: own elaboration with data from EIA [6].....	56
Figure 3.4. Residential Energy Demand Temperature Index 1990-2021. Source: own elaboration with data from NOAA [47].....	58
Figure 3.5. Electricity End Use 1990-2021. Source: own elaboration with data from EIA [6]. .....	59
Figure 3.6. Year-to-year percent change of World Industrial Production levels. Source: own elaboration with data from OECD data [11]. .....	61
Figure 3.7. Wu-Xia Shadow Rate 1990-2021. Source: own elaboration with data from Federal Reserve Bank of Atlanta [8]. .....	62
Figure 3.8. Historical nominal prices of natural gas (2001-2021) [25]. .....	64
Figure 3.9. Historical real prices of natural gas (2001-2021). Deflated with US CPI [25],[26].....	64
Figure 5.1. ML Prediction Procedures [36]. .....	77

Figure 5.2. Decision Trees Selection Procedure through Random Forest [36]....	81
Figure 5.3. Neural Network with 4 lags and one layer with 2 nodes.....	83
Figure 5.4. Example of RNN [41].....	85
Figure 5.5. Example of LSTM cell [41] .....	86
Figure 5.6. Historic price of natural gas since 2000. ....	90
Figure 5.7. Test set vs. training set data split. ....	90
Figure 5.8. September 2010 – January 2022 Natural Gas Price Prediction using LSTM Model vs. Actual Natural Gas Price.....	93
Figure 6.1. Figure 6.1. September 2010 – January 2022 Natural Gas Price Prediction using LSTM Model vs. Actual Natural Gas Price .....	99
Figure 6.2. Machine Learning vs AR Model predictions vs Actual Price.....	101
Figure A1.1. 17 ODSs. Fuente: Sustainable Development Goals. United Nations [20].....	111

## List of Tables

Table 1. Summary of R2Os and Clark & West for various models including different combinations of variable.....	73
Table 2. Keras Functions.....	88
Table 3. Summary of R2Os and Clark & West for various models including different combinations of variables. ....	97
Table 4. Summary of Machine Learning Model Parameters. ....	98
Table 5. Summary of the Machine Learning model's performance metrics. ....	99
Table 6. Performance metrics of the three different models .....	100
Table 7. Performance metrics of the three different models .....	106

# Chapter 1

## Introduction

### 1.1. Motivation

Natural gas is a hydrocarbon composed of several light gases found in nature. This gas was formed millions of years ago deep beneath the ground from the remains of plants and animals that built up in thick layers on the earth's surface and ocean floors [11].

Nowadays, this gas is used as a fossil energy fuel, and it has become an essential means of energy in today's world. Natural gas is used mainly to generate electricity, to run heating systems, as a fuel for vehicles, and for industrial production. The electric power sector, the industrial sector, the residential sector, the commercial sector, and the transportation sector are the ones that use this fuel the most [13].

Demand for natural gas is increasing due to its price-calorific value ratio. According to the U.S. Energy Information Association (EIA), the United States used about 31.5 quadrillion British thermal units (Btu) in 2020 which makes 34% of U.S. total energy consumption [13].

After a 4% drop in 2020 due to COVID-19 crisis, natural gas demand progressively recovered in 2021 as consumption returned near to its pre-crisis levels in mature markets. An average growth rate of 1.5% of demand per year is

expected for the period of 2019-2025. Half of this growth in consumption is attributable to the Asia Pacific region, due mainly by the raise of gas usage in China and India. Despite the current economic uncertainty, both countries still support reforms to increase the role of gas in the energy mix [31].

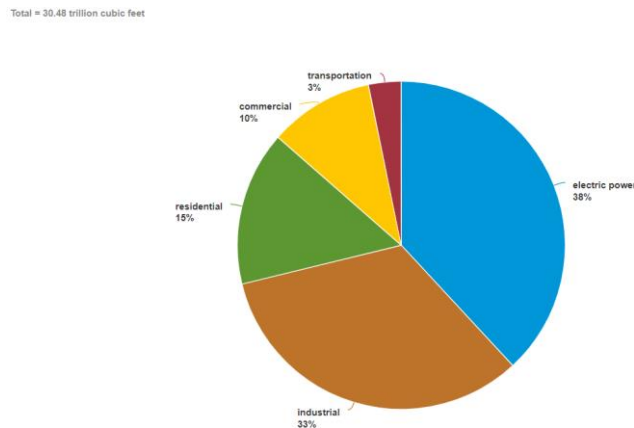


Figure 1.1. US 2020 Natural Gas Consumption by sector [26].

On the production side, North America and the Middle East are expected to contribute to half of the net increase in natural gas supply. In the US, however, spending on shale tight oil and gas declined almost by 50% y-o-y in 2020 due to the pandemics. In the Middle East, production growth will be driven by large conventional projects in Saudi Arabia, Iran, Israel, Iraq and Qatar [31].

Russia is the second largest producer and one of the biggest exporters of natural gas. On February 24<sup>th</sup>, 2022, Russia decided to launch an offensive to Ukraine after a long escalation of tension between the two countries. The sanctions imposed on Russia by the European Union and the United States, caused a rise in gas prices due to a decrease in Russian gas exports [32].

The Nord Stream 2 gas pipeline was closed as a result of a sanction imposed by Germany to Russia. This pipeline, which runs along the bottom of the Baltic Sea from Russia to Germany, supplies gas to many parts of Europe. However, after Putin's decision to recognize the territories of Donetsk and Lugansk as

independent from Ukraine, German Chancellor Olaf Scholz announced the cancelation of the certification that would give green light to the project [32].

U.S. President Joe Biden joined Germany on this initiative and sanctioned the companies involved in the project, including the Russian state-owned Gazprom, Shell and Engie. These sanctions caused the price of gas to rocket in EU territories, where almost 40% of the gas supply comes from Russia [32].

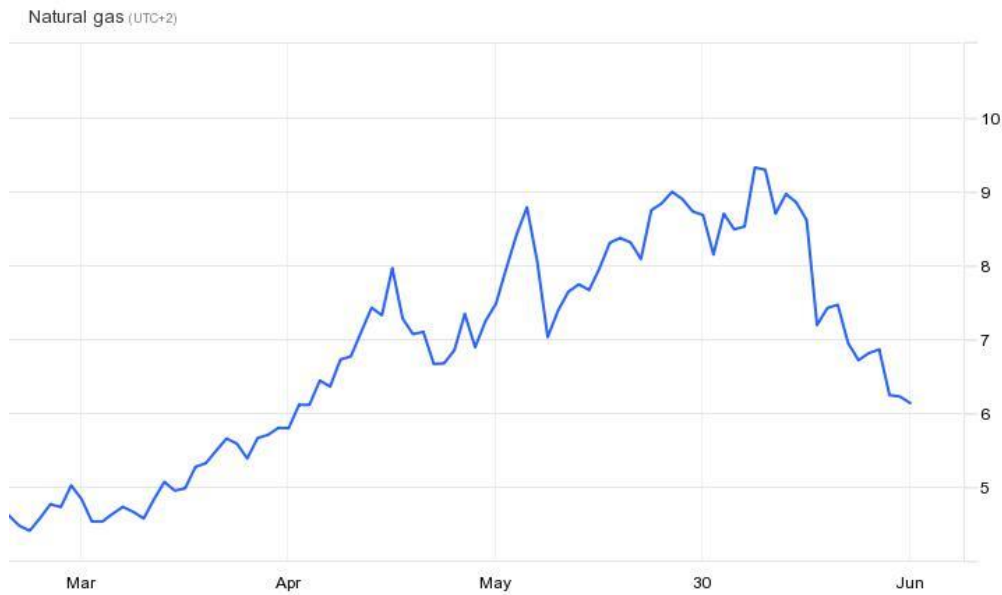


Figure 1.2. Natural gas prices in USD/MMBtu from February 24th 2022 to June 27th 2022 [33].

Russian energy company Gazprom announced on June 16 that daily gas shipments to Europe via the Nord Streamline will be reduced by 60%. The Yamal-Europe line stopped shipping gas to Europe, and exports through Ukraine were cut by about half.

On top of all the crisis caused by the war in Ukraine, the Freeport LNG terminal in Texas was closed on June 9<sup>th</sup>, 2022, due to a fire. The shutdown was initially scheduled to last three weeks, but the authorities later announced that it would be extended until the end of 2022 [33].

Given all these events, price per megawatt-hour for July contracts in Europe, traded on the Netherlands-based virtual natural gas trading point (TTF), rose to EUR 127.17 (USD 133.49) on June 22<sup>nd</sup>, 2022, up 60% from EUR 79.40 (USD 83.35) on June 8<sup>th</sup>, which was the lowest closing level in the last four months since the start of the Russia-Ukraine war on February 24<sup>th</sup> [33].



*Figure 1.3. Price in €/MWh for July contracts traded in Dutch TTF Gas Futures (February 21st, 2022 - June 27th, 2022) [34].*

Regarding environmental implications and the sustainability of this energy source in the long term, even though natural gas is considered the cleanest option among the fossil fuels due to its cleaner burning properties, there are some environmental implications attached to its drilling and extraction, transportation, burning and consumption [17]. Natural gas is mainly methane, which is a strong greenhouse gas, since its ability to trap heat is higher than that of carbon dioxide [18]. However, carbon dioxide stays longer in the atmosphere than methane molecules. When burned, natural gas emits lower levels of nearly all types of pollutants, including NO<sub>x</sub>, PM particles, and CO<sub>2</sub>, but it emits methane.

Drilling the wells from which natural gas is extracted can affect wildlife and land use, and radioactive materials, methane, and other underground gases

can leak into drinking water supplies. Moreover, hydraulic fracturing or fracking requires large quantities of water, which raises water-availability concerns in some communities.

Natural gas is not renewable, and, according to Worldometers, only 52 years of natural gas reserves are left to extract [21]. Yet, the natural gas resource base is more widely geographically dispersed than oil, which makes it a more reliable source of energy. This, along with its cleaner burning properties than other fossil fuels, makes natural gas a popular option to back up the transition to cleaner and renewable energy sources, such as solar or wind power, until the technology to store energy obtained from these sources is developed in order to meet demand even when there is no sun or wind.

## 1.2. Objectives

The main objective of this project is to forecast the price of natural gas, first by using an autoregressive model that takes into consideration a series of variables that affect demand and supply of natural gas and the price of natural gas itself in previous periods, and secondly, using machine learning techniques.

As it has been mentioned in the previous subchapter, natural gas has become an essential energy fossil fuel due to its cleaner burning capacities and a more even distribution of reserves across the globe. Many countries, with the aim of becoming more energy independent and looking to obtain energy in a cleaner way than with other fossil fuels such as petroleum or carbon, rely on natural gas, while they work on developing the technology to generate and store energy from renewable sources. This way, it is important to understand how the market for this commodity works and what might drive its price in the future.

Before venturing in developing a forecast model, it is fundamental to understand what natural gas is, how it is obtained, transformed into the final



product used as a source of energy, transported, distributed, the location of the reserves of this fuel and its environmental implications. All this information will help to understand what are the factors that affect the price of this commodity.

Natural gas price has a direct relationship with demand levels, while it has an inverse one with supply levels. The higher the demand, the higher the price, the lower the supply, the higher the price. The main forces affecting supply and demand for natural gas were studied in *TFG: Modelling the Natural Gas Market* [2]. Supply is affected by weather conditions, levels of storage and production, political relations and civil unrest, availability of workers and equipment, capacity of the pipeline system, access to natural gas deposits, and the financial environment. On the demand side, the cyclical nature of demand, weather, fuel switching, and the economic environment are the main factors to consider. The impact of each of these factors on the price of natural gas can be quantified using indicators related to each of these factors. These indicators were also identified in the *TFG: Modelling the Natural Gas Market* and are described and used in this study.

Once these indicators have been identified, the goal is to build a forecasting model that predicts the price of natural gas in one period using the price of the commodity in the previous one and the seven other variables described above and evaluate its forecasting performance.

In order to measure the model's performance, a benchmark model is first defined. This benchmark model is a first order autoregressive model in which the price of natural gas for a period is predicted using only the price for natural gas in the previous period. The forecasting performance of this model is then compared to the model that uses the other seven variables.

The next objective is to determine the forecasting model, which is an autoregressive model that uses exogenous inputs (ARX). The exogenous inputs are the seven variables described above. These variables appear in the model multiplied by a constant (beta) that needs to be calculated, as well as the independent term (alpha).

Then the goal is to evaluate the forecasting performance of these models. This is done by calculating the out-of-sample  $R^2$  statistic, which compares the error made in each of the forecasts using the benchmark model and the ARX model, by comparing the forecast and the actual historical price given by the EIA. After this, it is essential to determine if the models are statistically significant.

The next step is to study and describe machine learning techniques that would be useful in forecasting natural gas prices. Then, one of these techniques will be used to forecast the price of natural gas for the following year.

Finally, the autoregressive model and the machine learning techniques forecasting performance are compared.

### 1.3. Methodology

In order to achieve the objectives described above, a first chapter is dedicated to giving general information about natural gas. Firstly, the formation of natural gas is studied, as well as the different techniques to extract it from the ground, transform it into an energy fuel and transport it to the end-user. Then, global supply, the biggest producers, storers, and exporters are identified. Then the same is performed with global demand, studying how this demand is distributed across sectors. Finally, environmental implications of extracting, producing, and burning natural gas are flagged out.

Secondly, a chapter is dedicated to compile the variables that affect natural gas prices and describe how the indicators used in the model are built. These variables were identified in a previous study (*TFG: Modelling the Natural Gas Market*) and more detailed research can be there found.

Before building the ARX model, a benchmark model was constructed to measure the forecasting performance of this ARX model. This benchmark model consists of a first order autoregressive model, in which the price of natural gas for a period ahead is dependent on an independent term (alpha) and the price for natural gas in the current period multiplied by a scaling parameter (beta). The model is noted as AR(1) and defined by the following equation:

$$y_{t+1} = \alpha + \beta_1 y_t + \varepsilon_{t+1} \quad \varepsilon_t \sim (0, \sigma^2)$$

*Equation 4. Benchmark Model [1].*

where  $\varepsilon_{t+1}$  is white noise,  $\alpha$  is a constant, and  $\beta_1$  is the parameter of the model (1).

To estimate the alpha and beta of the model, the relation between natural gas observed price and the one observed the periods before is calculated from August 2010 to December 2021. For this, data of natural gas prices retrieved from the Energy Information Administration (EIA) website during the period of September 2000 to December 2021 is used. This way,  $y(t+1)$  and  $y(t)$  are known, and after performing a regression, the values for alpha and beta for each period are obtained.

This is performed with MATLAB and the function used is *regress(Y, X)*, where Y is a vector that takes for each month, starting August 2010, the monthly natural gas prices recorded until the moment of the calculation (since September 2000). X is a matrix that contains a first column of ones and a second column that, starting September 2010, takes monthly natural gas prices recorded until a

month before the moment of the calculation, also since September 2000. This means that  $X$  takes the lag of the observed data.

The regression is calculated for each month beginning in September 2010 and ending in December 2021 (135 times). For each of these calculations, the alphas and betas obtained are stored in a matrix called `BM_Parameters` with size  $135 \times 2$ . Finally, taking the alpha, beta, and natural gas observed price for each period from September 2010 to December 2021, we calculate the natural gas price forecast for the following month. This way, a vector under the name of `BM_Forecast` is created to store the forecasts for each of the periods calculated (September 2010-January 2022). The forecasting performance of this benchmark model will then be compared to the ARX model.

In the ARX model, the price of natural gas in the following period ( $y_{t+1}$ ) depends on the price of natural gas a period before the prediction ( $y_t$ ), and the year-to-year or month-to-month variation of seven other variables (*see Chapter 2*) also a month before the prediction ( $\Delta f_{1t} \dots \Delta f_{8t}$ ).

$$y_{t+1} = \alpha + \beta_1 y_t + \beta_2 \Delta f_{1t} + \dots + \beta_8 \Delta f_{8t} + \varepsilon_{t+1} \quad \varepsilon_t \sim (0, \sigma^2)$$

*Equation 5. ARX Model [1].*

The parameters of the model, the alpha and the eight betas, are determined the same way as in the benchmark model, with the difference that the matrix  $X$  now contains seven extra columns that gather the month-to-month or year-to-year variations of the other seven variables (REDTI, drilling activity, natural gas production, industrial production, electricity consumption, natural gas inventories, and monetary policy shadow rate) since September 2000 until the moment of the calculation.

The parameters obtained are stored in a 135x9 matrix called ARX\_Parameters. The natural gas price forecasts for September 2010-January 2022 are stored in a vector under the name of ARX\_Forecast.

The forecasting performance of each model is evaluated by the out-of-sample  $R^2$  statistic, suggested by Campbell and Thompson [4] to compare the  $\hat{y}_{t+1}$  and  $\bar{y}_{t+1}$  forecasts at a 1-month horizon.  $\hat{y}_{t+1}$  is the forecast using the predictive model described by *Equation 2* and  $\bar{y}_{t+1}$  is the forecast based on the benchmark model in *Equation 1*. This  $R_{OS}^2$  statistic is given by:

$$R_{OS}^2 = 1 - \frac{\sum_{t=t_0-1}^{T-1} (y_{t+1} - \hat{y}_{t+1})^2}{\sum_{t=t_0-1}^{T-1} (y_{t+1} - \bar{y}_{t+1})^2}$$

*Equation 3. R2 out of sample [1].*

where  $y_{t+1}$  is the realized and real value of the price of natural gas,  $t_0$  is the start of the forecasting sample (September 2010) and  $T$  is the end of the sample (December 2021). The  $R_{OS}^2$  measures the reduction in Mean Squared Prediction Error (MSPE) for the predictive regression model relative to the benchmark model, in percentage terms. Therefore, if  $R_{OS}^2 > 0$ , the predictive model forecasts ( $\hat{y}_{t+1}$ ) outperform the benchmark forecasts ( $\bar{y}_{t+1}$ ).

We then examine whether the results are statistically significant by testing against the null hypothesis of equal MSPE between the two models. The benchmark model, however, is nested in the predictive regression model, which makes the parameter  $\beta_2$  in *Equation 2* is zero in the population under the null. This produces an upward bias in the estimation of the MSPE of the predictive regression model produced by such parameter. Therefore, we estimate the MSPE-adjusted statistic developed by Clark and West (2007). This statistic adjusts for the bias by deducting the square difference in the point predictions generated by each model.

We first define

$$\varepsilon_{t+1} = (y_{t+1} - \bar{y}_{t+1})^2 - [(y_{t+1} - \hat{y}_{t+1})^2 - (\hat{y}_{t+1} - \bar{y}_{t+1})^2]$$

*Equation 4. MSPE-adjusted statistic [1].*

and then regress  $\{\varepsilon_{t+1}\}_{s=t_0}^{T-1}$  on a constant and calculate the t-statistic under the null hypothesis that the constant is zero. The p-value is obtained with a standard normal distribution, for a one-tailed test.

After building a complete ARX model using all the variables mentioned above, we also build different ARX models including in each of them only some and not all of the nine variables and measure and compare their performance.

Finally, the concept of “machine learning” is described as well as the different techniques within this field to forecast the behavior of any variable. We will use machine learning algorithms to train the machine on historical price records and predict the expected future natural gas price. We use the environment of Anaconda distribution, an open-source Python distribution platform. We will predict the Natural Gas price for the period of September 2010 – January 2022 and compare our predictions to the real values recorded at EIA’s website (51). This prediction will be done using Time Series with Long short-term memory (LSTM) Neural Networks. We use the Henry Hub natural gas spot price from September 2000 to December 2021 recorded at EIA’s website, monthly. As software frameworks, we will use TensorFlow and Keras. Throughout this application, stochastic gradient descent with backpropagation of errors to train deep neural networks will be used.

First, we extract historical natural gas prices from EIA’s website. Then we train the LSTM model, separating the data sample into a training set and a test set. We use 20% of the data for the test set and 80% for the training set. We then

define to get natural gas price 6 months back and predict the following month's price, this defines our number of time-steps to use in the LSTM models.

Next, we build an LSTM model with four hidden layers and an output layer, with 200 neurons. As activation function, we use the rectified linear activation function (ReLU), a dense layer with one neuron is used as output layer, the number of epochs is set to 50, as optimizer we use Adam and as loss function, we use the mean squared error (MSE).

Once the model is fit, we can estimate the performance of the model on the train and test datasets. As performance metrics we use R2 and MSE.

The last step in the study is to compare the results obtained with the autoregressive models and the LSTM model.

This project consists of six chapters that include the collection of the information obtained on natural gas, an explanation of the process followed to achieve the objectives described above, a description of the results obtained, and a chapter dedicated to conclusions and possible future lines of research (*Chapter 8*). Each chapter includes:

- **Chapter 2.** General description of natural gas, how it is obtained, transformed, and transported, as well as how reserves, production, and demand are distributed around the globe, and what natural gas is used for.
- **Chapter 3.** Recompilation of the variables that affect the price of natural gas, both on the supply and demand side, and explanation of the indicators used in the autoregressive model.
- **Chapter 4.** Construction of an autoregressive model to forecast natural gas prices, and evaluation of the performance of this model compared to a benchmark model.

- **Chapter 5.** Study of different machine learning techniques that can be used to forecast prices of the commodity. Application of one of these techniques to the forecasting of natural gas prices.
- **Chapter 6.** Summary of all the results obtained in the previous chapter, as well as a comparison between the forecasting performance of each of the techniques used throughout the document.
- **Chapter 7.** Compilation of the main conclusions of each chapter as well as the presentation of the results.
- **Chapter 8.** Proposal of several possible future lines of research.



# Chapter 2

## An Overview of Natural Gas

In this chapter, the composition of natural gas, how it is extracted from the ground, treated, transported, stored, and distributed is briefly described, as well as the main uses the world gives to this commodity.

### 2.1. Definition

Natural gas is a fossil fuel that forms deep beneath the earth's surface and is composed mainly of methane ( $\text{CH}_4$ ), although it also contains smaller amounts of natural gas liquids (NGL) and nonhydrocarbon gases, such as carbon dioxide and water vapor [12].

The formation of natural gas took place millions to hundreds of millions of years ago when the remains of plants and animals built up in thick layers on the earth's surface and ocean floors, sometimes mixed with sand, silt, and calcium carbonate. These layers were then buried under sand and rock and pressure and heat transformed this material into coal, oil, and natural gas [12].

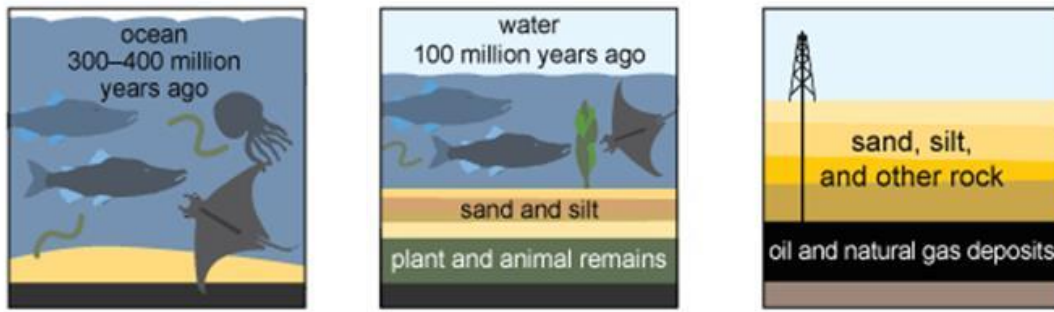


Figure 2.1. Petroleum and natural gas formation [12].

Depending on where natural gas is found, it receives a different name. The natural gas found in cracks and spaces between layers of rock is called conventional natural gas. The one that occurs in tiny pores within sedimentary rock is referred to as shale gas or tight gas, and it is sometimes called unconventional natural gas. Finally, the natural gas that occurs with deposits of crude oil is called associated natural gas and the one that is found in coal deposits is called coalbed methane [12].

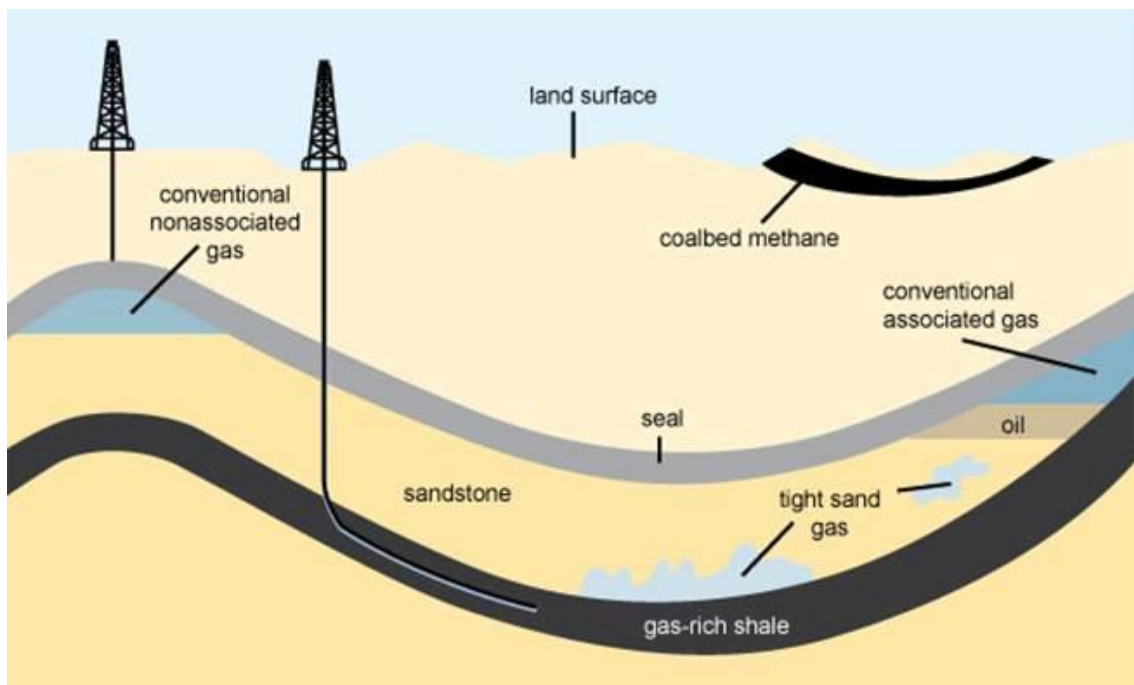


Figure 2.2. Schematic geology of natural gas resources [12].

There are two different ways to extract the natural gas contained in the rock depending on the permeability of the soil. If this permeability is high, the fuel can easily be extracted through wells since it flows up to the surface. However, if the permeability is low, natural gas is extracted by drilling both vertically and horizontally and forcing water, chemicals, and sand down the wells under high pressure. This high pressure breaks up the formation, releases the natural gas from the rock, and allows it to flow to and up wells to the surface. This technique is known as hydraulic fracturing or fracking [12].

Once the natural gas reaches the surface, there are two ways to transport it. If the distance it must travel is small and there is a pipeline network available, the gas is gathered into pipelines and sent to natural gas processing plants, where water vapor and nonhydrocarbon compounds found in the natural gas extracted from the ground are removed. This gas is often transported and stored at very high pressures to save space. If transporting it through pipelines is not feasible due to long distances or lack of infrastructure, the gas is liquified and transported in containers, at very low temperatures ( $-162^{\circ}\text{C}$ ), by boat. This type of gas is referred to as Liquified Natural gas (LNG) [2].

Natural gas is then sent through pipelines to underground storage fields or to distribution companies and then to consumers [12].

## 2.2. Supply

Natural gas must go through a series of treatments after extracting it from the well to remove impurities and obtain the final product, which is almost 100% methane. The main steps of this treatment process include oil and condensates removal, carbon dioxide and hydrogen sulfide removal, dehydration, and Natural Gas Liquids (NGL) recovery [13].

When this final product is achieved, some of this natural gas produced from a particular well might need to travel a great distance to reach its point of use. As explained in the subchapter above, this final product can be transported in its gaseous state at high pressures through pipelines or in its liquified state at very low temperatures on cargo ships.

There are three major types of pipelines depending on their function: the gathering system, the international pipeline system, and the distribution system. The gathering system consists of low pressure, small diameter pipelines that transport raw natural gas from the wellhead to the processing plant. Natural gas that is needed to travel long distances is transported through international pipelines, with the biggest diameter, high pressures, which reduces the volume of the natural gas by up to 600 times [25]. Once the raw natural gas has been processed, the final product is delivered to the end user through the distribution system.

Along the transportation route, to monitor malfunctions, leaks, or any other unusual activity along the pipeline, there are a series of equipment. Compressor stations along the network ensure that the fuel travels at the right pressure, while meter stations determine the conditions of the fuel being transported (flow rate, temperature, pressure, etc.). Finally, sophisticated control systems, such as Control and Data Acquisition (SCADA) systems, are used to monitor the gas as it travels through the lengthy pipeline network. These systems collect, assimilate, and manage data received from the metering stations, compressor stations and valves, and might even be able to remotely operate certain equipment, including compressor stations, giving the opportunity to adjust flow rates immediately and easily from the centralized control station [2].

According to the International Energy Agency (IEA), natural gas supplies approximately 23% of the world’s primary energy demand. Global production has risen since the 2008 financial crisis, boosted by the development of unconventional shale fracking technologies. According to the BP Statistical Review of World Energy 2021, natural gas production totaled to 3.85 trillion cubic meters (Tcm) in 2020, decreasing by 122 Bcm (-3.3%) with respect to 2019’s production [20].

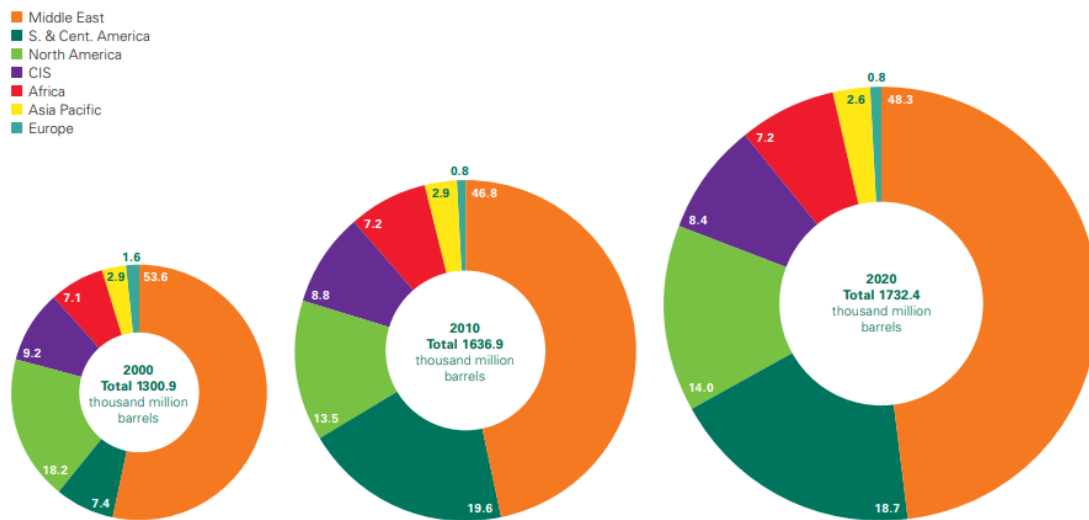


Figure 2.3. Distribution of proved reserves by geographical area in 2000, 2010 and 2020 in percentages [19].

According to the BP Statistical Review of World Energy 2020, the world’s total proved reserves of natural gas decreased by 2.2 Tcm to 188.1 Tcm in 2020, while in 2000 these proved reserves only amounted to 138.0 Tcm. Algeria (2.3 Tcm) provided the largest fall in reserves of 2 Tcm. Russia (37.4 Tcm), Iran (32.1 Tcm) and Qatar (24.7 Tcm) are the countries with the largest reserves. The current global R/P ratio shows that gas reserves in 2020 accounted for 48.8 years of current production. The Middle East (110.4 years) and CIS (70.5 years) are the regions with the highest R/P ratio [19].

The US accounted for 23.7% of the world’s natural gas production in 2020, with 914.6 Bcm, being the biggest natural gas producer in the world, followed by

Russia, supplying 638.5 Bcm in 2019, 16.6% of the global total. Russia is also the country that has the largest-known reserves of the fuel (est. 37.4 Tcm), most of which is located in Siberia.

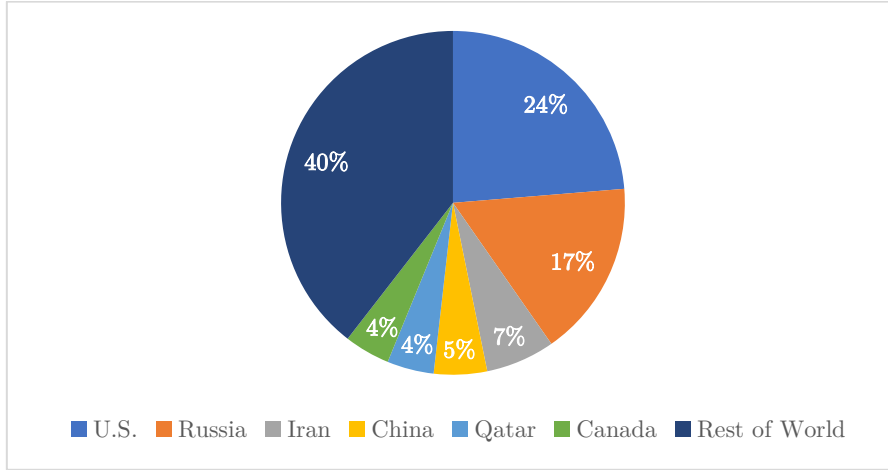


Figure 2.4. World's natural gas supply by country. Source: own elaboration using data from NS Energy [44].

The third biggest producer after the U.S.A and Russia was Iran, producing 6.5% of the world's natural gas in 2020, 250.8 Bcm [19]. Iran also has the second-biggest reserves globally, which means that even if as of today its production is not even close to that of the U.S., it is probably going to be a key player in the natural gas market in the near future [2].

The rest of the production is mainly attributable to China, accounting for a 5.0% of the global natural gas share in 2020 (194.0 Bcm), Qatar, with a 4.4% (171.3 Bcm), and Canada, with 4.3% of the global supply (165.2 Bcm).

Qatar is the world's biggest producer of liquefied natural gas (LNG), and was the second biggest LNG exporter in 2020, only after Australia. The country is working on the North Field East expansion project, to develop what will become the largest LNG project in the world with a capacity of 33 million tons per year [19].

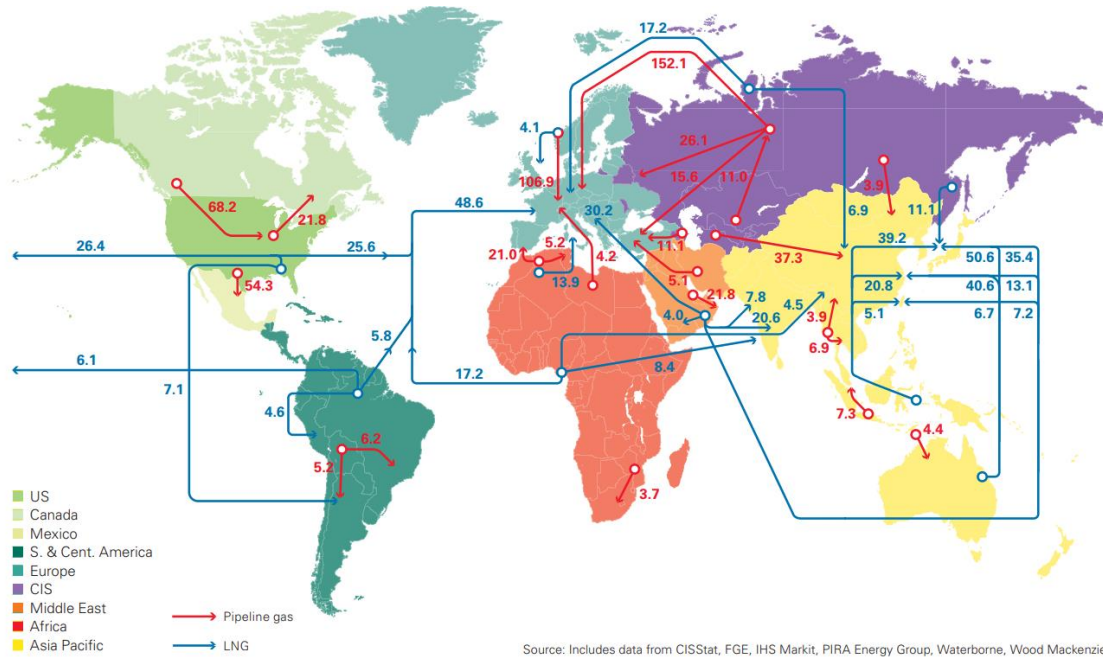


Figure 2.5. Major natural gas trade movements in 2020 (Bcm) [19].

### 2.3. Demand

Demand for natural gas has increased in recent years due to its price-calorific value ratio and the fact that is considered to be a “greener” option than other fossil fuels. According to the U.S. Energy Information Association (EIA), the United States used about 31.5 quadrillion British thermal units (Btu) in 2020, 34% of U.S. total energy consumption [14].

The most common uses of natural gas are electricity, heating, transportation, and production. The sectors that use this fossil fuel the most are the electric power, industrial, residential, commercial, and transportation sectors [14].

The electric power sector uses natural gas to generate electricity. Natural gas power plants usually generate electricity in gas turbines, directly using the hot gases of fuel combustion. “Combined-cycle” (NGCC) plants, with two turbines, a gas turbine operated with combustion gases, and a steam turbine

operated with the steam from boiling water that has been heated up with the hot gases from the first turbine. Natural gas-fired plants are among the cheapest power plants to build and have greater operational flexibility than coal plants because they can be fired up and turned down rapidly [16]. In 2020, the electric power sector accounted for about 38% of total U.S. natural gas consumption, according to the EIA, and natural gas was the source of about 33% of the U.S. electric power sector's primary energy consumption. It accounted for 40% of total utility-scale U.S. electricity generation by all sectors in 2020 [14].

Total = 30.48 trillion cubic feet

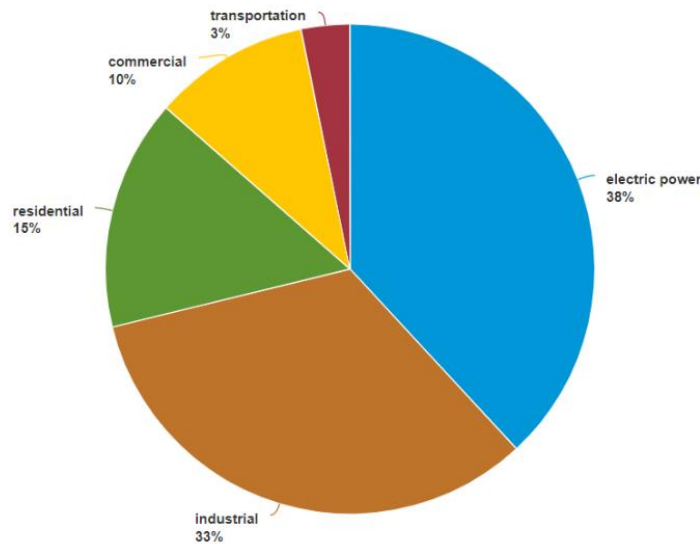


Figure 2.6. US 2020 Natural Gas Consumption by sector [14].

The industrial sector uses natural gas as a fuel for process heating, and as a raw material to produce chemicals, fertilizers, and hydrogen. One of the most important industrial products derived from natural gas is ammonia, used for fertilizer, which is produced by making high temperature water vapor react with methane. In 2020, the industrial sector accounted for about 33% of total U.S. natural gas consumption, and natural gas was the source of about 34% of the U.S. industrial sector's total energy consumption [14].



The residential sector uses natural gas to heat buildings and water, to cook, and to dry clothes. According to the EIA, around 50% of the homes in the United States use natural gas for heating. Natural gas heating is more effective than electric heating pumps, since heat from natural gas is delivered from forced-air systems at temperatures between 50-60°C, whereas the air from an electric heat pump is typically delivered at 30-35°C [15]. In 2020, the residential sector accounted for about 15% of total U.S. natural gas consumption, and natural gas was the source of about 23% of the U.S. residential sector's total energy consumption [14].

The commercial sector uses natural gas to heat buildings and water, to operate refrigeration and cooling equipment, to cook, to dry clothes, to provide outdoor lighting, and as a fuel in combined heat and power systems. In 2020, the commercial sector accounted for about 10% of total U.S. natural gas consumption, and natural gas was the source of about 19% of the U.S. commercial sector's total energy consumption [14].

The transportation sector uses natural gas as a vehicle fuel (in the form of CNG and LNG) and to operate compressors that move natural gas through pipelines, as mentioned in the last chapter. In 2020, the transportation sector accounted for about 3% of total U.S. natural gas consumption and natural gas was the source of about 4% of the U.S. transportation sector's total energy consumption, of which 94% was for natural gas transportation and distribution [14].

Another increasing use of natural gas is cogeneration and trigeneration. Cogeneration is the simultaneous generation of electricity and heat, while trigeneration is the simultaneous generation of electricity, heat and cooling. These forms of generation greatly increase the efficiencies of natural gas [16].

## 2.4. Environmental Implications

Natural gas is considered the cleanest option among fossil fuels, due to its clean burning properties. The process of burning natural gas to obtain energy results in fewer emissions of nearly all types of air pollutants, including NO<sub>x</sub>, PM particles, and CO<sub>2</sub> than burning coal or petroleum to produce an equal amount of energy. Around 117 pounds of carbon dioxide are produced per MMBtu equivalent of natural gas compared to more than 200 pounds of CO<sub>2</sub> per MMBtu of coal and more than 160 pounds per MMBtu of fuel oil [20].

However, there are some environmental implications attached to natural gas drilling and extraction, transportation, burning and consumption [20]. Drilling the wells from which natural gas is extracted can affect wildlife and land use, as clearing and leveling of the area is required. To make transportation feasible, some pipelines are buried in the land, affecting the ecosystem around them, breaking migration patterns, and causing pollution of dirt and pollutants [22].

Moreover, hydraulic fracturing or fracking requires large quantities of water, that cannot be then returned to rivers or lakes. It has been estimated that a typical well may require 2–10 million gallons of water for fracking, although this amount may vary depending on the quality of the water and the shale formation [24]. This water also poses health risks to nearby communities, since more than 1,000 chemical additives are used in fracking fluids, and radioactive materials, methane, and other underground gases have sometimes leaked into drinking water supplies from improperly cased wells. In Ohio and Pennsylvania, for example, there have been documented cases of groundwater near oil and gas wells being contaminated with fracking fluids as well as with gases [2].

Natural gas is mainly methane, which is a strong greenhouse gas that leaks to the atmosphere in big amounts. The U.S. EPA estimates that in 2018, methane

emissions from natural gas and petroleum systems and from abandoned oil and natural gas wells were the source of about 29% of total U.S. methane emissions and about 3% of total U.S. greenhouse gas emissions [23]. A methane molecule is approximately 90 times more effective at trapping heat in the atmosphere than a molecule of carbon dioxide, although methane disintegrates faster. According to energy experts, CO<sub>2</sub> emissions need to be eliminated because they stay longer in the atmosphere, but to keep air temperatures from raising more than 2 degrees Celsius by 2050 it is also essential to keep any extra methane from leaking into the atmosphere [24].

Natural gas is flared at well sites in areas where it is not economical to transport the fuel for sale or when it contains high concentrations of hydrogen sulfide. This process of burning natural gas produces CO<sub>2</sub>, carbon monoxide, sulfur dioxide, nitrogen oxides, and many other compounds [22].

This commodity is also not renewable. According to Worldometers, only 52 years of natural gas reserves are left to extract [21]. Yet, the natural gas resource base is more widely geographically dispersed than oil, which makes it a more reliable source of energy. This, along with its cleaner burning properties than other fossil fuels, makes natural gas a popular option to back up the transition to cleaner and renewable energy sources, such as solar or wind power, until we develop the technology to store energy obtained from these sources in order to meet demand even when there is no sun or wind.

# Chapter 3

## Variable Selection

In order to model the market for natural gas and forecast natural gas prices, we gathered a few indicators or variables which could explain the variations in the price of this fuel. This selection is based on thorough research performed for the elaboration of the *TFG – Modelling the natural gas market* [2]. In the mentioned document, a description of the natural gas supply and demand process is exposed, which justifies the selection of these variables, and the natural gas market is modelled based on historic natural prices and the variation of these variables.

We divided these indicators into two groups: indicators of supply and indicators of demand. In this chapter, we define each of these indicators and explain how they are constructed.

In the *TFG – Modelling the natural gas market* [2], it was found out that the variables that could explain the most part of natural gas price fluctuations were the residential demand due to change in temperatures, the levels of natural gas production and inventories held, and the amount of electricity consumed.

Using an autoregressive model (ARX), we forecast the price of natural gas one period ahead ( $t+1$ ). We first include all eight variables in the model and measure its forecasting performance, then we include different combinations of the eight variables based on the results we obtained in the *TFG – Modelling the natural gas market* [2] and measure the forecasting performance of the model each time, to try to reach the most significant and most accurate forecasting model.

### 3.1. Indicators of supply

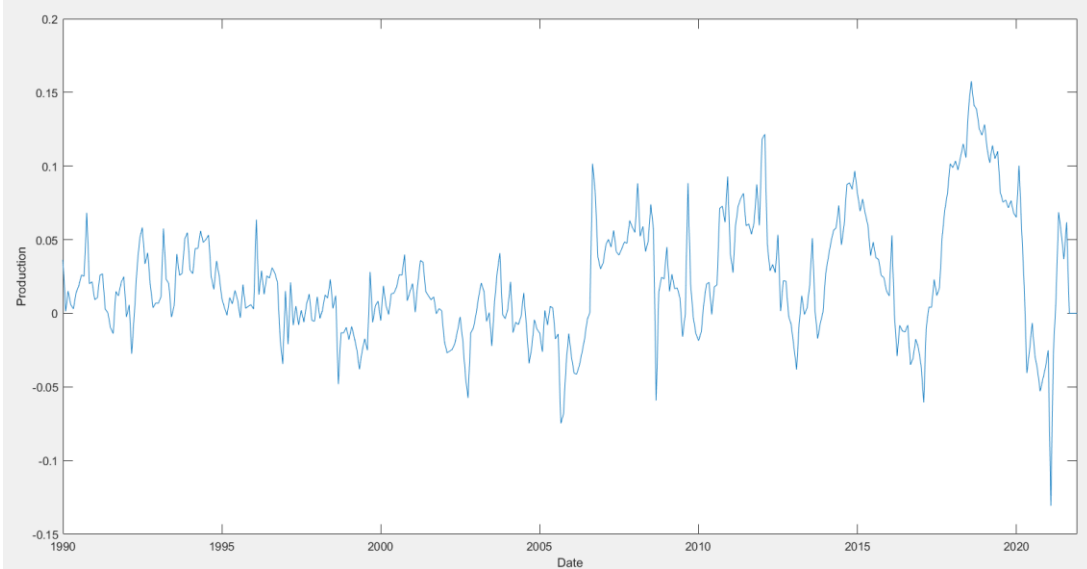
Increases in natural gas supply normally result in lower prices of this resource, while decreases in supply generally lead to higher prices. The amount of natural gas supplied depends on a series of external factors, among which weather conditions, levels of storage and production, political relations and civil unrest, availability of workers and equipment, capacity of the pipeline system, access to natural gas deposits, and the financial environment, might be the most influential ones [17]. All these factors are considered in the forecasting model through some indicators described below.

#### U.S. Natural Gas Gross Withdrawals

The level of production is the most direct factor affecting supply. Normally, production tends to increase when demand is high, or reserves are low [17]. When production is disrupted due to several factors, supply is interrupted or lowered, depending on the amount of natural gas held in storage.

U.S. Natural Gas Gross Withdrawals is an indicator is used to determine the amount of natural gas produced each month in the U.S. This amount is measured in MMcf and the information is available at the EIA website beginning 1980 [2].

Then, year-to-year percent change of the amount of natural gas that was withdrew each year in the U.S. is computed to include the impact of the change on production levels on the fluctuations of the price of natural gas.

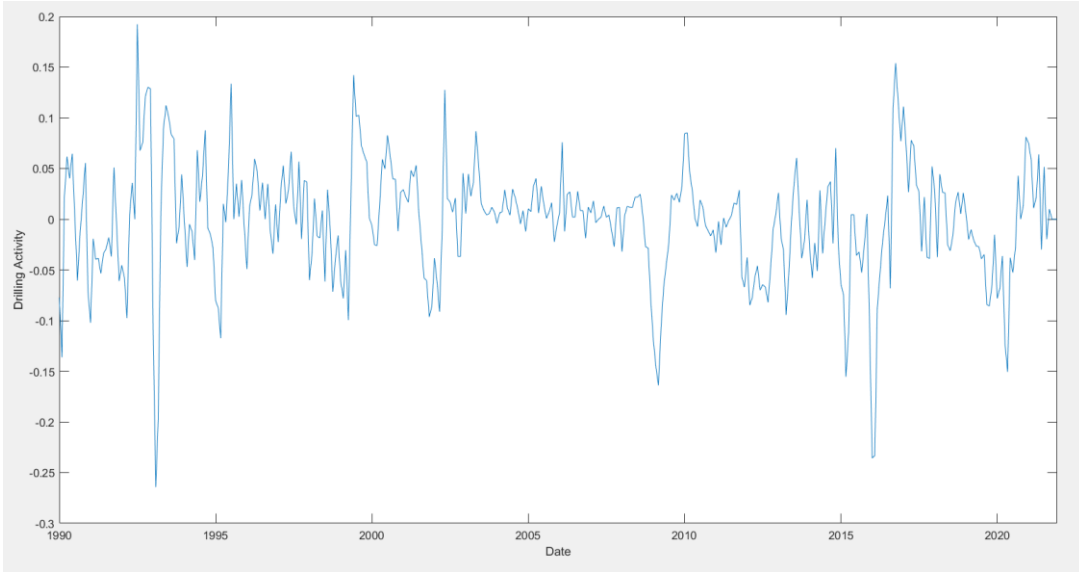


*Figure 3.1. Year-to-year percent change of U.S. Natural Gas Gross Withdrawals.  
Source: own elaboration with data from EIA [6].*

### U.S. Natural Gas Rotary Rigs in Operation

A rotary rig is the equipment used for drilling in most wells, which includes an engine and a hoisting, rotating and mud circulating system [2]. This indicator counts the number of rotary rigs being operated in the U.S. and it is used to determine the level of drilling activity in the U.S. The historical information of the number of rigs in operation is available at the EIA website since 1987.

In the model, the percentage change month-to-month of the number of rigs in operation is used to include the impact of the change in the level of drilling activity on the price of natural gas.



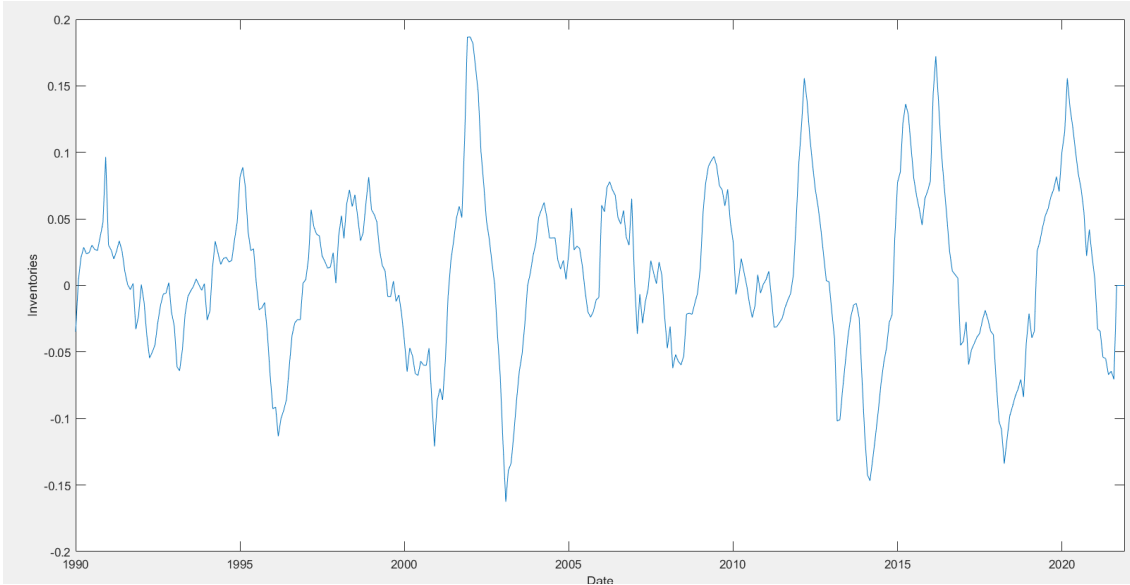
*Figure 3.2. Percentage change month-to-month of the number of rotary rigs in operation in the U.S. Source: own elaboration with data from EIA [6].*

### U.S. Natural Gas Underground Storage Volume

The level of natural gas in underground storage fields has a large influence on overall supply. When demand is higher than production, natural gas held in storage is used to satisfy the needs for this fuel. On the other hand, when supply is higher than demand, storage serves to absorb this excess in production and saves it for when it is needed. If there is a shortage in production and storage levels are not high enough, supply might fall, failing to meet demand [2].

To determine the levels of inventories of natural gas, the US Natural Gas underground storage volume is the measure used. This information can be found at the EIA website and is updated every month since 1979. The volume of underground storage is measured in MMcf.

In the forecast model, the change year-to-year of this volume is used, in order to include the effect of the change of this variable on the price of natural gas.



*Figure 3.3. Percentage change year-to-year of U.S. Natural Gas underground storage volume. Source: own elaboration with data from EIA [6].*

## 3.2. Indicators of demand

Natural gas is believed to be a key fuel for the transition to a zero-emission world. Natural gas, therefore, is expected to play an increasingly important role in meeting demand for energy [2].

Demand for natural gas depends highly on the time of year, and changes from season to season. The highest demand has traditionally been registered for the coldest months (January and February) while the lowest demand has been associated with the warmest months of summer (July and August). The main driver for this primary cycle of natural gas demand is the need for residential and commercial heating [18].

Even if normally the peak demand for natural gas is registered in the winter during the cold months, untypical weather conditions might reverse the cycle. An extremely hot summer, for example, can result in a rise in natural gas demand during the summer months, while a warmer winter might decrease the needs for this fuel [2].



Demand for natural gas is also affected by the price of natural gas versus other fuels, such as coal. Consumers tend to choose the cheapest fuels. While most residential and commercial customers can only rely on natural gas to meet many of their energy requirements, some industrial and electric generation consumers have the capacity to switch between fuels [18].

Another factor that might affect natural gas demand is the state of the global economy. When the economy is expanding, output from industrial sectors generally increases at a similar rate, and when the economy is in recession, output from industrial sectors drops. The amount of output coming out from industries determines the energy needed by these industrial users, and therefore affects demand for natural gas [2]. For instance, during the economic downturn of 2008, industrial natural gas consumption fell by over 7% [18].

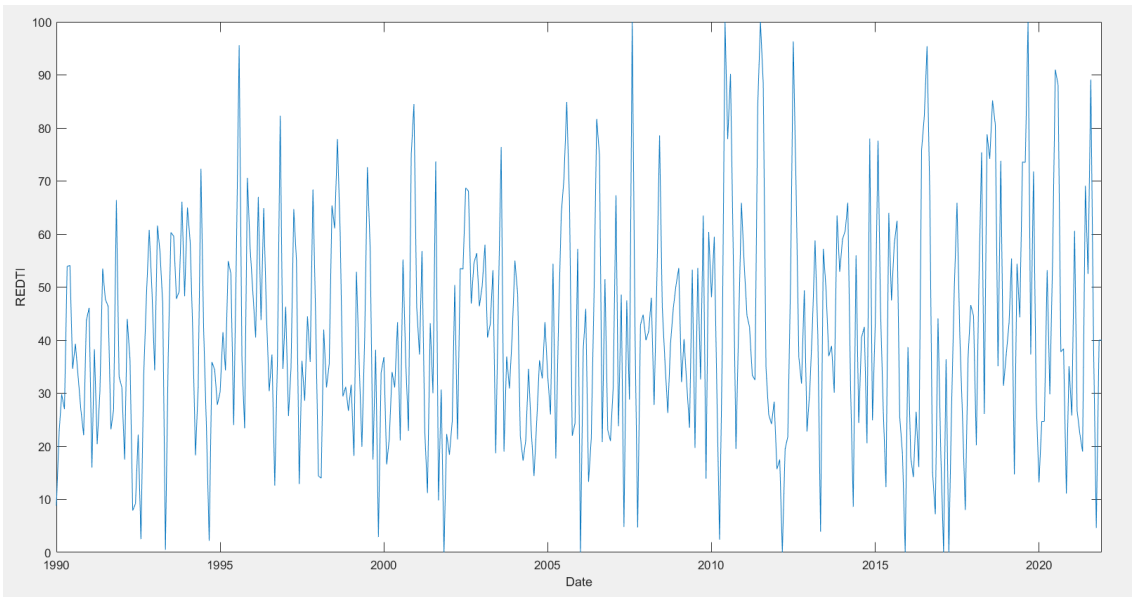
Five variables were considered in this study to consider the impact of the change on the demand of natural gas in its price. These variables include the effects on the demand of natural gas of weather conditions, electricity consumption, world industrial production, and monetary policy.

### **REDTI**

The Residential Energy Demand Temperature Index (REDTI) is based on population weighted heating and cooling degree days. This index is a relevant tool for explaining year-to-year fluctuations in energy demand for residential heating and cooling. Residential energy consumption is greatly correlated to the number of heating and cooling degree days (HDD's and CDD's). Diaz and Quayle (1980) found this correlation between energy use and heating degree days to be as high as 0.97 at the household level. Because of this strong relationship, seasonal changes in the REDTI can provide a good indication of the fluctuating energy demands [7].

The REDTI is calculated on a seasonal basis, using the sum of population weighted HDD's and CDD's (base 65), to provide retrospective information on the impact of seasonal temperatures on residential energy demand from 1895 to present. To simplify year-to-year comparisons, the index is scaled from 0 to 100. An index of 100 is assigned to the year with the greatest population weighted degree day average while the year with the smallest degree day average receives an index of 0 [7].

The winter season index, for example, is created by calculating the population weighted degree day totals for each winter season. From this series, the maximum and minimum yearly degree day totals are identified. The minimum value is then subtracted from all years in the series so that a new series of values is created which ranges from zero to a value equal to the arithmetic difference between the maximum and minimum. This new series is then scaled to have a range of 0 to 100 using a common scaling factor [7].



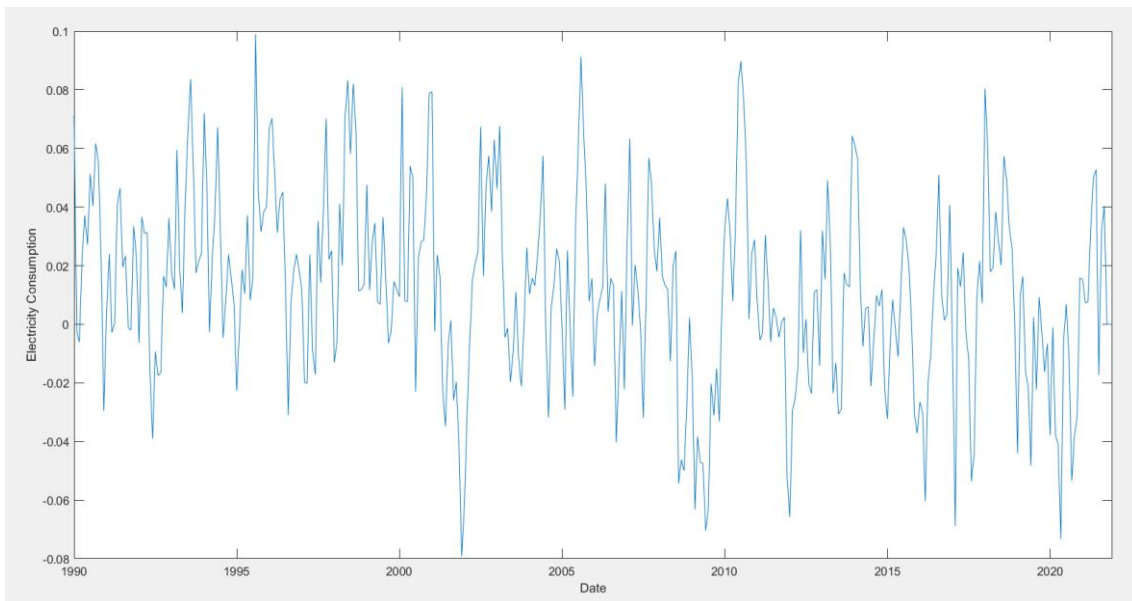
*Figure 3.4. Residential Energy Demand Temperature Index 1990-2021. Source: own elaboration with data from NOAA [47].*

## Electricity End Use

In order to take into consideration how the variations in electricity demand affect the price of natural gas, we included in the model the amount of electricity used historically in the U.S., measured in millions of kilowatts per hour.

This data was obtained from the Monthly Energy Review published by the U.S. Energy Information Administration (EIA). This publication includes total energy production, consumption, stocks, and trade; energy prices; overviews of petroleum, natural gas, coal, electricity, nuclear energy, renewable energy, and carbon dioxide emissions; and data unit conversions values [6].

The percent change year to year is then computed, obtaining the transformed data shown in *Figure 3.5*.



*Figure 3.5. Electricity End Use 1990-2021. Source: own elaboration with data from EIA [6].*

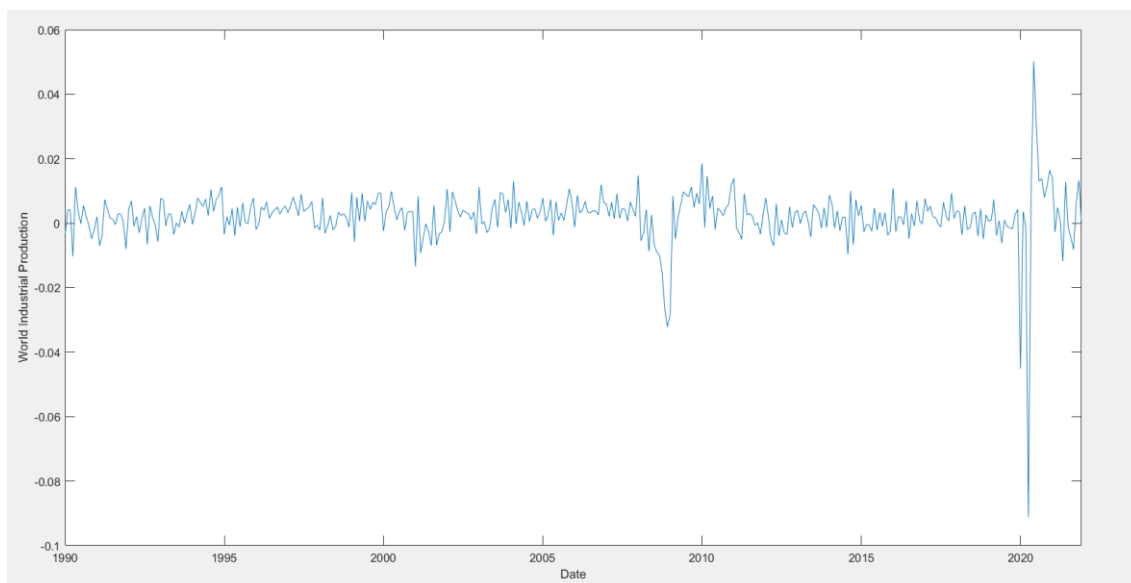
## World Industrial Production

The Baumeister's World Industrial Production index is an indicator of the world's economic activity. For empirical analysis of the determinants of variables like commodity prices researchers often use a monthly measure of economic

activity. This global economic activity can be measured either using a market price such as the cost of shipping as a proxy (Kilian, 2009) or an index of global industrial. OECD Main Economic Indicators published an estimate of monthly industrial production for the OECD plus 6 other major countries (Brazil, China, India, Indonesia, the Russian Federation and South Africa). The OECD series begins in January 1958 and ends in October 2011. Baumeister and Hamilton (2019) reproduced the methodology by which the original index was constructed to extend the series up to July 2018 [10]. Although shipping costs offer a plausible option for measuring the level of real economic activity, in practice they do not work nearly as well as estimates of global industrial production, particularly after 2015.

The world industrial production index measures levels of production in the manufacturing, mining, oil and gas field drilling services, and electrical and gas utilities sectors. It also measures capacity, which is an estimate of the production levels that could be sustainably maintained, and capacity utilization, which is the ratio of actual output to capacity. Industrial production and capacity levels are expressed as an index level relative to a base year.

The year-to-year percent change in this index of world industrial production is used in our forecast model to take into account the impact global economic activity has on the demand and therefore on the price of natural gas.



*Figure 3.6. Year-to-year percent change of World Industrial Production levels. Source: own elaboration with data from OECD data [11].*

### Wu-Xia Shadow Federal Funds Rate

After the 2007-2010 financial crisis, a tool was proposed, a “shadow rate,” that can be used in established economic models to measure the economic effects of unconventional monetary policies, such as the Federal Reserve’s quantitative-easing program. Traditional economic models used for research do not capture these effects when key interest rates sit at or near zero, as they do in much of the developed world today [9].

In normal economic times, economists use the federal funds rate, which is the interest rate banks use to lend to each other overnight, in many economic models. But in 2009, the fed funds rate hit zero, and monetary policy entered the zone termed the “zero lower bound”, and the fed funds rate stopped working in models [9].

Chicago Booth’s Jing Cynthia Wu and Fan Dora Xia, who now work at the Bank of International Settlements, proposed an alternate shadow fed funds rate that can be negative, reflecting the Fed’s additional easing through unconventional policies [9].

The model of Cynthia Wu and Fan Dora Xia uses one-month forward rates beginning in years hence. Wu and Xia use forward rates corresponding to  $n = 1/4, 1/2, 1, 2, 5, 7,$  and 10 years. These forward rates are constructed with end-of-month Nelson-Siegel-Svensson yield curve parameters from the Gurkaynak, Sack, and Wright (2006) dataset. The shadow rate is assumed to be a linear function of three latent variables called factors, which follow a VAR process. The latent factors and the shadow rate are estimated with the extended Kalman filter [8].

By plugging the shadow rate into a vector autoregression model (VAR), the effects of quantitative easing on economic aggregates such as the unemployment rate, industrial production, and housing starts, can be measured [9].

Shadow rate models are used by many researchers to characterize the term structure of interest rates (Kim and Singleton [2012] and Bauer and Rudebusch [2013]) or quantify the stance of monetary policy (Bullard [2012] and Krippner [2013]) [8]. In this model estimates from the Wu and Xia (2016) model of the shadow rate are used.

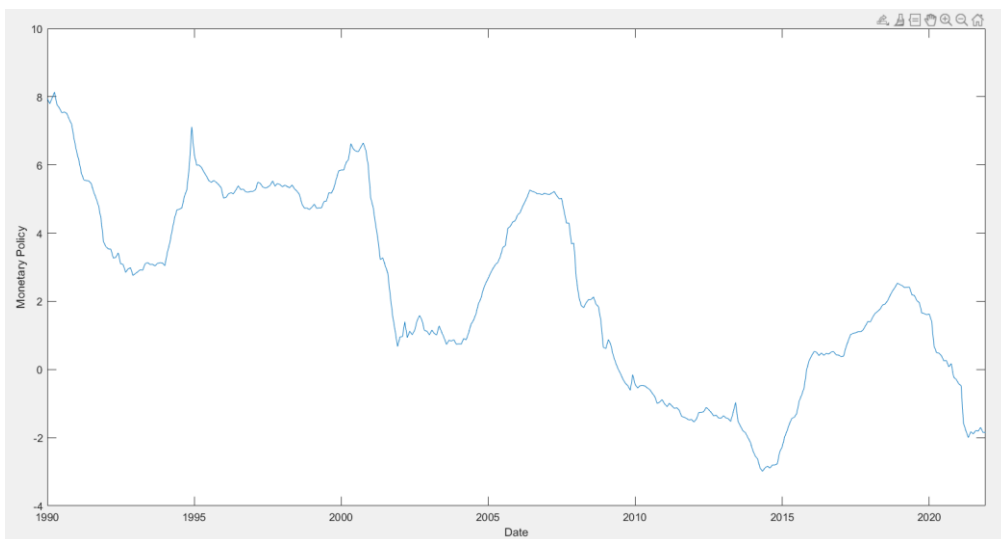


Figure 3.7. Wu-Xia Shadow Rate 1990-2021. Source: own elaboration with data from Federal Reserve Bank of Atlanta [8].

### 3.3. Natural Gas Historical Price

There are three main natural gas trading markets, the European average import border price, the Henry Hub spot price of the US, and the imported liquefied natural gas (LNG) price of Japan [2]. In this study, however, the Henry Hub spot price of natural gas of the US for the period of September 2000-December 2021 was the one used.

These markets trade based on contracts of various lengths (spot, short-term, medium-term, and long-term contracts). U.S. natural gas prices tend to be set by reference to the price of natural gas at the Henry Hub (HH), European natural gas prices are usually set by reference to the price of natural gas at either the National Balancing Point (NBP) in the UK or the Title Transfer Facility (TTF) in the Netherlands and Asian LNG prices are typically set by reference to the price of the Japan Customs-cleared Crude (known as the ‘Japanese Crude Cocktail’ or JCC) [27].

The differences between these reference prices have always existed, but they became greater with the increase in global spot natural gas trading activity. This trading activity takes place when the buyer agrees to pay a negotiated price for the natural gas to be delivered by the seller at a specified delivery point [27].

Natural gas spot prices reflect daily supply and demand balances and can be volatile. Over the past 12 months, for example, there have been sharp fluctuations, with all natural gas markets hitting record lows. The drop in the prices was due to the slump in demand, as a result of the Covid-19 crisis, as well as to the abundant supplies of LNG. However, as it can be appreciated in *Figure 3.8* and *Figure 3.9*, all prices recovered when the economic activity resurged [28].

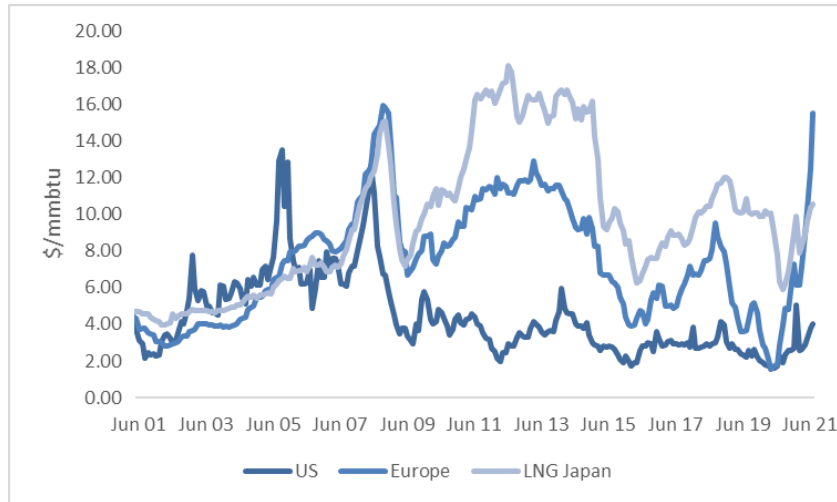


Figure 3.8. Historical nominal prices of natural gas (2001-2021) [25].

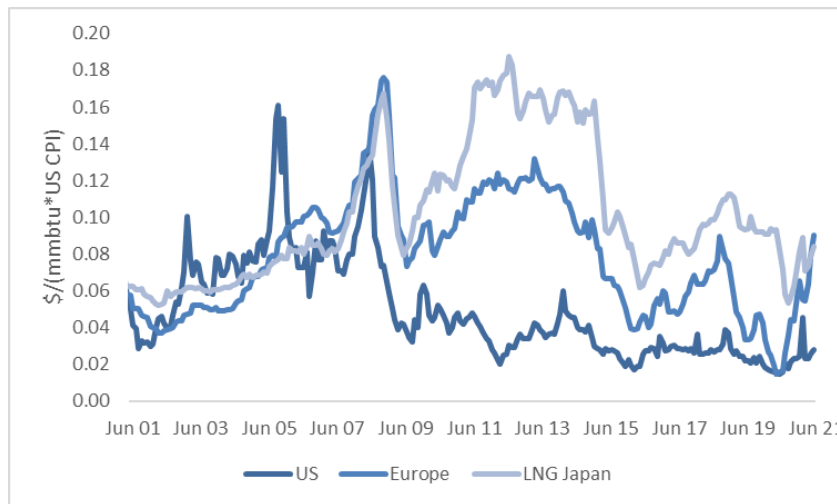


Figure 3.9. Historical real prices of natural gas (2001-2021). Deflated with US CPI [25],[26].

As mentioned above, for this study, the Henry Hub reference was used. Henry Hub is a pipeline that transports natural in Erath, Louisiana and serves as the official delivery location for futures contracts on the New York Mercantile Exchange (NYMEX), because of its central location and its high degree of interconnectedness. Sabine Pipeline LLC is the proprietary of the hub which has access to many of the major gas markets in the United States and is interconnected with 13 different intra and interstate pipelines [29].



The NYMEX contract for deliveries at Henry Hub began trading in 1990 and is deliverable 18 months in the future. The settlement price at Henry Hub constitutes a benchmark for North American natural gas market and part of the global LNG market [29].

The Henry Hub spot prices were retrieved from the U.S. Energy Information Administration (EIA) website. The historic price of natural gas is then used to build the forecast model along with the changes of other seven factors that we considered have affected the price of natural gas historically.

# Chapter 4

## Methodology I: Autoregressive Model

To forecast natural gas prices, we first use an autoregressive model (ARX), that includes initially the eight variables described in the previous chapter and forecasts the price of natural gas one period ahead ( $t+1$ ).

In this chapter, the methodology followed to obtain the parameters needed to construct the model and the forecasting techniques used in this study are described.

### 4.1. Benchmark Model

In order to determine if including the variables that might affect to natural gas prices in the forecasting exercise delivers a better prediction of natural gas prices than when taking into account only the price for the commodity in previous periods, a benchmark model is used. This benchmark model consists of a first order autoregressive model, in which the price of natural gas for a period ahead is dependent on an independent term ( $\alpha$ ) and the price for natural gas in the current period multiplied by a scaling parameter ( $\beta$ ).

An autoregressive model relates a time series variable to its past values and predicts future behavior based on past behavior. This model specifies that the output variable depends linearly on its own previous values and on a stochastic term which is imperfectly predictable. The simplest autoregressive

model (first order) uses only the most recent outcome of the time series observed to predict future values. The model is noted as AR(1) and defined by the following equation:

$$y_{t+1} = \alpha + \beta_1 y_t + \varepsilon_{t+1} \quad \varepsilon_t \sim (0, \sigma^2)$$

*Equation 5. Benchmark Model (1)*

where  $\varepsilon_{t+1}$  is white noise,  $\alpha$  is a constant, and  $\beta_1$  is the parameter of the model (1).

There are three main natural gas trading markets, the European average import border price, the Henry Hub spot price of the US, and the imported liquefied natural gas (LNG) price of Japan. In this model, however, the Henry Hub spot price of natural gas of the US was the one used (2).

To estimate the alpha and beta of the model, the relation between natural gas observed price in a period and the ones observed before is calculated from September 2010 to December 2021. For this, data of natural gas prices retrieved from the Energy Information Administration (EIA) website during the period of September 2000 to December 2021 is used. This way,  $y(t+1)$  and  $y(t)$  are known, and after performing a regression, the values for alpha and beta for each period are obtained.

The exercise is carried out for the sample spanned from September 2010 to December 2021, using the natural gas price since September 2000. To simulate the situation of a real forecaster, only data available before September 2010 is used for an in-sample estimation of parameters, which is then updated recursively each period as new information is obtained for an out-of-sample estimation [1].

The recursive regression is performed with MATLAB and the function used is *regress(Y,X)*, where Y is a vector that takes for each month, starting August 2010, the monthly natural gas prices recorded until the moment of the calculation

(since September 2000).  $X$  is a matrix that contains a first column of ones and a second column that, starting September 2010, takes monthly natural gas prices recorded until a month before the moment of the calculation, also since September 2000. This means that  $X$  takes the lag of the observed data.

The regression is calculated for each month beginning in September 2010 and ending in December 2021 (135 times). For each of these calculations, the alphas and betas obtained are stored in a matrix called `BM_Parameters` with size  $135 \times 2$ . Finally, taking the alpha, beta, and natural gas observed price for each period from September 2010 to December 2021, we calculate the natural gas price forecast for the following month. This way, a vector under the name of `BM_Forecast` is created to store the forecasts for each of the periods calculated (September 2010-January 2022). The forecasting performance of this benchmark model will then be compared to the ARX model.

For September 2021, for example, we obtain a beta of 0.9347 and an alpha of 0.2838. Therefore, the equation of the benchmark model to forecast the price of natural gas for October 2021, results in the following:

$$y_{t+1} = 0.2838 + 0.9347 y_t + \varepsilon_{t+1} \quad \varepsilon_t \sim (0, \sigma^2)$$

*Equation 6. Benchmark Model with estimated alpha and beta.*

Using this equation and the observed price for natural gas in September 2021 (5.1111), the price of natural gas for October 2021, would be:

$$y_{t+1} = 0.2838 + 0.9347 * 5.1111 + \varepsilon_{t+1} = 5.0611 + \varepsilon_{t+1}$$

and, considering that the recorded price of natural gas for October 2021 on the EIA website is 5.4780, the stochastic term in this example would be 0.4169.

In subchapter *Forecasting Statistics*, the forecasting performance of this model will be evaluated and compared to the ARX model by the calculation of

the out-of-sample  $R^2$  statistic, which compares the error made in each of the forecasts using this model and the ARX model, by comparing the forecast and the actual historical price given by the EIA.

## 4.2. ARX

An Autoregressive Exogenous model (ARX) is an autoregressive model which has exogenous inputs. This means that the model relates the current value of a time series to both past values of the same series, and current and past values of the driving (exogenous) series, which are externally determined series that influences the series of interest. In addition, the model contains an error ( $\varepsilon_{t+1}$ ) term which depicts that the current value of the time series cannot be predicted exactly, even if the other terms are known with precision (3).

In our first model, the times series being predicted is the price of natural gas ( $y_{t+1}$ ), and this prediction depends on the price of natural gas a period before the prediction ( $y_t$ ), and the year-to-year or month-to-month variation of seven other variables (*see Chapter 2*) also a month before the prediction ( $\Delta f_{1t} \dots \Delta f_{8t}$ ).

$$y_{t+1} = \alpha + \beta_1 y_t + \beta_2 \Delta f_{1t} + \dots + \beta_8 \Delta f_{8t} + \varepsilon_{t+1} \quad \varepsilon_t \sim (0, \sigma^2)$$

*Equation 7. ARX Model (1).*

To estimate the parameters of the model, the alpha and the eight betas, similarly as with the benchmark model, the data of natural gas prices retrieved from the Energy Information Administration (EIA) website during the period of September 2000 to December 2021 is used, and also the monthly or yearly variation of the other seven variables with the corrections explained in *Chapter 2*, for the same period of time. This way,  $y_{t+1}$ ,  $y_t$ , and  $\Delta f_{1t} \dots \Delta f_{8t}$  are known, and after performing a regression, the values for alpha and the eight betas for each of the periods between September 2010 and January 2022 are obtained.

Like the benchmark model, the exercise is carried out for the sample spanned from September 2010 to December 2021, using the data available for the price of natural gas and the variation of the other seven variables since September 2000. To simulate the situation of a real forecaster, only data available before September 2010 is used for an in-sample estimation of parameters, which is then updated recursively each period as new information is obtained for an out-of-sample estimation [1].

As explained in the subchapter *Benchmark Model*, this is performed with a MATLAB function called *regress(Y,X)*, where Y is a vector that takes for each month, starting September 2010, the monthly natural gas prices recorded until the moment of the calculation (since September 2000). X in this model is a matrix that contains a first column of ones, a second column that, starting September 2010, takes monthly natural gas prices recorded until a month before the moment of the calculation, also since September 2000, and seven other columns that gather the month-to-month or year-to-year variations of the other seven variables (REDTI, drilling activity, natural gas production, industrial production, electricity consumption, natural gas inventories, and monetary policy shadow rate) during that same period.

The regression is calculated for each month beginning in September 2010 and ending in December 2021 (135 times). For each of these calculations, the alpha and betas obtained for each of the variables are stored in a matrix called *ARX\_Parameters* with size 135x9. Finally, taking the alpha, eight betas, and natural gas observed price for each period from September 2010 to December 2021, we calculate the natural gas price forecast for the following month. This way, a vector under the name of *ARX\_Forecast* is created to store the forecasts for each of the periods calculated (September 2010-January 2022).

For September 2021, for example, we obtain the following alpha and betas:

$$\alpha = 0.2178; \beta_1 = 0.9481; \beta_2 = -0.0011; \beta_3 = -0.1841;$$

$$\beta_4 = -0.0088; \beta_5 = 8.0844; \beta_6 = 1.3883; \beta_7 = -0.0778; \beta_8 = -0.0057$$

The model, therefore, can be re-defined with the following equation:

$$y_{t+1} = 0.2178 + 0.9481y_t - 0.0011\Delta f_{1t} - 0.1841\Delta f_{2t} - 0.0088\Delta f_{3t} + 8.0844\Delta f_{4t} \\ + 1.3883\Delta f_{5t} - 0.0778\Delta f_{6t} - 0.0057\Delta f_{7t} + \varepsilon_{t+1} \quad \varepsilon_t \sim (0, \sigma^2)$$

*Equation 8. ARX model with estimated alpha and betas.*

This is not the only ARX model we built. After comparing the ARX model performance to the benchmark model and determining the significance of the forecasts, various models with different combinations of variables are built and their forecasting performance is compared to determine how the model improves depending on the variables used to forecast the natural gas price.

### 4.3. Forecasting Statistics

The forecasting performance of each model is evaluated by the out-of-sample  $R^2$  statistic, suggested by Campbell and Thompson [4] to compare the  $\hat{y}_{t+1}$  and  $\bar{y}_{t+1}$  forecasts at a 1-month horizon.  $\hat{y}_{t+1}$  is the forecast using the predictive model described by *Equation 5* and  $\bar{y}_{t+1}$  is the forecast based on the benchmark model in *Equation 4*. This  $R_{OS}^2$  statistic is given by:

$$R_{OS}^2 = 1 - \frac{\sum_{t=t_0-1}^{T-1} (y_{t+1} - \hat{y}_{t+1})^2}{\sum_{t=t_0-1}^{T-1} (y_{t+1} - \bar{y}_{t+1})^2}$$

where  $y_{t+1}$  is the realized and real value of the price of natural gas,  $t_0$  is the start of the forecasting sample (September 2010) and  $T$  is the end of the sample (December 2021). The  $R_{OS}^2$  measures the reduction in Mean Squared Prediction Error (MSPE) for the predictive regression model relative to the

benchmark model, in percentage terms. Therefore, if  $R_{OS}^2 > 0$ , the predictive model forecasts ( $\hat{y}_{t+1}$ ) outperform the benchmark forecasts ( $\bar{y}_{t+1}$ ).

After performing this calculation in Matlab, we obtain a value of  $R_{OS}^2$ , which for the model that contains all eight variables is -0.0315. Therefore, the predictive model forecasts ( $\hat{y}_{t+1}$ ) do not outperform the benchmark forecasts ( $\bar{y}_{t+1}$ ) for this model. This is why we build other models using only some of the variables. If we use the variables that we found out to be most significant in the *TFG – Modelling the natural gas market* [2], which are REDTI, natural gas gross withdrawals, natural gas storage levels and electricity end-use, we obtain a  $R_{OS}^2$  of +0.0348. In the table below we summarize all the models tested and the  $R_{OS}^2$  obtained for each of them.

We then examine whether the results are statistically significant by testing against the null hypothesis of equal MSPE between the two models. The benchmark model, however, is nested in the predictive regression model, which makes the parameter  $\beta_2$  in Equation 5 zero in the population under the null. This produces an upward bias in the estimation of the MSPE of the predictive regression model produced by such parameter. Therefore, we estimate the MSPE-adjusted statistic developed by Clark and West (2007). This statistic adjusts for the bias by deducting the square difference in the point predictions generated by each model.

We first define

$$\varepsilon_{t+1} = (y_{t+1} - \bar{y}_{t+1})^2 - [(y_{t+1} - \hat{y}_{t+1})^2 - (\hat{y}_{t+1} - \bar{y}_{t+1})^2]$$

and then regress  $\{\varepsilon_{t+1}\}_{s=t_0}^{T-1}$  on a constant and calculate the t-statistic under the null hypothesis that the constant is zero. The p-value is obtained with a standard normal distribution, for a one-tailed test.



If this statistic is less than 10% the results of the model can be considered to have a statistical significance higher than 90%. For the model that uses all eight variables, we obtain a Clark and West statistic or p-value of 0.0512, which means that the model has a significance of nearly 95%.

In the table below we summarize the  $R_{Os}^2$  and Clark and West statistics for each model and rank them according to its forecasting performance compared to the benchmark model and statistical significance.

Rank	REDTI	Drilling Activity	Production	World Industrial Production	Electricity Consump.	Inventories	Monetary Policy	$R_{Os}^2$	Clark & West
Model 1	X		X		X	X		<b>0.0348</b>	<b><math>7.894 \cdot 10^{-5}</math></b>
Model 2	X							<b>0.0389</b>	<b><math>2.827 \cdot 10^{-5}</math></b>
Model 3	X	X	X		X	X		<b>0.0311</b>	<b><math>2.687 \cdot 10^{-4}</math></b>
Model 4	X		X			X		<b>0.0300</b>	<b>0.0010</b>
Model 5	X		X	X	X	X		-0.0076	<b><math>3.147 \cdot 10^{-4}</math></b>
Model 6	X	X	X	X	X	X		-0.0155	<b>0.0020</b>
Model 7	X		X			X	X	-0.0181	<b>0.0168</b>
Model 8	X	X	X	X	X	X	X	-0.0315	<b>0.0512</b>

*Table 1. Summary of R2Os and Clark & West for various models including different combinations of variable.*

# Chapter 5

## Methodology II: Machine Learning

Machine learning is a branch of artificial intelligence that allows machines to learn without being specifically programmed to do so. These machines or systems identify patterns among data and make predictions [37].

Deep learning, a subcategory of 'machine learning' first described in 1960, allows systems to not only learn from experience, but to also train themselves to use data in the most efficient way. The three main currents used to train these models are:

- *Reinforcement learning* occurs when a machine learns by trial and error until it reaches the best way to complete a given task. The system is given "rewards" for solving the assigned task correctly, without specifically programming it to perform it in a certain way. Therefore, the system modifies its behavior according to the rewards it obtains.
- *Supervised learning* takes place when machines are trained with labeled data, such as photos with descriptions of the items in them. This way, once the machine is trained, the algorithm is able to select those labels from other databases.
- *Unsupervised learning* occurs when machines look for similarities in unlabeled databases, instead of identifying patterns in labeled ones. In this case, the algorithms are not programmed to detect a specific type of data, but to look for items that look similar and can be grouped together. Facial

recognition is an example of unsupervised learning, where the algorithm searches for a series of common patterns in a face to recognize it every time.

In the last few decades, the real-time execution capacity of computers and their processing speed have grown dramatically, drastically reducing the cost of computation. Also, systems to store and process huge amounts of data, known as big data (BD), have been developed. These two phenomena have led to the emergence of new prediction procedures based on supervised Machine Learning (ML) [36].

Even though there is no well-defined boundary between the so-called traditional prediction procedures and ML procedures, the main difference between the two resides in that the latter use statistical models and flexible algorithms that allow extracting information from high-dimensional data. In general, these ML procedures do not try to represent the mechanism that generated the data, they just predict and do not try to explain [36]. The ML or deep learning methods can improve prediction accuracy when there is a lot of data, but traditional models can be more efficient for solving simpler problems, or problems that by their nature require analytical solutions [37].

ML algorithms are computationally intensive and are designed to systematically identify patterns and relationships in the data that serve as the basis for making prediction [36]. In this thesis we focus on prediction procedures for dynamic data in which the lags of the variable to be predicted themselves may appear among the predictors. In the following subchapters we briefly review the main empirical applications of ML-based forecasting procedures in the context of economic and financial time series.

## 5.1. The prediction problem

The objective is to obtain predictions of  $y_{T+h}$ , where  $h$  is the prediction horizon, based on observations  $\{y_1, \dots, y_T\}$ , normally obtained in equidistant spaces of time. In our case, the prediction is univariate since the objective is the prediction of the future value of a single variable (natural gas price). In addition to observations, one can have a set of predictors,  $x_{i,t}$ , where  $i=1, \dots, N$ ,  $t=1, \dots, T$  [36]. The prediction problem is to obtain  $h$ -periods-forward predictions by the following straightforward procedure:

$$\hat{y}_{T+h|T} = g_h(y_T, \dots, y_1, x_{1,T}, \dots, x_{1,1}, \dots, x_{N,T}, \dots, x_{N,1})$$

ML prediction procedures require selecting an architecture and/or some parameters, called hyperparameters. For example, in Artificial Neural Networks (ANN) procedures, one must select the number of nodes and layers and in regularization procedures, one has to pre-select the regularization parameters) [36].

These hyperparameters reduce the complexity of the model and must be selected prior to the actual estimation of the parameters of the model. The sample needs to first be divided into three subsamples: the training subsample, in which, given the hyperparameters, the parameters are estimated; the validation subsample, in which the hyperparameters are chosen by minimizing a given loss function; and the contrast subsample, in which the predictions are evaluated [36].

In a large majority of ML prediction applications, hyperparameters are selected by cross-validation and predictions are evaluated using the mean square prediction error (MSE). To allow finding patterns in the data, both the training and test samples must be large enough [36].

1	+ + $T_1 \quad T_1+1$	+ $T$	+ $T_1+1$	$T+H$
---	--------------------------	----------	--------------	-------

<p>Training sample (estimation): Given the hyperparameters, the model is estimated.</p>	<p>Validation sample (model selection): <b>Selection of hyperparameters</b> that minimize the loss function of the predictions obtained with the estimated model in the training sample.</p>	<p>Contrast sample (prediction): Evaluation of out-of-sample predictions: EPCM.</p>
<p>Hyperparameters:</p> <ul style="list-style-type: none"> <li>- <b>Regularization:</b> Constants</li> <li>- <b>Trees:</b> Number of branches</li> <li>- <b>Neural Networks:</b> Activation function.</li> </ul>	<p>Loss function: Mean square prediction error (MSE). Procedure: Cross-validation</p>	

Figure 5.1. ML Prediction Procedures [36].

ML prediction procedures can be classified into two main groups depending on whether they try to reduce the dimension of the predictors (regression models with regularization and the SVM-based procedures) or whether they try to represent complex functions (decision tree models, neural networks, or deep learning) [36]. The most popular procedures within each of these two groups are briefly described below.

## 5.2. Dimensionality reduction: regression with regularization and Support Vector Machine

Sometimes, before estimating, it is necessary to regularize, that is, to force the parameters of some predictors to take relatively small values or to select some predictors, in order to recover degrees of freedom and improve the predictive capacity of the model. This is done when the number of predictors in a regression model is very high and some of them may be redundant or not have enough information to be included in the predictive model. When the function  $g$  is linear, the predictive model is the following regression model:

$$y_{T+1} = \beta' z_T + \varepsilon_{T+1}$$

Where  $z_T = (y_T, \dots, y_{T-p}, x_{1,T}, \dots, x_{1,T-p}, \dots, x_{N,T}, \dots, x_{N,T-p})$  is the vector of predictors, with  $p$  being the number of lags in the model.

A convex penalty to force the parameters in the vector  $\beta$  toward zero is introduced to estimate the  $(N+1) \times p$  unknown parameters  $\beta$ . This is known as regression with regularization. Among the most popular regularized regression procedures are ridge regression which reduce overfitting but do not select predictors, so they do not generate a more interpretable model. Alternatively, with the Least Absolute Shrinkage and Selection Operator (LASSO), both parameter reduction and variable selection occur, improving both the accuracy of the predictions and the interpretability of the model by selecting only a subset of the available predictors. When both a high number of variables and their lags appear among the predictors the sparse group LASSO (sg-LASSO) is introduced, reducing the dimension both among variables and among their lags. Another popular regression procedure with regularization is the elastic net, in which a second penalty is introduced [36].

Finally, among the models we have SVM, a data mining procedure originally developed by Vapnik at Bell Labs in 1995 for classification. When applied to regression, SVM is referred to as Support Vector Regression (SVR) [36]. Support Vector Regression gives the flexibility to define how much error is acceptable in the model and finds an appropriate line to fit the data. Instead of minimizing the sum of the squared errors, the objective function of SVR is to minimize the coefficients, while the error term is handled in the constraints, where the absolute error is set to less than or equal to a specified margin( $\epsilon$ ). SVR is a powerful algorithm that allows us to choose how tolerant we are of errors, both through an acceptable error margin( $\epsilon$ ) and through tuning our tolerance of falling outside that acceptable error rate [45].

### 5.3. Decision trees

Decision trees are used to segment a range of predictor values into a finite number of subregions so that, within each subregion, the predictors are more homogeneous, and prediction can be performed more easily [36].

From each node of the decision tree only two branches can emerge. The thresholds are successively determined to minimize the MSE. Finally, the value of the predictors is used to guide us through the tree to one of the subregions (terminal nodes of the tree). This way the prediction of  $y_{t+1}$  is obtained [36].

The best predictors and delays are selected to minimize the MSE when building a decision tree. However, this process often results in trees with many branches that can generate overfitting and, consequently, suboptimal predictions. For this reason, in practice, the tree is "pruned" by introducing a penalty on the number of terminal nodes in the objective function [36]. The most common procedures to generate decision trees are explained below.

Bagging (Bootstrap Aggregation) is a procedure designed to generate decision trees by avoiding the large MSE that is usually observed in predictions obtained using decision trees. It is based on generating bootstrap samples of the predictors and building the corresponding decision tree on each of the bootstrap replicates. In the case of temporal data, the resampling to obtain bootstrap replicates of the training sample must be carried out with special care so as not to destroy the existing temporal dependence in the data [36]. There are several bootstrap techniques, such as the wild bootstrap which was proposed by Wu (1986) for regression models with independent but heteroskedastic errors.

Bülmann and Yu (2002) have also proposed subsampling-based bagging (sub bagging) instead of Bootstrap and show that although it is computationally simpler, it has the same level of accuracy as Bagging. Subsampling-based

ensemble methods differ from traditional ensemble methods in that subsampling-based ensemble methods use a subset randomly drawn from the original learning sample without replacement to train each learning machine (e.g., regression tree) in the ensemble. In comparison, the conventional ensemble methods aggregate predictions based on repeated bootstrap samples where each sample is of the same size as the learning sample, and it is drawn from the learning sample with replacement. Subsampling-based ensemble methods use a smaller training set to train each individual learning machine, resulting in computational advantages [46].

Random Forest (RF), originally proposed by Breiman (2001), is a procedure designed to reduce the correlation between decision trees when they are generated by Bagging. The same predictors appear in all bootstrap replicates and decision trees are built with bootstrap replicates of the training sample. Bagging uses in each tree all existing regressors ( $p$ ), while RF selects randomly, only  $m < p$  of the  $p$  predictors, usually  $m = \sqrt{p}$ , being  $m$  another hyperparameter to be selected by cross validation in the training sample. Along with  $m$ , it is also necessary to determine other hyperparameters such as the number of Bootstrap replicates, the number of decision trees and the minimum number of observations at the terminal nodes (pruning) [36].

Finally, another procedure to generate decision trees is Boosting, which generates trees sequentially from a relatively simple initial tree. At each step, the fit of the previous tree is improved by constructing a new tree for the residuals instead of for the variable to be predicted directly. The tree is improved by improving specific areas where the previous tree fit did not work properly. At the end of the process, the prediction is a weighted average of the predictions of all the intermediate trees, giving more weight to the last trees, which are the best



calibrated. Again, the number of trees and the Boosting learning ratio must be calibrated by cross-validation [36].

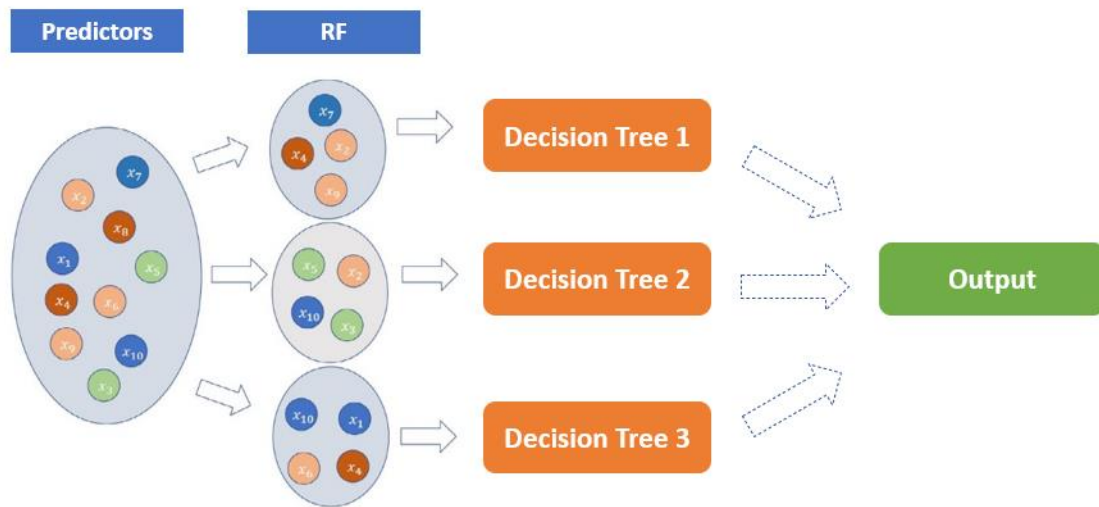


Figure 5.2. Decision Trees Selection Procedure through Random Forest [36].

## 5.4. Artificial Neural Networks

Artificial Neural Networks (ANNs) are data-driven procedures with very few prior assumptions about the models generating the data. They are computationally intensive mathematical models that attempt to mimic the way humans learn when they receive information. Their structure makes it possible to capture complex relationships between the predictors and the variable being predicted without the need to specify the correct form of these relationships, since the neural network itself will try to identify these relationships from the data available for training [36].

ANNs learn inductively through examples and mimic the ability to learn of humans [38]. Traditional approaches to time series forecasting assume that the time series under investigation is matched by a linear process, as opposed to ANNs [38].

The first prediction application using ANN appears in Hu's (1964) thesis for weather forecasting. Recent research has validated the good predictive results of ANNs by showing that, if the networks are sufficiently large, almost all local minima are very similar to the global minimum [36].

Among the numerous types of neural networks, one of the most popular is the one known as "multilayer feedforward networks". These networks consist of an input layer, with several nodes (or neurons) that merely receive the values of the predictors; an output layer related to the variable to be predicted; and, between the two, a series of hidden layers, each with several nodes that receive information only from the nodes of the immediately preceding layer and pass the information on to the nodes of the next layer. All the nodes of a layer are linked to each of the nodes of the next layer through values known as "weights". Finally, at each node it is necessary to define an activation function that transforms the information arriving from each of the nodes of the previous layer. Among the most popular activation functions are linear, exponential, and logistic [36]. In most cases, input level neurons have no activation function because their function is to transmit inputs to the hidden layers [38]. The most used activation function for the output level is the linear function since the nonlinear activation function can insert a distortion in the previous outputs. Logistic and hyperbolic functions are often used at the hidden layers [39].

The relationship between the output ( $y_T$ ) and the inputs ( $y_{t-1}, \dots, y_{t-p}$ ) of an ANN has the following mathematical representation:

$$y_t = w_o + \sum_{j=1}^q w_j \cdot g \left( w_{0j} + \sum_{i=1}^p w_{ij} \cdot y_{t-i} \right) + \varepsilon_t$$

where,  $w_{ij}$  ( $i= 0,1,2,\dots,p; j=1,2,\dots,q$ ) and  $w_j$  ( $j = 0, 1, 2, \dots, q$ ) are model parameters often referred to as connection weights;  $p$  is the number of input nodes; and  $q$  is the number of hidden nodes.

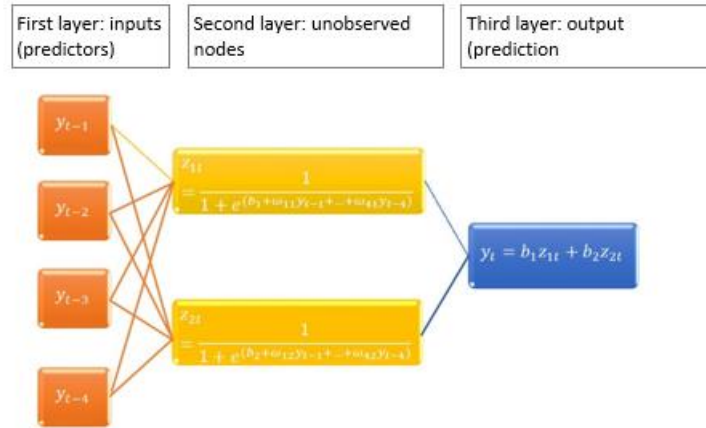


Figure 5.3. Neural Network with 4 lags and one layer with 2 nodes

Once the functional form and architecture of the network has been specified, the parameter estimation process can begin. All the hyperparameters appearing in the network are determined using an iterative estimation procedure to update the weights using the errors. Specifically, the error made at the final or exit node is distributed backwards through all the hidden nodes that reach that final node, such that each node is assigned a part of that error by iteratively adjusting the weights of each node. This sequential optimization procedure is known as "back propagation of error", and the hyperparameter that controls the change of the weights from one iteration to the next is known as the learning ratio, which is often set to 0.1 [36].

The goal is to estimate the parameters that minimize the cost function of the neural network, a standard measure of accuracy where the mean square error of both the training set and the test set is agreed [39].

$$E = \frac{1}{N} \sum_{n=1}^N (e_i)^2 = \frac{1}{N} \sum_{n=1}^N \left( y_i - \left( w_0 + \sum_{j=1}^Q w_j g \left( w_{0j} + \sum_{i=1}^P w_{ij} y_{t-4} \right) \right) \right)^2$$

where  $N$  is the number of error terms. In the basic backpropagation training algorithm, the neural network parameters ( $w_{ij}$ ) are modified in  $\Delta w_{ij}$ , according to the following formula:

$$\Delta w_{\delta j} = -\eta \frac{\partial E}{\partial w_{ij}}$$

where, the parameter  $\eta$  is the learning rate and  $\frac{\partial E}{\partial w_{ij}}$  is the partial derivative of the function  $E$  with respect to the weight  $w_{ij}$ . This derivative is calculated in two steps. In the forward step, an input vector from the training set is introduced in the input units of the network and propagated through it, layer by layer, producing the final output. During the backward step, the obtained output is compared to the desired output and the resulting error is then propagated backward through the network by adjusting the weights in each process. To speed up the learning process and avoid the instability of the algorithm, the  $\Delta w_{ij}$  is defined as:

$$\Delta w_{ij}(t+1) = -\eta \frac{\partial E}{\partial r_{ij}} + \delta \Delta w_{ij}(t)$$

After this, the estimated model is evaluated using a separate reserve sample that is not exposed to the training process [39].

A specific type of these neural networks are recurrent neural networks (RNN) are networks that were conceived in the 1980s but did not become popular at that time due to their high computational requirements. Recurrent networks have a certain form of memory, in contrast to the networks described above, which have an activation function that only acts in one direction, forward, from

one layer to the next, and do not remember previous values. Each recurrent neuron has two sets of hyperparameters, one applied to the input data it receives from the previous layer and another set applied to the output vector of the previous instant. This way, the structure of these networks allows a recurrent connection in time [36].

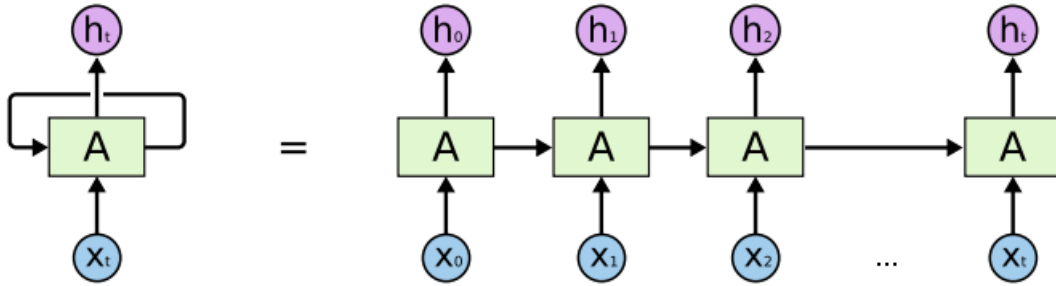


Figure 5.4. Example of RNN [41]

Fully recurrent neural networks connect the outputs of all neurons to the inputs of all neurons. This is the most general most general neural network topology because all other topologies can be represented by setting some connection weights to zero to simulate the lack of connections between those neurons [44.]

Long Short-Term Memory (LSTM) is a recurrent artificial neural network architecture that can process not just a single data point (such as an image), but the entire data stream (such as audio, video, or any series) [40]. In this type of architecture, the model learns to maintain its "cellular memory" state over time, adding one or more new gates that allow it to add or remove information in the "cell memory" [43].

The internal memory unit of the LSTM records historical information up to the present time and is controlled by three gates: input, forget and output. The input gate controls the new information entered into the internal memory unit, the forgetting gate controls how much information from the previous

moments needs to be saved, and the output gate produces and reproduces the required information [39].

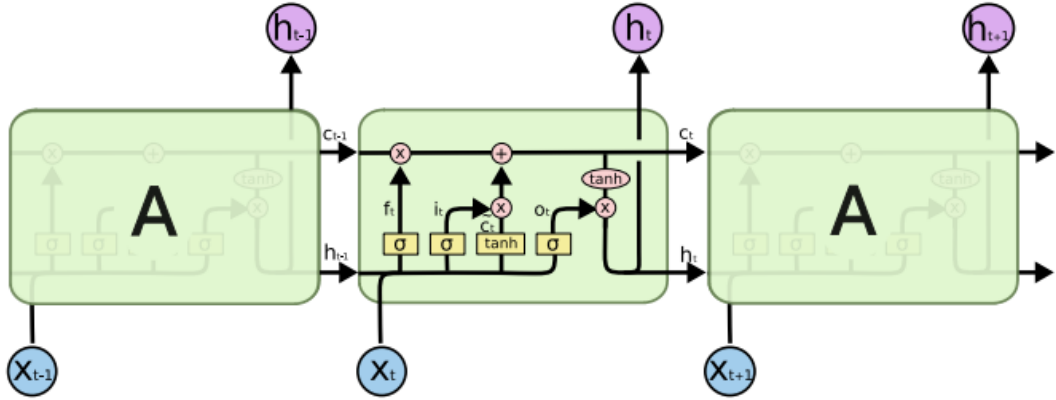


Figure 5.5. Example of LSTM cell [41]

Networks need many observations to learn and making predictions outside the range of the variables' values the networks have learned with may result in bad properties. In addition, convergence of the estimation process to a global minimum is often complicated, so the search process must be initialized at different random points. Neural networks can have problems related to overfitting and large computational effort. To solve these problems, procedures such as Extreme Learning Machine (ELM) have been proposed [36].

## 5.5. Tensor Flow and Keras

Implementing a successful complex deep neural network model is not an easy task and was often limited to deep learning practitioners. Fortunately, several software frameworks recently became available where the design and training of deep neural networks is easier [39].

Given the increase in popularity of deep learning, Google researchers developed and launched TensorFlow, a deep learning framework and open-source software [39]. TensorFlow constitutes a stable platform for deep learning applications and research, providing scientists and practitioners with a

standardized system that has the potential to improve the reproducibility of deep learning techniques [40]. TensorFlow also offers a variety of enhancements, including improved graph rendering and compiling time.

TensorFlow uses a declarative programming model that allows researchers to focus on the symbolic definition of what needs to be computed or modeled, rather than the exact method or the specific order in which these computations should be performed [41]. This model can be optimized for digital stability and performance, and individual parts can be converted to be executed by the CPU, GPU, or other TPUs [42]. The processing framework automatically manages the abstract representation of the model, which makes TensorFlow suitable for developing new models using gradient-based optimization [39].

Another of these types of deep learning software frameworks is Keras, a high-level open-source neural network application program interface written in Python language. Keras was originally developed by François Chollet, who now works at Google [43]. This interface can be run on several machine learning libraries such as TensorFlow (developed by Google), CNTK (developed by Microsoft) and Theano (developed by the University of Montreal). Keras aims to provide users with fast and easy prototyping through ease of use, modularity, and extensibility [41].

In 2018, Keras ranked second only after TensorFlow in effectiveness evaluations based on usage, interest, and popularity. It provides all standard loss functions, automatic backpropagation, and standard optimization algorithms [43].

Some other very widely used deep learning frameworks are Torch, Theano, and Caffe, which are particularly suitable for complex neural networks. Many startups and research groups have released deep learning tools such as Neon (a

Python-based framework), Deeplearning J (Skymind's Java engine) and H2O3 (Skymind in Java-based machine learning) [42].

<code>keras_model()</code>	Create model
<code>keras_model_sequential()</code>	Create model composed of a linear stack of layers
<code>compile()</code>	Set Keras model for training
<code>fit()</code>	Initiate training
<code>evaluate()</code>	Evaluate Keras model
<code>predict()</code>	Prediction method for Keras model
<code>summary()</code>	Prints a model summary
<code>save_model_hdf5()</code>	Save/ load models using HDF5 files
<code>get_layer()</code>	Retrieves layer according to its name
<code>pop_layer()</code>	Delete the last layer of the model
<code>save_model_weights_hdf5()</code>	Save model weights using HDF5 files
<code>load_model_weights_hdf5()</code>	Load model weights using HDF5 files
<code>get_weights()</code> <code>set_weights()</code>	Retrieve or set model weights
<code>get_config()</code> <code>from_config()</code>	Retrieve layer/ model configuration
<code>model_to_json()</code>	Set up model as JSON
<code>model_to_yaml()</code>	Set up model as YAML

*Table 2. Keras Functions*

## 5.6. Forecasting Natural Gas Prices with Machine Learning

In this chapter, we will use machine learning algorithms to train the machine on historical price records and predict the expected future natural gas price. We will use the environment of Anaconda distribution, an open-source Python distribution platform.



We will predict the Natural Gas Price for the period of September 2010 to January 2022, similarly to what we did with the Autoregressive model, and compare the predictions with the real values obtained at EIA's website. This prediction will be done using Time Series with Long short-term memory (LSTM) Neural Networks. We use Natural Gas historical monthly price downloaded from EIA's website (51) from September 2000 to December 2021. As software frameworks, we will use TensorFlow and Keras.

Throughout this application, stochastic gradient descent with backpropagation of errors to train deep neural networks will be used. Stochastic Gradient Descent (SDG) is an optimization algorithm used to train machine learning algorithms. Its job is to find a set of internal model parameters that perform well against some performance measure such as logarithmic loss or mean squared error. In our models we use mean squared error. The algorithm is iterative, which means that the search process occurs over multiple discrete steps. Each step uses the model with the current set of internal parameters to make predictions on some samples, comparing the predictions to the real expected outcomes, calculating the error, and using the error to update the internal model parameters. As an update procedure we use the backpropagation update algorithm, explained in the previous subchapters.

After importing all the packages that we are going to use during the programming of the model, we import the historical gas price since 2000 from the EIA Website and plot it (Figure 5.6).

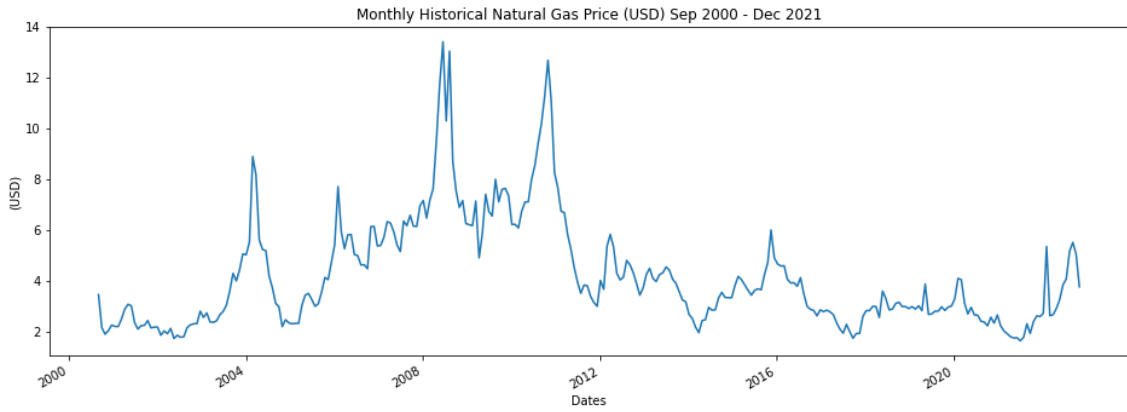


Figure 5.6. Historic price of natural gas since 2000.

Then we train the LSTM model, separating the data sample into a training set and a test set. We use 20% of the data for the test set and 80% for the training set. We then define to get natural gas price 6 months back and predict the following month's price, this defines our number of time-steps to use in the LSTM models.

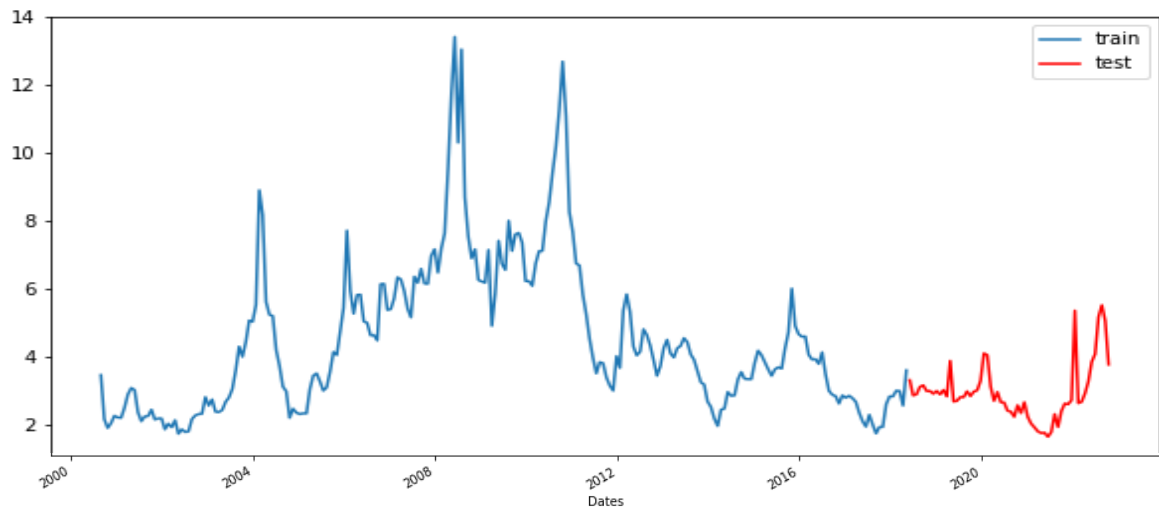


Figure 5.7. Test set vs. training set data split.

Next, we build an LSTM model with 4 hidden layers, each with 200 neurons, and one output layer. As activation function, we use the rectified linear activation function (ReLU), which is a piecewise linear function that will output the input directly if it is positive, otherwise, it will output zero. It looks and acts like a linear function, but is, in fact, a nonlinear function allowing complex relationships in the data to be learned. ReLU has become the default activation

function for many types of neural networks because a model that uses it is easier to train and often achieves better performance.

$$g(z) = \max\{0, z\}$$

We set the output layer as a dense layer with one neuron and the number of epochs to 50. The number of epochs is a hyperparameter that defines the number of times that the learning algorithm will work through the entire training dataset. One epoch means that each sample in the training dataset has had an opportunity to update the internal model parameters.

As optimizer we use Adam, proposed by *Kingma and Lei Ba in Adam: A Method For Stochastic Optimization*. “Adam is a method for efficient stochastic optimization that only requires first-order gradients with little memory requirement. The method computes individual adaptive learning rates for different parameters from estimates of first and second moments of the gradients; the name Adam is derived from adaptive moment estimation” [47]. Default parameters are those suggested in the paper,

- Learning rate (lr): float  $\geq 0$ .
- beta\_1, beta\_2: floats,  $0 < \beta < 1$ . Generally close to 1.
- Fuzz factor (epsilon): float  $\geq 0$ . Fuzz factor.

As loss function, we use the mean squared error (MSE), which is calculated as the average of the squared differences between the predicted and actual values. This loss function compares the target and predicted output values and measures how well the neural network models the training data [48]. We also use *Early Stopping*, a form of regularization to avoid overfitting when training a learner with an iterative method, such as gradient descent. We set the patience to 16 iterations, meaning that after 16 iterations without improvement on the MSE, the training is stopped.

A dropout regularization technique is also used. Dropout is a regularization technique for neural network models where randomly selected neurons are ignored during training proposed by Srivastava et al. in their 2014 paper “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. Some neurons’ contribution to the activation of downstream neurons is temporally removed on the forward pass, and no weight updates are applied to the neuron on the backward pass. As a neural network learns some neurons become specialized and neighboring neurons come to rely on this specialization, which can result in a fragile model too specialized for the training data. If neurons are randomly dropped out of the network during training, other neurons have to step in and make predictions for the missing neurons. This results in the network learning multiple independent internal representations, less sensitive to the specific weights of neurons, able to better generalize and less likely to overfit the training data. [49].

For training the model, a `model.fit()` training loop checks at the end of every epoch whether the loss is no longer decreasing, considering the minimum delta and patience (16 in our model). Once the loss is found no longer decreasing the training terminates.

After training the model, we predict the natural gas price from September 2010 to January 2022 and compare it to the actual historical natural gas price for those months, retrieved from the EIA’s website (*Figure 5.10*).

Next, we compute two performance metrics:  $R^2$  (variance score) and Mean Squared Error (RMSE, test score). For this, we import the package `math` and `skelearn.metrics`.

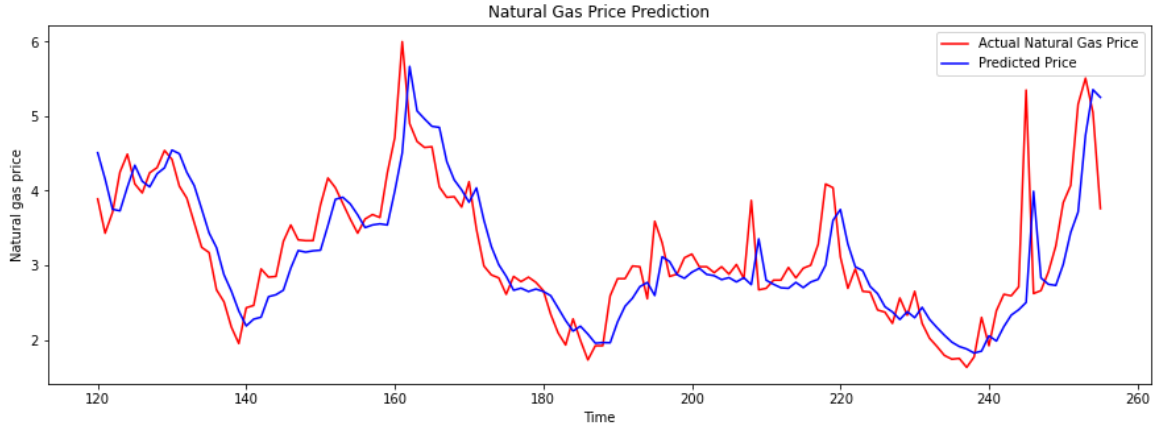


Figure 5.8. September 2010 – January 2022 Natural Gas Price Prediction using LSTM Model vs. Actual Natural Gas Price

R-squared is a statistical measure that represents the proportion of the variance for a dependent variable that is explained by an independent variable.

$$R^2 = 1 - \frac{\text{Unexplained Variation}}{\text{Total Variation}} = \frac{\sum_{i=1}^N (\text{Predicted Values} - \text{Actual Values})^2}{\sum_{i=1}^N (\text{Actual Values} - \text{Average Actual Value})^2}$$

where N is the sample size.

We obtain a variance score of 0.64, which means that 64% of the observed variation can be explained by the model inputs.

RMSE is the standard deviation of the residuals or prediction errors, it is a measure of how concentrated the data points are around the line of best fit.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\text{Predicted Values}_i - \text{Actual Values}_i)^2}{N}}$$

We obtain a test score of 0.52 RMSE, which means that the average error of our predictions is +/- 0.52€ with respect to the actual price.

We will compare these values to the ones obtained with the AR benchmark model, which also only uses historical natural gas price to predict natural gas price for the following month.

# Chapter 6

## Summary of Empirical Results

### 6.1. Natural Gas Price Forecasts with Autoregressive

#### Model

The first step in forecasting natural gas with the Autoregressive Model (AR) is to build a benchmark model whose prediction performance will then be compared to that of the ARX model. The parameters of the model are calculated by performing a regression for each month beginning in September 2010 and ending in December 2021 (135 times), since we have the data for natural gas prices for all these months. For each of these calculations, the alphas and betas obtained are stored in a matrix called `BM_Parameters` with size `135x2`. Finally, taking the alpha, beta, and natural gas observed price for each period from September 2010 to December 2021, we calculate the natural gas price forecast for the following month. This way, a vector under the name of `BM_Forecast` is created to store the forecasts for each of the periods calculated (September 2010-January 2022).

For September 2021, for example, we obtain a beta of 0.9347 and an alpha of 0.2838. Therefore, the equation of the benchmark model to forecast the price of natural gas for October 2021, results in the following:

$$y_{t+1} = 0.2838 + 0.9347 y_t + \varepsilon_{t+1} \quad \varepsilon_t \sim (0, \sigma^2)$$

*Equation 6. Benchmark Model with estimated alpha and beta.*

Using this equation and the observed price for natural gas in September 2021 (5.1111), the price of natural gas for October 2021, would be:

$$y_{t+1} = 0.2838 + 0.9347 * 5.1111 + \varepsilon_{t+1} = 5.0611 + \varepsilon_{t+1}$$

and, considering that the recorded price of natural gas for October 2021 on the EIA website is 5.4780, the stochastic term in this example would be 0.4169.

The ARX model is built similarly by performing a regression using data of historical monthly natural gas price and monthly change of other seven variables that affect natural gas prices (mentioned in previous chapters) starting September 2010 and ending in December 2021 (135 times). This way, the alpha and betas for each of the variables of the model are calculated and are stored in a matrix called ARX\_Parameters with size 135x9. Finally, taking the alpha, eight betas, and natural gas observed price for each period from September 2010 to December 2021, we calculate the natural gas price forecast for the following month. This way, a vector under the name of ARX\_Forecast is created to store the forecasts for each of the periods calculated (September 2010-January 2022).

For September 2021, for example, we obtain the following alpha and betas:

$$\alpha = 0.2178; \beta_1 = 0.9481; \beta_2 = -0.0011; \beta_3 = -0.1841 ;$$

$$\beta_4 = -0.0088; \beta_5 = 8.0844; \beta_6 = 1.3883; \beta_7 = -0.0778; \beta_8 = -0.0057$$

The model, therefore, can be re-defined with the following equation:

$$y_{t+1} = 0.2178 + 0.9481y_t - 0.0011\Delta f_{1t} - 0.1841\Delta f_{2t} - 0.0088\Delta f_{3t} + 8.0844\Delta f_{4t} \\ + 1.3883\Delta f_{5t} - 0.0778\Delta f_{6t} - 0.0057\Delta f_{7t} + \varepsilon_{t+1} \quad \varepsilon_t \sim (0, \sigma^2)$$

Equation 7. ARX model with estimated alpha and betas.

The forecasting performance of each model is evaluated by the out-of-sample  $R^2$  statistic, suggested by Campbell and Thompson [4] to compare the  $\hat{y}_{t+1}$  and  $\bar{y}_{t+1}$  forecasts at a 1-month horizon.  $\hat{y}_{t+1}$  is the forecast using the ARX model and  $\bar{y}_{t+1}$  is the forecast based on the benchmark model. This  $R_{OS}^2$  statistic is given by:

$$R_{OS}^2 = 1 - \frac{\sum_{t=t_0-1}^{T-1} (y_{t+1} - \hat{y}_{t+1})^2}{\sum_{t=t_0-1}^{T-1} (y_{t+1} - \bar{y}_{t+1})^2}$$

where  $y_{t+1}$  is the realized and real value of the price of natural gas,  $t_0$  is the start of the forecasting sample (September 2010) and  $T$  is the end of the sample (December 2021).

After performing this calculation in Matlab, we obtain a value of  $R_{OS}^2$  that compares the predictive regression model to the benchmark model predictions. In the case of the model that uses eight variables, this value is -0.0315, which means that the predictive model forecasts ( $\hat{y}_{t+1}$ ) do not outperform the benchmark forecasts ( $\bar{y}_{t+1}$ ).

In light of these results, we build different models using different combinations of variables. We start building a model using the four variables that appeared to be more significant when studied in the *TFG – Modelling the Natural Gas Market* [2], the REDTI, the U.S. gross withdrawals, the U.S. storage levels and U.S. electricity end-use. The  $R_{OS}^2$  for this model is significantly better, 0.0348, which means that the ARX model outperforms the benchmark model. We build some more models to identify the best combination of variables to predict natural gas prices.



We then examine whether the results are statistically significant by testing against the null hypothesis of equal MSPE between the two models. We estimate the MSPE-adjusted statistic developed by Clark and West (2007).

We first define

$$\varepsilon_{t+1} = (y_{t+1} - \bar{y}_{t+1})^2 - [(y_{t+1} - \hat{y}_{t+1})^2 - (\hat{y}_{t+1} - \bar{y}_{t+1})^2]$$

and then regress  $\{\varepsilon_{t+1}\}_{s=t_0}^{T-1}$  on a constant and calculate the t-statistic under the null hypothesis that the constant is zero. The p-value is obtained with a standard normal distribution, for a one-tailed test.

If this statistic is less than 10% the results of the model can be considered to have a statistical significance higher than 90%. For the model that uses all eight variables, we obtain a Clark and West statistic or p-value of 0.0512, which means that the model has a significance of nearly 95%.

Rank	REDTI	Drilling Activity	Production	World Industrial Production	Electricity Consump.	Inventories	Monetary Policy	$R_{0s}^2$	Clark & West
<b>Model 1</b>	X		X		X	X		<b>0.0348</b>	<b><math>7.894 \cdot 10^{-5}</math></b>
<b>Model 2</b>	X							<b>0.0389</b>	<b><math>2.827 \cdot 10^{-5}</math></b>
<b>Model 3</b>	X	X	X		X	X		<b>0.0311</b>	<b><math>2.687 \cdot 10^{-4}</math></b>
<b>Model 4</b>	X		X			X		<b>0.0300</b>	<b>0.0010</b>
Model 5	X		X	X	X	X		-0.0076	<b><math>3.147 \cdot 10^{-4}</math></b>
Model 6	X	X	X	X	X	X		-0.0155	<b>0.0020</b>
Model 7	X		X			X	X	-0.0181	<b>0.0168</b>
Model 8	X	X	X	X	X	X	X	-0.0315	<b>0.0512</b>

Table 3. Summary of  $R^2$ Os and Clark & West for various models including different combinations of variables.

In Table 3 we summarize the  $R_{0s}^2$  and Clark and West statistics for each model and rank them according to its forecasting performance compared to the benchmark model and statistical significance. We highlighted in bold the models

that both outperform the benchmark model and are statistically significant (90% significance or above).

## 6.2. Natural Gas Price Forecasts with Machine Learning

Using Time Series with Long short-term memory (LSTM) Neural Networks, we predicted Natural Gas price for the period of September 2010 to January 2022, similarly to what we did with the Autoregressive model, and compared the predictions with the real values obtained at EIA’s website. We use Natural Gas historical monthly price downloaded from EIA’s website (51) from September 2000 to December 2021. As software frameworks, we will use TensorFlow and Keras.

The parameters of the model are summarized in the following table:

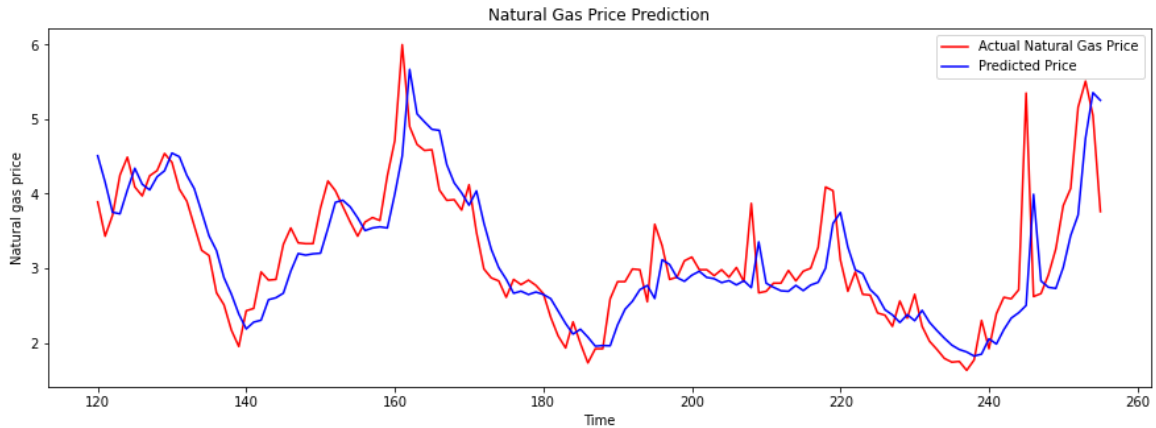
• Stochastic gradient descent with backpropagation error	✓
• Test vs. Training Set	20% vs. 80%
• Time-lags	6 months
• Number of hidden layers	4
• Number of neurons per layer	200
• Optimizer	Adam
• Activation Function	ReLU
• Output layer	Dense
• Number of epochs	50
• Loss function	MSE
• Dropout Regularization Technique	✓

*Table 4. Summary of Machine Learning Model Parameters.*

For training the model, a `model.fit()` training loop checked at the end of every epoch whether the loss was no longer decreasing, considering the minimum

delta and patience (16 in our model). Once the loss was found no longer decreasing the training terminates.

After training the model, we predicted the natural gas price from September 2010 to January 2022 and compare it to the actual historical natural gas price for those months, retrieved from the EIA’s website (*Figure 5.10*).



*Figure 6.1. September 2010 – January 2022 Natural Gas Price Prediction using LSTM Model vs. Actual Natural Gas Price*

Next, we computed two performance metrics:  $R^2$  (variance score) and Mean Squared Error (RMSE, test score). We obtained the following values:

<b><math>R^2</math></b>	0.64 (64%)	64% of observed variation explained by model inputs
<b>RMSE</b>	0.52€	Average error of predictions: +/- 0.52€

*Table 5. Summary of the Machine Learning model's performance metrics.*

We will compare in the next subchapter these values to the ones obtained with the AR model.

## 6.1. Results Comparison: AR Model vs. Machine Learning

In order to compare the performance of the Machine Learning model to the autoregressive models, the  $R^2$  (variance score) and Root Mean Squared Error (RMSE, test score) is calculated for the benchmark autoregressive model and the

ARX Model 1. We used Model 1 because it is the one with the best performance metrics.

As explained in *Chapter 5.6*, the formulas for  $R^2$  and Root Mean Squared Error are the following:

$$R^2 = 1 - \frac{\text{Unexplained Variation}}{\text{Total Variation}} = \frac{\sum_{i=1}^N (\text{Predicted Values} - \text{Actual Values})^2}{\sum_{i=1}^N (\text{Actual Values} - \text{Average Actual Value})^2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\text{Predicted Values}_i - \text{Actual Values}_i)^2}{N}}$$

We obtain the following values for all three models:

	Machine Learning	BM AR Model	ARX Model 1
$R^2$	0.64 (64%)	0.32 (32%)	0.34 (34%)
RMSE	0.52€	0.51€	0.50€

*Table 6. Performance metrics of the three different models*

In terms of the variance explained by the model, the machine learning model is way superior to the autoregressive model, explaining approximately 30% more variance than the other two models. However, in terms of RMSE, all models perform a very similar error in average when predicting natural gas, around +/- 0.50€.

Taking these performance metrics, the ARX model that includes, apart from the historical natural gas price, the variations in residential demand due to weather conditions, in natural gas production, in natural gas inventories and in electricity consumption, performs slightly better than the benchmark model that includes only historical natural gas prices to predict future prices.

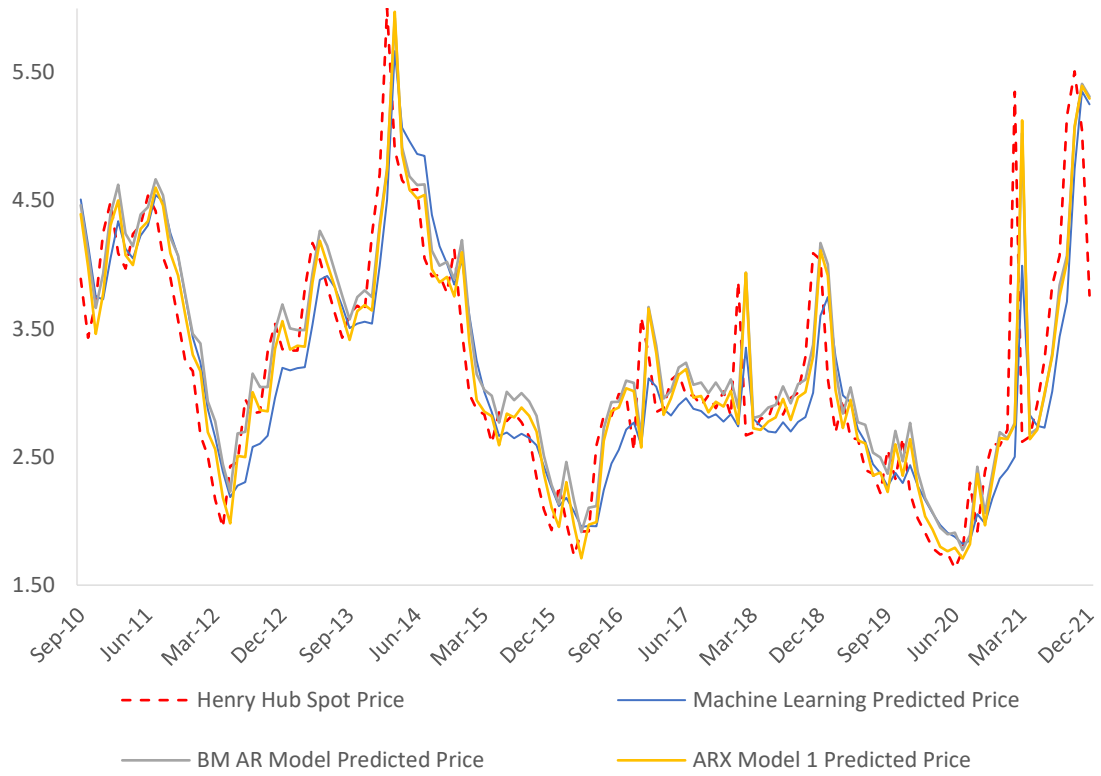


Figure 6.2. Machine Learning vs AR Model predictions vs Actual Price

# Chapter 7

## Conclusions

Aware of the length of this document, the most important conclusions and results are presented as a summary in this chapter, so that the reader can refer to the data of most interest to them without having to go through the entire document.

The price of natural gas has been a topic of especial interest in the past few months, due to the European energetic crisis caused by the outburst of the Russian-Ukraine conflict. Europe imposed economic sanctions to Russia to show disapproval for the invasion and this led to a reduction of natural gas supply coming from Russia, which resulted in the rocketing of natural gas prices in June 2022, which increased in 60% in only two weeks [2]. However, this paper uses data until December 2021, so it does not take into consideration the recent fluctuations in the natural gas market due to the political unrest.

Demand for natural gas is increasing due to its price-calorific value ratio. According to the U.S. Energy Information Association (EIA), the United States used about 31.5 quadrillion British thermal units (Btu) in 2020 which makes 34%

of U.S. total energy consumption [13]. The most common uses of natural gas are for electricity, heating, transportation, and production.

This paper first explained what natural gas is, how it is obtained, transformed, and transported, as well as how reserves, production, and demand are distributed around the globe, and what natural gas is used for (see *Chapter 2* above), as well as which factors might affect to the supply and demand of this commodity (see *Chapter 3* above). Then, having described which factors affect the demand and supply of gas and how indicators were built to use in the forecasting model, an autoregressive model is built to forecast natural gas prices for September 2010 to January 2022, to compare the forecasting results to the real values recorded at the EIA website and to determine the forecasting performance of the model (see *Chapter 4* above). Machine learning techniques are studied in *Chapter 5. Methodology II: Machine Learning* and one of them is then used to forecast natural gas prices for the same period. Finally, *Chapter 6* is dedicated to the summary of all the results obtained in the two previous chapters.

In order to model the market for natural gas and forecast natural gas prices using the autoregressive model, we gathered eight indicators or variables which could explain the variations in the price of this fuel. This selection is based on thorough research performed for the elaboration of the *TFG – Modelling the natural gas market* [2]. In this *TFG – Modelling the natural gas market* [2], it was found out that the variables that could explain the most part of natural gas price fluctuations were the residential demand due to change in temperatures, the levels of natural gas production and inventories held, and the amount of electricity consumed.

To forecast natural gas prices, we first use an autoregressive model (ARX), that includes initially the eight variables gathered in the *TFG – Modelling the*

*natural gas market* [2] and forecasts the price of natural gas one period ahead ( $t+1$ ). Then we include different combinations of the eight variables based on the results we obtained in the *TFG – Modelling the natural gas market* [2] and measure the forecasting performance of the model each time, trying to reach the most significant and most accurate forecasting model.

In order to determine if including the variables that might affect to natural gas prices in the forecasting exercise delivers a better prediction of natural gas prices than when taking into account only the price for the commodity in previous periods, a benchmark model is used. This benchmark model consists of a first order autoregressive model, in which the price of natural gas for a period ahead is dependent on an independent term (alpha) and the price for natural gas in the current period multiplied by a scaling parameter (beta). To simulate the situation of a real forecaster, only data available before September 2010 is used for an in-sample estimation of parameters, which is then updated recursively each period as new information is obtained for an out-of-sample estimation [1]. The recursive regression takes place for each month, starting August 2010, the monthly natural gas prices recorded until the moment of the calculation (since September 2000). The regression is then calculated for each month beginning in September 2010 and ending in December 2021 (135 times) and forecasts for these months are obtained and compared to the prices recorded in the EIA website for this period.

After estimating the benchmark model, an Autoregressive Exogenous model (ARX) is built, using initially all eight variables identified in *TFG – Modelling the natural gas market*. In this model, the times series being predicted is the price of natural gas ( $y_{t+1}$ ), and this prediction depends on the price of natural gas a period before the prediction ( $y_t$ ), and the year-to-year or month-to-month variation of seven other variables (*see Chapter 2*) also a month before the prediction ( $\Delta f_{1t} \dots \Delta f_{8t}$ ). Like the benchmark model, to simulate the situation of



a real forecaster, only data available before September 2010 is used for an in-sample estimation of parameters. The regression is calculated for each month beginning in September 2010 and ending in December 2021 (135 times) and taking the alpha, eight betas, and natural gas observed price for each period from September 2010 to December 2021, we calculate the natural gas price forecast for the following month, for the period spanned between September 2010-January 2022.

This is not the only ARX model we built. After comparing the ARX model performance to the benchmark model and determining the significance of the forecasts, various models with different combinations of variables were built and their forecasting performance was compared to determine how the model improves depending on the variables used to forecast the natural gas price.

The forecasting performance of each model is evaluated by the out-of-sample  $R^2$  statistic, suggested by Campbell and Thompson [4]. We then examine whether the results are statistically significant by estimating the MSPE-adjusted statistic developed by Clark and West (2007). Using these two metrics, the model with the best forecasting performance resulted to be the one that uses REDTI, natural gas levels of production and inventories and electricity consumption to forecast the price of natural gas, apart from the historical price of natural gas itself (see *Table 1* in *Chapter 4* for a summary of the results of each model).

As mentioned above, Machine Learning forecasting techniques were also studied in this paper and one of them, LSTM Neural Networks, is used to forecast natural gas prices for the same period of September 2010-January 2022. This way, the results of Machine Learning and the Autoregressive models were compared. In this case, like in the benchmark autoregressive model, natural gas price was predicted using only historical natural gas prices, no other variables were included

to make the training of the machine less complicated. TensorFlow and Keras were used as software frameworks and the parameters of the model are summarized in *Table 4* in *Chapter 6*.

In order to compare the performance of the Machine Learning model to the autoregressive models, the  $R^2$  (variance score) and Root Mean Squared Error (RMSE, test score) is calculated for the benchmark autoregressive model and the ARX Model 1. We used Model 1 because it is the one with the best performance metrics.

We obtained the following values for all three models:

	Machine Learning	BM AR Model	ARX Model 1
$R^2$	0.64 (64%)	0.32 (32%)	0.34 (34%)
RMSE	0.52€	0.51€	0.50€

*Table 7. Performance metrics of the three different models*

In terms of the variance explained by the model, the machine learning model is way superior to the autoregressive model, explaining approximately 30% more variance than the other two models. However, in terms of RMSE, all models perform a very similar error in average when predicting natural gas, around +/- 0.50€.

Taking these performance metrics, the ARX model that includes, apart from the historical natural gas price, the variations in residential demand due to weather conditions, in natural gas production, in natural gas inventories and in electricity consumption, performs slightly better than the benchmark model that includes only historical natural gas prices to predict future prices.

## Chapter 8

# Future Lines of Investigation

The forecasting industry is a very wide one, with many techniques available to predict the behavior of almost every variable we could think of. In this study, we decided to use two techniques that are commonly used to forecast shares and commodity prices, the Autoregressive Model (ARX) with certain variables and Long-Short-Term-Memory Neural Networks. However, there are other methodologies that can be used for the same purpose and that might be of interest to study in the future.

Concerning the variables considered to build the ARX Model, it could be of interest for future studies to include two more variables: the supply variations of nuclear power and the indicator “container trade volume”. As studied in the TFG: Modelling the Natural Gas Market [2], after the Fukushima earthquake and the consequent temporary reduction of nuclear power generation, natural gas prices rocketed, which suggests that the price of natural gas might be affected by the supply levels of nuclear power. Also, supply bottlenecks seem to have affected throughout history to the fluctuations in the price of natural gas. In order to include this supply problems in the model, a possible future modification of the model could be to include the indicator “container trade volume”, which counts the containers traded between certain regions at a certain point in time [2].

In Chapter 5, we studied different Machine Learning techniques, but we decided to predict natural gas prices for the next year using LSTM Neural Networks. However, in future studies it could be of interest to use another Machine Learning techniques, such as Convolutional Neural Networks (CNN) and compare the results to those obtained in this study. These neural networks consist of filters that are applied along the input data to de-noise it, so that it can be fed through a basic neural network. CNNs have always been considered as the best algorithm for image data processing, however, some studies suggest that CNNs could also be applied to an array of 1D values and give similar prediction results as LSTM Neural Networks using less training time and computational intensity [50]. Another option is to combine both techniques, LSTM and CNN, to obtain optimal results.

To make the model less complex, we only used historical natural gas prices to predict with Machine Learning. As we saw with the autoregressive model, the model that also included monthly variations of residential demand due to temperature changes, natural gas production, inventories, and electricity consumption, performed better than the benchmark model that only used historical natural gas price to predict future natural gas prices. Therefore, another future variation could be adding these variables to the Machine Learning model to see if it performs better than the one, we built and so it could be compared to the ARX model we built.

# Annexes

# Annex I

## Sustainable Development Goals

Throughout history, there have been many moments in which humans had to join forces to overcome the difficulties that laid ahead on a global scale. Examples include the 1948 Marshall Plan developed to rebuild the economies of Western Europe after World War II, or the achievement of the UN Millennium Development Goals, which succeeded, among other things, in halving the under-five mortality rate and the number of people living in extreme poverty.

Now, in the 21<sup>st</sup> century, the world is facing one of its greatest challenges. Everyone needs to work together to end global poverty, address social inequalities and tackle the serious problem of climate change by 2030. This requires investment in innovative solutions, bold finance to drive progress, and the commitment of entrepreneurs, the scientific community, activists, governments and, in general, the entire population.

The United Nations (UN), aware of the situation, developed the so-called Sustainable Development Goals (SDGs), a set of 17 global goals designed to achieve a better and more sustainable future for all [48].



Figure A1.1. 17 ODSs. Fuente: Sustainable Development Goals. United Nations [20].

With these 17 goals encompassed in the 2030 Agenda, the UN presents a plan to follow to address the problems facing the international community today, giving all humans the opportunity to participate. This plan aims to make the world a place where resources are cared for in order to achieve a fairer distribution. The main components of this agenda are the five Ps:

- *People*: ending poverty and world hunger in all their forms and ensuring dignity and equality among people.
- *Prosperity*: to ensure prosperous and fulfilling lives in harmony with nature.
- *Peace*: to implement justice, peace and inclusion in communities.
- *Partnership*: to implement the Agenda through a strong global partnership.
- *Planet*: protect the planet's resources and climate to ensure a prosperous future for generations to come.

Aware of the responsibility that this implies for the Engineering Community, which must, among other things, develop innovative processes and products to

achieve the goals set, this master’s thesis studies one of the cleanest burning fossil fuels available at the moment, natural gas. Burning natural gas for energy results in fewer emissions of nearly all types of air pollutants and carbon dioxide (CO<sub>2</sub>) than burning coal or petroleum products to produce an equal amount of energy. According to the EIA, about 117 pounds of CO<sub>2</sub> are produced per million British thermal units (MMBtu) equivalent of natural gas compared with more than 200 pounds of CO<sub>2</sub> per MMBtu of coal and more than 160 pounds per MMBtu of distillate fuel oil [2]. Although the main objective of this master’s thesis is to predict natural gas prices and understand how the price is affected by various variables, it also aims to educate readers about natural gas, its properties, its possibilities, and its environmental implications.

The Sustainable Development Goals included in this project “Forecasting Natural Gas Prices”, by educating the reader about a fossil fuel that is cleaner than other fossil fuels and might be key in the transition to producing energy only from renewable sources, are the following:



The SDGs addressed, and more information about their application in this project, are detailed below:

<i>SDG Dimension</i>	<i>SDG Specification</i>	<i>Role</i>	<i>Objective</i>
<b>Economic</b>	SDG 7: Ensure access to affordable, reliable, sustainable and modern energy for all	Secondary	Educate the reader about the possibilities of natural gas and how



			it can help many countries to obtain energy independence and reduce GHG emissions while technologies to rely only on renewable energy sources are developed
<b>Economic</b>	SDG 12: Ensure sustainable consumption and production patterns	Secondary	Educate the reader about the uses given to natural gas and how daily activities affect to the greater or lower consumption of this fuel
<b>Environmental</b>	SDG 13: Take urgent action to combat climate change and its impacts	Secondary	Popularize natural gas as a cleaner option to other fossil fuels, such as petroleum or carbon, with lower GHG emissions, that used responsibly can constitute a good option in the transition to a fully renewable energy portfolio

## Annex II

### MATLAB Code

```
%%
clear all
clc
close all
%% Load Data
infomat = load ('Matrix.mat');
Matrix = infomat.Matrix;
infomatx = load ('Matrixx.mat');
Matrixx = table2array (infomatx.Matrixx);

%% Set up forecasting exercise
infomat = load ('NGPrice.mat');
NGPrice=infomat.NGPrice1;
% Target Variable

start = 248;
% Forecasting starts here
sample = size(NGPrice,1)-start-1;
% Sample size to perform forecasting
BM_Forecast = zeros(sample,1);
% Empty Matrix to save benchmark model forecasts
BM_Parameters = zeros(sample,2);
% Empty matrix to save alpha and beta

%% Benchmark Model - AR(1)

for i = 0:sample-1
% For each observation starting Dec 2004
    Y = NGPrice{129:start+i,:};
% Take natural gas recorded prices until that moment
    X = [ones(size(Y,1),1) Y];
% Take lag of observed data
```

```

    b = regress(Y,lagmatrix(X,1));
% Calculate relation between natural gas observed
price and the one observed the period before
    BM_Parameters(i+1,:) = b';
% Save estimated alpha and beta
    BM_Forecast(i+1,1) = X(end,:)*b;
% Forecast(with the last observed data in Dec 2004)
the value for Jan 2005....and do the same for each
observation aggregating the new data point observed
each month
end

%% Estimate ARX model
Variables= Matrixx(:,[1:6 9]);
% Variables= Matrixx(:,[1 3 5 6 9]);
% Variables= Matrixx(:,[1]);
ARX_Forecast = zeros(sample,1);
% Empty Matrix to save ARX model forecasts
ARX_Parameters = zeros(sample,size(Variables,2)+2);
% Empty matrix to save alpha and beta from ARX model

for i = 0:sample-1
    Y = NGPrice{1:start+i,:};
    X = [ones(size(Y,1),1) Y Variables(1:start+i,:)];
    b = regress(Y,lagmatrix(X,1));
    ARX_Parameters(i+1,:) = b';
    ARX_Forecast(i+1,1) = X(end,:)*b;
end

%% Estimate R2OS

ERRORS = cell(sample,1);
ERRORS2 = cell(sample,1);
SE = cell(sample+1,1);
SE2 = cell(sample+1,1);
R2OS = zeros(sample,1);

for i = 1:sample
    ERRORS{i,1} = NGPrice{start+i,1}-ARX_Forecast(i,1);
    SE{i,1} = ERRORS{i,1}.^2;
end

for i = 1:sample
    ERRORS2{i,1} = NGPrice{start+i,1}-BM_Forecast(i,1);
    SE2{i,1} = ERRORS2{i,1}.^2;
end

SE = cell2mat(SE);
SE2 = cell2mat(SE2);

```

```
temp1 = nansum(SE);
temp2 = nansum(SE2);

R2OS = 1-temp1./temp2;

%% Estimate Clark and West (1996) statistic

F = cell(sample,1);
CandW = zeros(sample,1);

F = SE2 - (SE - (ARX_Forecast - BM_Forecast).^2);

Y2 = F;
X2 = ones(size(Y2,1),1);
mdl = LinearModel.fit(X2,Y2,'Intercept',false);
tstat = table2array(mdl.Coefficients(1,3));
CandW = (1-tcdf(tstat,sample-5));

%
% save('FC_PMI_AR.mat','CandW','FORECAST','R2OS');
```

## **Annex III**

### **Variables Matrix**

ANNEX III. VARIABLES MATRIX

Date	Weather Conditions	World Ind			Electricity Consumption	Inventories	Natural Gas	Monetary	
		REDTI	Drilling Activity	Production			Prod	Price	Policy
01/1990		8.8	(7.74%)	3.63%	(0.40%)	7.12%	(3.47%)	202.52%	7.95
02/1990		22.8	(13.63%)	0.13%	0.41%	(0.24%)	0.40%	150.18%	7.80
03/1990		29.8	2.18%	1.51%	0.42%	(0.61%)	2.08%	134.29%	7.95
04/1990		27	6.18%	0.58%	(1.02%)	2.26%	2.86%	120.43%	8.14
05/1990		53.9	4.03%	0.30%	1.12%	3.71%	2.37%	101.11%	7.77
06/1990		54.1	6.45%	1.40%	0.36%	2.71%	2.48%	95.80%	7.67
07/1990		34.6	0.20%	1.86%	(0.01%)	5.13%	3.03%	94.41%	7.53
08/1990		39.3	(6.05%)	2.60%	0.55%	4.04%	2.70%	93.01%	7.55
09/1990		33	(1.29%)	2.52%	0.20%	6.17%	2.63%	110.11%	7.51
10/1990		26.9	1.96%	6.82%	(0.07%)	5.53%	3.67%	135.34%	7.34
11/1990		22.1	5.54%	2.01%	(0.48%)	2.25%	4.74%	146.68%	7.19
12/1990		43.7	(7.07%)	2.13%	(0.21%)	(2.95%)	9.66%	110.12%	6.76
01/1991		46.1	(10.22%)	0.93%	0.20%	0.46%	3.04%	(33.86%)	6.39
02/1991		16	(1.94%)	1.06%	(0.70%)	2.40%	2.64%	(32.03%)	6.10
03/1991		38.3	(3.95%)	2.57%	(0.40%)	(0.28%)	1.99%	(17.52%)	5.76
04/1991		20.4	(3.86%)	2.68%	0.73%	0.04%	2.59%	(14.83%)	5.56
05/1991		30.2	(5.35%)	0.30%	0.45%	4.07%	3.35%	(15.16%)	5.54
06/1991		53.5	(3.39%)	0.03%	0.16%	4.64%	2.58%	(22.03%)	5.53
07/1991		47.7	(2.92%)	(0.98%)	0.13%	1.95%	1.09%	(19.13%)	5.45
08/1991		46.4	(1.81%)	(1.38%)	(0.04%)	2.34%	0.07%	(7.20%)	5.21
09/1991		23.2	(3.68%)	1.48%	0.30%	(0.12%)	(0.31%)	9.48%	5.02
10/1991		26.9	5.10%	1.19%	0.28%	(0.20%)	0.15%	1.83%	4.80
11/1991		66.4	(0.61%)	2.09%	0.03%	3.35%	(3.28%)	(16.37%)	4.41
12/1991		33.1	(6.10%)	2.48%	(0.80%)	2.32%	(2.28%)	(11.64%)	3.75
01/1992		31.1	(4.55%)	(0.25%)	0.43%	(0.63%)	0.06%	(25.35%)	3.60
02/1992		17.5	(5.78%)	0.57%	0.69%	3.66%	(1.29%)	(13.47%)	3.53
03/1992		44	(9.75%)	(2.75%)	(0.09%)	3.10%	(3.71%)	(7.43%)	3.53

04/1992	36.3	0.40%	0.09%	0.21%	3.13%	(5.45%)	7.12%	3.27
05/1992	7.9	3.59%	3.09%	(0.30%)	(1.58%)	(5.02%)	17.81%	3.29
06/1992	9.2	0.00%	5.08%	0.14%	(3.91%)	(4.46%)	26.20%	3.42
07/1992	22.2	19.23%	5.82%	0.47%	(0.93%)	(2.82%)	42.56%	3.10
08/1992	2.5	6.77%	3.35%	(0.65%)	(1.75%)	(1.50%)	45.90%	3.08
09/1992	33.6	7.55%	4.11%	0.54%	(1.64%)	(0.64%)	38.79%	2.85
10/1992	48.3	12.08%	1.90%	0.17%	1.64%	(0.57%)	32.38%	2.95
11/1992	60.8	13.03%	0.38%	(0.06%)	1.26%	0.21%	20.01%	2.98
12/1992	49.6	12.86%	0.69%	(0.57%)	3.63%	(2.02%)	9.26%	2.76
01/1993	34.3	(10.81%)	0.67%	0.77%	1.74%	(3.06%)	42.25%	2.82
02/1993	61.6	(26.43%)	1.13%	0.71%	1.19%	(6.13%)	35.28%	2.87
03/1993	56.7	(19.76%)	5.75%	(0.09%)	5.95%	(6.41%)	65.32%	2.93
04/1993	46.8	0.75%	2.31%	0.28%	1.89%	(4.84%)	54.97%	2.91
05/1993	0.5	8.89%	2.02%	0.29%	0.38%	(2.17%)	32.22%	3.10
06/1993	36.9	11.22%	(0.26%)	(0.35%)	4.13%	(0.83%)	22.61%	3.14
07/1993	60.3	10.09%	0.48%	0.01%	6.47%	(0.42%)	14.46%	3.08
08/1993	59.6	8.33%	4.03%	(0.12%)	8.36%	(0.08%)	11.55%	3.09
09/1993	47.8	7.95%	2.58%	0.37%	5.28%	0.49%	(5.20%)	3.04
10/1993	49.2	(2.38%)	2.71%	0.00%	1.75%	0.06%	(18.76%)	3.11
11/1993	66.1	(0.73%)	5.06%	0.30%	2.16%	(0.37%)	(1.80%)	3.13
12/1993	48.3	4.41%	5.48%	0.58%	2.42%	0.13%	5.37%	3.13
01/1994	65	(0.23%)	2.92%	(0.04%)	7.20%	(2.60%)	21.49%	3.05
02/1994	58.5	(4.71%)	2.70%	0.33%	4.59%	(1.91%)	56.42%	3.43
03/1994	41.8	(0.49%)	4.40%	0.78%	(0.28%)	1.30%	(1.24%)	3.69
04/1994	18.3	(1.24%)	4.41%	0.65%	2.23%	3.32%	(15.20%)	4.07
05/1994	31.1	(4.02%)	5.60%	0.52%	3.75%	2.45%	(13.50%)	4.43
06/1994	72.3	6.81%	4.81%	0.75%	6.72%	1.57%	(5.90%)	4.68
07/1994	42	1.72%	5.01%	0.23%	3.27%	2.03%	(7.35%)	4.70
08/1994	23.6	4.34%	5.31%	1.04%	(0.46%)	2.11%	(28.62%)	4.74

ANNEX III. VARIABLES MATRIX

09/1994	2.2	8.78%	2.48%	0.36%	0.85%	1.77%	(36.25%)	5.08
10/1994	35.9	(0.85%)	1.62%	0.74%	2.38%	1.87%	(27.15%)	5.27
11/1994	34.4	(1.50%)	3.56%	0.87%	1.55%	3.43%	(31.86%)	5.97
12/1994	27.8	(2.83%)	2.55%	1.13%	0.73%	4.77%	(28.36%)	7.12
01/1995	30.3	(8.05%)	1.01%	(0.35%)	(2.27%)	8.16%	(37.27%)	6.32
02/1995	41.5	(8.76%)	0.36%	0.20%	(0.16%)	8.88%	(43.32%)	5.99
03/1995	34.3	(11.73%)	(0.14%)	(0.06%)	1.86%	7.41%	(32.21%)	6.00
04/1995	54.9	1.51%	1.06%	0.45%	1.04%	4.04%	(22.99%)	5.93
05/1995	52.5	(0.30%)	0.64%	(0.39%)	3.72%	2.63%	(17.17%)	5.79
06/1995	24	5.07%	1.55%	0.50%	0.82%	2.75%	(17.25%)	5.68
07/1995	56.1	13.35%	0.89%	(0.11%)	1.57%	0.19%	(28.55%)	5.54
08/1995	95.6	0.00%	(0.31%)	0.63%	9.90%	(1.83%)	(8.42%)	5.49
09/1995	36.4	3.51%	1.95%	0.01%	4.48%	(1.69%)	7.34%	5.54
10/1995	23.4	0.24%	0.33%	(0.00%)	3.15%	(1.28%)	14.09%	5.49
11/1995	70.6	3.86%	0.48%	0.52%	3.84%	(3.56%)	25.84%	5.42
12/1995	57.5	(0.70%)	0.58%	0.78%	3.99%	(6.66%)	53.67%	5.33
01/1996	49	(4.92%)	0.29%	(0.19%)	6.70%	(9.26%)	88.77%	5.03
02/1996	40.5	1.23%	6.36%	(0.03%)	7.02%	(9.14%)	171.11%	5.05
03/1996	67	2.43%	1.27%	0.50%	5.21%	(11.33%)	86.26%	5.15
04/1996	43.8	5.94%	2.89%	0.40%	3.12%	(9.99%)	33.86%	5.19
05/1996	64.9	4.71%	1.29%	0.66%	4.27%	(9.38%)	32.83%	5.15
06/1996	44.5	0.86%	2.54%	0.15%	4.51%	(8.56%)	49.49%	5.26
07/1996	30.4	3.61%	2.39%	0.31%	1.37%	(6.14%)	67.40%	5.39
08/1996	37.3	0.00%	3.09%	0.42%	(3.11%)	(3.77%)	26.57%	5.28
09/1996	12.6	3.48%	2.73%	0.50%	0.72%	(2.81%)	8.92%	5.29
10/1996	36.4	(1.19%)	2.10%	0.29%	1.81%	(2.57%)	29.92%	5.21
11/1996	82.3	(3.41%)	(1.78%)	0.45%	2.39%	(2.57%)	43.85%	5.20
12/1996	34.6	1.45%	(3.44%)	0.54%	1.85%	0.17%	36.35%	5.22
01/1997	46.3	(2.25%)	1.52%	0.32%	1.17%	0.45%	9.64%	5.23



02/1997	25.7	2.93%	(2.10%)	0.56%	(1.99%)	2.08%	(51.03%)	5.27
03/1997	34.6	5.28%	2.09%	0.81%	(2.02%)	5.69%	(37.66%)	5.49
04/1997	64.7	1.54%	(0.81%)	0.54%	2.38%	4.39%	(11.09%)	5.46
05/1997	55	2.85%	0.49%	0.22%	(0.90%)	3.85%	(2.19%)	5.35
06/1997	12.9	6.65%	(0.79%)	0.90%	(1.72%)	3.71%	(13.58%)	5.32
07/1997	36.1	1.21%	0.20%	0.36%	3.53%	2.23%	(13.57%)	5.35
08/1997	28.6	(0.51%)	(0.60%)	0.44%	1.42%	1.80%	18.95%	5.40
09/1997	44.5	5.68%	0.63%	0.50%	3.99%	1.30%	51.00%	5.53
10/1997	35.9	(1.95%)	1.31%	0.68%	7.02%	1.38%	25.65%	5.38
11/1997	68.4	3.82%	(0.48%)	(0.15%)	2.21%	2.46%	(2.18%)	5.46
12/1997	44.2	3.68%	(0.53%)	(0.06%)	2.51%	0.17%	(40.02%)	5.42
01/1998	14.4	(6.02%)	1.12%	(0.21%)	(1.32%)	3.70%	(37.57%)	5.36
02/1998	14	(3.28%)	(0.35%)	0.79%	(0.55%)	5.24%	(1.42%)	5.41
03/1998	42	2.04%	0.17%	(0.31%)	4.11%	3.54%	16.39%	5.37
04/1998	31.1	(1.66%)	1.24%	(0.07%)	2.01%	6.15%	17.46%	5.33
05/1998	35.5	(1.86%)	1.04%	0.24%	7.08%	7.18%	(6.05%)	5.41
06/1998	65.4	0.86%	2.31%	(0.20%)	8.32%	5.93%	(2.94%)	5.29
07/1998	61.1	(6.15%)	0.33%	(0.10%)	5.81%	6.81%	(2.61%)	5.23
08/1998	77.9	2.91%	1.19%	0.34%	8.20%	5.12%	(26.29%)	5.13
09/1998	60	(1.06%)	(4.81%)	0.24%	6.57%	3.36%	(30.22%)	4.85
10/1998	29.4	(7.16%)	(1.32%)	0.30%	1.14%	3.99%	(38.74%)	4.73
11/1998	31.2	(3.85%)	(1.33%)	0.16%	1.18%	6.24%	(31.38%)	4.74
12/1998	26.7	(1.60%)	(0.95%)	(0.12%)	1.37%	8.14%	(26.50%)	4.69
01/1999	31.6	(6.11%)	(1.79%)	0.95%	4.76%	5.69%	(12.88%)	4.76
02/1999	18.2	(7.81%)	(0.90%)	(0.58%)	1.16%	5.32%	(21.58%)	4.85
03/1999	52.9	(3.06%)	(1.59%)	0.81%	2.77%	4.79%	(21.09%)	4.73
04/1999	35.3	(9.95%)	(2.47%)	0.06%	3.46%	2.64%	(13.14%)	4.74
05/1999	19.9	2.43%	(3.80%)	0.93%	0.77%	1.50%	2.99%	4.74
06/1999	40.7	14.21%	(2.56%)	0.06%	0.68%	1.09%	3.95%	4.93

ANNEX III. VARIABLES MATRIX

07/1999	72.6	10.14%	(1.73%)	0.69%	3.65%	(0.84%)	4.22%	4.93
08/1999	57.3	10.25%	(2.51%)	0.43%	1.58%	(0.85%)	47.47%	5.19
09/1999	17.5	7.21%	2.81%	0.65%	(0.64%)	0.33%	23.13%	5.17
10/1999	38.2	6.37%	(0.61%)	0.58%	(0.23%)	(1.20%)	40.32%	5.31
11/1999	2.9	5.66%	0.52%	0.94%	1.46%	(0.72%)	9.36%	5.59
12/1999	33.9	0.16%	0.82%	0.93%	1.20%	(2.12%)	32.10%	5.83
01/2000	36.8	(0.63%)	(0.50%)	(0.24%)	0.94%	(4.14%)	26.57%	5.85
02/2000	16.6	(2.53%)	1.86%	0.36%	8.09%	(6.48%)	45.05%	5.85
03/2000	21.4	(2.60%)	0.54%	0.50%	0.80%	(4.70%)	50.21%	6.06
04/2000	34	1.50%	(0.08%)	0.99%	0.78%	(5.30%)	36.81%	6.16
05/2000	31.1	5.91%	1.32%	0.50%	5.40%	(6.62%)	54.28%	6.62
06/2000	43.4	4.96%	1.37%	0.10%	5.02%	(6.76%)	79.39%	6.46
07/2000	21.1	8.27%	1.77%	0.36%	(2.31%)	(5.70%)	65.47%	6.41
08/2000	55.2	6.28%	2.61%	0.49%	2.25%	(5.99%)	52.94%	6.39
09/2000	35.7	3.98%	2.59%	(0.22%)	2.82%	(6.00%)	92.56%	6.52
10/2000	22.9	3.95%	3.99%	0.33%	2.88%	(4.72%)	78.40%	6.65
11/2000	75	(1.19%)	0.85%	0.36%	4.45%	(8.62%)	127.34%	6.44
12/2000	84.5	2.64%	1.46%	0.36%	7.87%	(12.09%)	266.64%	5.99
01/2001	46.2	2.93%	2.02%	(1.34%)	7.94%	(8.63%)	225.49%	5.04
02/2001	37.3	2.16%	0.08%	0.83%	(0.25%)	(7.75%)	105.21%	4.76
03/2001	56.8	1.67%	1.97%	(0.93%)	2.36%	(8.61%)	79.59%	4.30
04/2001	22.8	4.82%	3.58%	(0.44%)	1.61%	(5.60%)	65.31%	3.83
05/2001	11.2	4.18%	3.47%	(0.02%)	(2.26%)	(0.79%)	13.55%	3.23
06/2001	43.2	5.32%	1.49%	(0.24%)	(3.47%)	1.87%	(16.00%)	3.27
07/2001	30	0.76%	1.19%	(0.70%)	(0.72%)	3.32%	(23.79%)	3.05
08/2001	73.7	(2.46%)	0.92%	0.55%	0.13%	5.01%	(34.88%)	2.83
09/2001	9.8	(5.81%)	1.12%	(0.68%)	(2.59%)	5.94%	(58.58%)	2.17
10/2001	30.7	(6.07%)	(0.03%)	(0.33%)	(1.97%)	5.10%	(52.21%)	1.57
11/2001	0	(9.64%)	0.32%	(0.27%)	(4.23%)	11.05%	(58.27%)	1.13

12/2001	22.3	(8.61%)	0.18%	0.08%	(7.92%)	18.67%	(73.50%)	0.67
01/2002	18.4	(3.85%)	(1.87%)	1.06%	(6.11%)	18.68%	(72.79%)	0.95
02/2002	24.7	(6.34%)	(2.69%)	(0.27%)	(2.88%)	18.18%	(59.43%)	0.97
03/2002	50.4	(9.13%)	(2.58%)	0.97%	(0.58%)	16.48%	(42.07%)	1.40
04/2002	21.3	(0.81%)	(2.46%)	0.70%	1.55%	14.48%	(34.92%)	0.93
05/2002	53.5	12.75%	(2.07%)	0.38%	2.07%	10.09%	(18.12%)	1.12
06/2002	53.4	2.03%	(1.20%)	0.20%	2.46%	7.60%	(14.13%)	1.02
07/2002	68.7	1.70%	(0.25%)	0.40%	6.74%	4.87%	(5.26%)	1.15
08/2002	68.1	0.70%	(1.86%)	0.34%	1.64%	3.50%	2.95%	1.43
09/2002	46.9	2.08%	(4.30%)	0.30%	4.76%	1.68%	63.57%	1.59
10/2002	54.5	(3.67%)	(5.75%)	0.11%	5.75%	0.06%	64.82%	1.44
11/2002	56.4	(3.67%)	(1.33%)	0.35%	3.84%	(3.73%)	67.41%	1.14
12/2002	46.4	4.54%	(1.01%)	(0.34%)	6.30%	(6.78%)	91.92%	1.12
01/2003	50.2	0.56%	(0.05%)	1.12%	4.62%	(11.88%)	133.81%	1.02
02/2003	58	4.46%	1.18%	(0.02%)	6.77%	(16.25%)	226.26%	1.16
03/2003	40.5	2.27%	2.05%	0.05%	2.50%	(13.91%)	90.50%	1.06
04/2003	43	3.65%	1.48%	(0.29%)	(0.45%)	(13.36%)	51.37%	1.01
05/2003	53.2	8.68%	(0.55%)	(0.16%)	(0.14%)	(11.18%)	64.17%	1.28
06/2003	18.7	5.32%	0.04%	0.46%	(1.97%)	(8.60%)	75.83%	1.10
07/2003	47.7	1.54%	(2.22%)	0.74%	(0.96%)	(6.46%)	64.57%	0.94
08/2003	76.4	0.87%	0.75%	(0.13%)	1.10%	(5.15%)	57.38%	0.74
09/2003	19	0.43%	2.80%	0.95%	(1.18%)	(2.85%)	26.21%	0.86
10/2003	36.9	0.53%	4.08%	0.91%	(2.11%)	(0.02%)	10.26%	0.83
11/2003	30.9	1.17%	(0.07%)	0.33%	(0.00%)	0.95%	9.95%	0.87
12/2003	40.7	0.74%	(0.37%)	0.74%	2.62%	2.24%	26.67%	0.74
01/2004	55	(0.42%)	0.29%	(0.16%)	1.03%	3.17%	10.50%	0.75
02/2004	48.6	0.63%	2.13%	1.30%	1.58%	5.11%	(31.95%)	0.75
03/2004	21.9	0.73%	(1.31%)	(0.02%)	1.31%	5.66%	(10.80%)	0.91
04/2004	17.3	2.89%	(0.61%)	0.65%	2.36%	6.23%	5.65%	0.87

ANNEX III. VARIABLES MATRIX

05/2004	21.2	1.10%	(0.77%)	0.33%	3.69%	5.15%	5.47%	1.07
06/2004	34.6	0.40%	(0.22%)	(0.08%)	5.75%	3.56%	5.32%	1.32
07/2004	23	2.97%	1.38%	0.69%	(0.05%)	3.57%	15.07%	1.44
08/2004	14.4	2.11%	(0.84%)	0.05%	(3.19%)	3.58%	5.85%	1.63
09/2004	25	0.94%	(3.40%)	0.40%	0.59%	1.91%	8.65%	1.93
10/2004	36.2	(0.47%)	(2.33%)	0.45%	1.32%	1.23%	34.04%	2.08
11/2004	32.8	0.84%	(0.44%)	0.15%	2.57%	1.88%	29.75%	2.35
12/2004	43.4	(1.21%)	(1.08%)	0.35%	2.14%	0.46%	3.95%	2.53
01/2005	33.2	1.03%	(1.35%)	0.78%	(0.20%)	2.44%	(1.81%)	2.68
02/2005	26	0.74%	(2.62%)	0.06%	(2.91%)	5.81%	10.84%	2.84
03/2005	54.4	3.23%	0.19%	0.24%	2.52%	2.67%	25.10%	2.96
04/2005	17.7	4.03%	(0.79%)	0.73%	(0.09%)	2.96%	21.05%	3.08
05/2005	43.2	0.60%	0.46%	(0.38%)	(2.47%)	2.78%	(0.74%)	3.15
06/2005	63.8	3.25%	0.37%	0.67%	3.32%	1.45%	11.74%	3.32
07/2005	70.5	1.49%	(1.73%)	0.09%	6.02%	(0.42%)	24.81%	3.58
08/2005	84.9	0.08%	(1.41%)	0.32%	9.13%	(1.99%)	72.19%	3.63
09/2005	65.9	0.73%	(7.47%)	0.08%	6.46%	(2.38%)	139.21%	4.14
10/2005	22	1.62%	(6.87%)	0.58%	4.53%	(1.96%)	102.11%	4.20
11/2005	24.3	(2.23%)	(3.43%)	1.07%	0.78%	(1.08%)	65.83%	4.33
12/2005	57.2	(0.65%)	(1.37%)	0.66%	1.57%	(0.90%)	88.61%	4.36
01/2006	0	0.66%	(2.94%)	(0.13%)	(1.42%)	6.03%	35.29%	4.53
02/2006	38.3	7.57%	(4.08%)	0.87%	0.38%	5.54%	17.58%	4.60
03/2006	45.9	(1.21%)	(4.14%)	0.31%	0.84%	7.37%	(4.26%)	4.77
04/2006	13.3	2.45%	(3.56%)	0.41%	1.25%	7.80%	(4.28%)	4.91
05/2006	21.4	2.69%	(2.69%)	0.68%	4.80%	7.18%	(7.86%)	5.07
06/2006	48.3	0.22%	(1.73%)	0.33%	0.43%	6.77%	(17.28%)	5.26
07/2006	81.7	0.22%	(0.45%)	0.33%	1.57%	5.11%	(21.28%)	5.23
08/2006	74.9	2.76%	0.07%	0.39%	1.33%	4.62%	(30.11%)	5.20
09/2006	20.8	0.85%	10.16%	0.38%	(4.03%)	5.63%	(62.98%)	5.15

10/2006	51.5	0.84%	8.26%	0.26%	(1.58%)	3.63%	(56.52%)	5.15
11/2006	23.1	(1.87%)	3.82%	1.19%	1.13%	3.03%	(29.92%)	5.13
12/2006	21	1.20%	3.01%	0.64%	(2.22%)	6.52%	(49.98%)	5.17
01/2007	31.1	0.63%	3.43%	0.58%	2.81%	0.40%	(25.52%)	5.14
02/2007	67.3	1.81%	4.67%	0.21%	6.33%	(3.64%)	3.96%	5.13
03/2007	23.8	(0.34%)	5.02%	0.65%	(0.04%)	(0.66%)	0.44%	5.18
04/2007	48.6	0.00%	4.50%	0.14%	2.03%	(2.84%)	4.25%	5.22
05/2007	4.8	0.21%	5.63%	0.92%	1.20%	(1.23%)	19.60%	5.10
06/2007	47.5	1.30%	4.18%	0.15%	(0.19%)	(0.27%)	14.85%	5.01
07/2007	28.8	0.20%	3.95%	0.45%	(3.19%)	1.86%	(2.72%)	5.02
08/2007	100	0.40%	4.36%	0.43%	0.79%	0.95%	(13.08%)	4.64
09/2007	42.1	(1.14%)	4.84%	0.06%	5.67%	0.14%	21.93%	4.29
10/2007	4.7	(2.71%)	4.76%	0.66%	4.84%	1.75%	10.04%	4.29
11/2007	42.9	1.11%	6.29%	0.42%	2.57%	0.76%	(8.25%)	3.69
12/2007	44.8	1.17%	5.85%	0.21%	1.80%	(2.30%)	4.40%	3.70
01/2008	40	(3.20%)	5.49%	1.49%	3.63%	(4.72%)	16.50%	2.66
02/2008	41.7	0.35%	8.83%	(0.54%)	1.63%	(3.09%)	2.98%	2.14
03/2008	48	1.26%	5.23%	(0.31%)	1.33%	(6.21%)	26.94%	1.87
04/2008	27.8	1.18%	5.90%	0.43%	1.19%	(5.19%)	28.52%	1.81
05/2008	50.5	1.16%	4.19%	(0.88%)	(1.26%)	(5.70%)	41.71%	1.96
06/2008	78.6	2.17%	4.90%	0.26%	2.00%	(5.97%)	65.38%	2.05
07/2008	46.3	2.19%	7.40%	(0.67%)	2.51%	(5.30%)	69.88%	2.05
08/2008	35.6	2.46%	5.65%	(0.90%)	(5.42%)	(2.16%)	26.45%	2.13
09/2008	26.3	0.25%	(5.92%)	(1.01%)	(4.62%)	(2.09%)	20.26%	1.91
10/2008	39.3	(2.71%)	1.47%	(1.54%)	(4.99%)	(2.17%)	(4.51%)	1.84
11/2008	46	(2.85%)	2.44%	(2.70%)	(2.77%)	(1.34%)	(7.51%)	1.42
12/2008	50.3	(7.88%)	2.34%	(3.21%)	0.23%	(0.57%)	(18.93%)	0.65
01/2009	53.6	(11.96%)	4.50%	(2.83%)	(1.84%)	1.30%	(34.41%)	0.61
02/2009	32.1	(14.65%)	1.49%	0.85%	(6.32%)	5.53%	(47.15%)	0.88

ANNEX III. VARIABLES MATRIX

03/2009	40.2	(16.39%)	2.65%	(0.49%)	(3.83%)	7.67%	(57.79%)	0.75
04/2009	31.5	(10.61%)	1.67%	0.19%	(4.71%)	8.95%	(65.26%)	0.43
05/2009	23.5	(6.71%)	1.72%	0.57%	(4.74%)	9.35%	(65.73%)	0.21
06/2009	53.3	(4.43%)	1.03%	0.98%	(7.04%)	9.69%	(69.62%)	0.02
07/2009	19.7	(2.32%)	(1.60%)	0.88%	(6.29%)	9.02%	(68.98%)	-0.12
08/2009	53.6	2.37%	(0.04%)	0.82%	(2.03%)	7.51%	(61.27%)	-0.28
09/2009	32.6	1.88%	8.84%	1.13%	(3.12%)	7.18%	(60.97%)	-0.41
10/2009	63.5	2.56%	2.03%	0.49%	(1.52%)	5.98%	(40.12%)	-0.47
11/2009	13.9	1.66%	(0.26%)	0.93%	(3.31%)	7.23%	(45.74%)	-0.61
12/2009	60.4	3.27%	(1.38%)	0.59%	0.79%	4.72%	(9.95%)	-0.15
01/2010	48.1	8.44%	(1.86%)	1.86%	3.30%	3.29%	7.96%	-0.45
02/2010	59.5	8.52%	(1.27%)	(0.13%)	4.30%	(0.67%)	15.59%	-0.54
03/2010	31.2	4.60%	0.62%	1.46%	3.17%	0.38%	6.23%	-0.48
04/2010	2.4	2.79%	1.97%	0.45%	0.78%	2.02%	11.99%	-0.47
05/2010	38.1	0.10%	2.10%	0.85%	3.26%	1.01%	6.94%	-0.48
06/2010	100	(0.73%)	(0.08%)	(0.19%)	8.19%	(0.02%)	24.64%	-0.54
07/2010	77.9	1.89%	1.78%	0.47%	8.98%	(1.37%)	34.66%	-0.59
08/2010	90.2	1.24%	1.94%	0.38%	7.68%	(2.41%)	35.20%	-0.70
09/2010	55.8	(0.61%)	7.13%	0.23%	5.39%	(1.52%)	30.36%	-0.80
10/2010	19.5	(1.13%)	7.27%	0.48%	0.18%	0.80%	(15.69%)	-1.00
11/2010	43.5	(1.66%)	6.18%	0.58%	2.43%	(0.58%)	(0.09%)	-0.96
12/2010	65.9	(1.05%)	9.29%	1.17%	2.88%	0.08%	(22.15%)	-0.88
01/2011	54.2	(3.30%)	4.03%	1.39%	0.70%	0.44%	(23.91%)	-1.01
02/2011	44.6	(0.22%)	2.76%	(0.17%)	(0.54%)	1.06%	(25.22%)	-1.09
03/2011	42.5	(2.54%)	5.90%	(0.29%)	(0.31%)	(0.86%)	(9.84%)	-0.99
04/2011	33.4	0.11%	7.28%	(0.50%)	3.05%	(3.14%)	2.74%	-1.07
05/2011	32.5	(0.79%)	7.80%	0.92%	1.47%	(3.10%)	0.21%	-1.14
06/2011	84.4	(0.11%)	8.14%	0.26%	(0.60%)	(2.78%)	(8.34%)	-1.12
07/2011	100	0.34%	5.96%	0.30%	0.56%	(2.45%)	(7.91%)	-1.19

08/2011	88.4	1.59%	6.06%	0.22%	0.25%	(1.66%)	(9.24%)	-1.38
09/2011	34.7	1.45%	5.36%	(0.06%)	(0.45%)	(1.08%)	(3.74%)	-1.40
10/2011	26	2.87%	6.17%	0.01%	0.06%	(0.61%)	0.44%	-1.44
11/2011	24.2	(5.68%)	8.75%	(0.34%)	0.25%	0.80%	(15.91%)	-1.48
12/2011	28.4	(6.70%)	5.97%	0.23%	(5.18%)	4.74%	(27.54%)	-1.47
01/2012	15.7	(3.78%)	11.83%	0.78%	(6.58%)	9.24%	(42.11%)	-1.54
02/2012	17.5	(8.48%)	12.15%	0.21%	(2.91%)	12.19%	(39.92%)	-1.45
03/2012	0	(7.75%)	4.70%	(0.50%)	(2.55%)	15.57%	(46.83%)	-1.27
04/2012	19.5	(5.70%)	2.88%	(0.70%)	(1.50%)	13.86%	(55.13%)	-1.26
05/2012	21.8	(4.61%)	3.31%	0.60%	3.21%	11.22%	(44.40%)	-1.24
06/2012	51.4	(7.00%)	2.76%	(0.40%)	(0.97%)	9.11%	(46.91%)	-1.11
07/2012	96.3	(6.45%)	5.33%	0.11%	0.17%	7.25%	(34.19%)	-1.18
08/2012	69.1	(6.70%)	0.15%	(0.30%)	(2.03%)	5.94%	(31.15%)	-1.26
09/2012	36.8	(8.21%)	2.21%	(0.35%)	(2.37%)	4.25%	(28.56%)	-1.36
10/2012	31.8	(4.92%)	2.19%	0.51%	1.09%	2.33%	(8.99%)	-1.34
11/2012	49.4	(0.94%)	(0.19%)	(0.14%)	1.19%	0.36%	7.19%	-1.42
12/2012	22.8	0.48%	(0.79%)	0.32%	(1.41%)	0.28%	3.71%	-1.43
01/2013	30.5	2.60%	(2.31%)	0.40%	3.20%	(1.97%)	22.14%	-1.36
02/2013	43.4	(1.84%)	(3.82%)	(0.02%)	1.53%	(4.07%)	29.54%	-1.42
03/2013	58.8	(3.05%)	(0.80%)	0.27%	4.91%	(10.19%)	73.12%	-1.44
04/2013	41	(9.44%)	1.19%	0.38%	2.69%	(10.09%)	111.64%	-1.52
05/2013	3.9	(5.61%)	(0.08%)	0.07%	(2.35%)	(7.82%)	63.49%	-1.27
06/2013	57.2	(0.28%)	0.36%	(0.43%)	(1.31%)	(5.72%)	53.18%	-0.97
07/2013	49.7	3.41%	1.92%	0.57%	(3.07%)	(3.77%)	20.71%	-1.52
08/2013	37	6.04%	5.10%	0.47%	(2.90%)	(2.32%)	18.85%	-1.67
09/2013	38.9	0.78%	0.17%	0.31%	1.76%	(1.46%)	25.86%	-1.80
10/2013	30.1	(3.86%)	(1.72%)	(0.15%)	1.36%	(1.36%)	9.78%	-1.85
11/2013	63.5	(2.14%)	(0.66%)	0.48%	1.28%	(2.46%)	1.00%	-2.00
12/2013	52.9	1.91%	0.09%	(0.13%)	6.43%	(6.81%)	24.97%	-2.13

ANNEX III. VARIABLES MATRIX

01/2014	59	(2.95%)	2.71%	0.87%	6.10%	(11.14%)	39.08%	-2.38
02/2014	60.7	(5.80%)	3.84%	0.57%	5.64%	(14.24%)	77.43%	-2.54
03/2014	65.9	(2.35%)	4.79%	(0.15%)	1.48%	(14.67%)	26.11%	-2.62
04/2014	31.1	(5.11%)	5.62%	0.39%	(0.76%)	(13.04%)	9.01%	-2.89
05/2014	8.6	2.85%	5.81%	(0.23%)	0.54%	(11.33%)	10.53%	-2.99
06/2014	56	(3.38%)	7.32%	0.19%	0.60%	(9.44%)	17.03%	-2.89
07/2014	24.4	0.00%	4.65%	0.18%	(2.11%)	(7.41%)	8.55%	-2.84
08/2014	40.5	3.18%	6.01%	(0.97%)	(0.53%)	(5.80%)	11.49%	-2.89
09/2014	42.5	3.70%	8.74%	1.00%	0.98%	(4.69%)	6.55%	-2.81
10/2014	20.6	(2.38%)	8.86%	(0.67%)	0.62%	(2.77%)	1.04%	-2.80
11/2014	78	7.01%	8.42%	0.73%	1.19%	(2.22%)	11.97%	-2.77
12/2014	24.9	(2.56%)	9.66%	0.23%	(2.22%)	3.46%	(19.50%)	-2.42
01/2015	43	(6.43%)	8.14%	0.56%	(3.23%)	7.75%	(36.58%)	-2.27
02/2015	77.6	(7.50%)	6.94%	(0.27%)	(0.67%)	8.52%	(52.28%)	-1.97
03/2015	45.7	(15.54%)	7.76%	(0.04%)	0.84%	12.17%	(42.58%)	-1.81
04/2015	26.5	(11.20%)	6.77%	(0.06%)	(0.19%)	13.63%	(44.27%)	-1.59
05/2015	12.3	0.45%	5.98%	(0.25%)	(1.11%)	12.84%	(37.81%)	-1.43
06/2015	64	0.45%	3.91%	0.47%	1.23%	10.34%	(39.50%)	-1.40
07/2015	47.5	(3.57%)	4.83%	(0.21%)	3.30%	8.04%	(29.58%)	-1.29
08/2015	58.1	(3.24%)	3.77%	0.34%	2.87%	6.72%	(29.03%)	-0.92
09/2015	62.5	(5.26%)	3.65%	(0.10%)	1.91%	5.71%	(32.45%)	-0.74
10/2015	25.6	(2.53%)	2.57%	0.33%	(0.10%)	4.54%	(38.70%)	-0.53
11/2015	18.5	0.52%	2.42%	(0.37%)	(3.07%)	6.53%	(49.55%)	0.00
12/2015	0	(10.31%)	1.52%	(0.24%)	(3.72%)	7.09%	(44.34%)	0.26
01/2016	38.7	(23.56%)	1.18%	1.08%	(2.65%)	7.85%	(24.47%)	0.40
02/2016	17.9	(23.31%)	5.29%	(0.26%)	(3.02%)	14.41%	(31.85%)	0.53
03/2016	14.2	(8.82%)	(0.09%)	0.20%	(6.03%)	17.22%	(39.77%)	0.51
04/2016	26.5	(5.38%)	(2.90%)	0.18%	(1.99%)	13.75%	(27.02%)	0.41
05/2016	16.1	(2.27%)	(0.82%)	(0.05%)	(1.06%)	10.13%	(33.01%)	0.48



06/2016	75.9	0.00%	(1.20%)	0.69%	0.99%	7.59%	(8.31%)	0.42
07/2016	82.2	2.33%	(1.27%)	(0.49%)	2.35%	5.23%	(2.34%)	0.47
08/2016	95.4	(6.82%)	(0.81%)	0.30%	5.10%	2.58%	(0.03%)	0.45
09/2016	66.6	10.98%	(3.50%)	(0.09%)	0.99%	1.08%	10.49%	0.51
10/2016	15.1	15.38%	(3.06%)	0.69%	0.14%	0.82%	25.27%	0.52
11/2016	7.2	11.43%	(1.76%)	0.07%	0.34%	0.55%	18.39%	0.43
12/2016	44.1	7.69%	(2.33%)	(0.03%)	4.07%	(4.50%)	82.54%	0.42
01/2017	18.7	11.11%	(3.52%)	0.76%	(0.76%)	(4.20%)	39.86%	0.38
02/2017	0	7.14%	(6.05%)	0.37%	(6.88%)	(2.74%)	40.39%	0.39
03/2017	36.4	2.67%	(0.99%)	0.53%	1.92%	(5.94%)	65.73%	0.64
04/2017	0	7.79%	0.42%	0.20%	1.28%	(4.84%)	58.31%	0.85
05/2017	27.2	7.23%	0.40%	0.18%	2.44%	(4.38%)	59.58%	1.03
06/2017	50	3.37%	2.29%	(0.02%)	(0.33%)	(3.89%)	12.71%	1.06
07/2017	65.9	2.72%	1.20%	(0.11%)	(1.15%)	(3.60%)	4.41%	1.08
08/2017	40.9	(3.17%)	1.73%	0.65%	(5.35%)	(2.63%)	1.07%	1.11
09/2017	25	2.19%	5.22%	0.22%	(4.45%)	(1.87%)	(2.37%)	1.10
10/2017	8	(3.74%)	7.03%	0.07%	0.92%	(2.55%)	(4.71%)	1.16
11/2017	36.6	(3.89%)	8.17%	0.93%	2.15%	(3.43%)	17.07%	1.28
12/2017	46.6	5.20%	10.15%	0.15%	0.70%	(3.70%)	(24.58%)	1.41
01/2018	44.6	2.75%	9.90%	0.38%	8.05%	(7.18%)	16.56%	1.40
02/2018	20.2	(3.74%)	10.35%	0.35%	5.95%	(10.20%)	(7.42%)	1.55
03/2018	50.2	4.44%	9.73%	(0.37%)	1.79%	(10.82%)	(8.93%)	1.65
04/2018	75.4	2.66%	10.64%	0.54%	1.93%	(13.39%)	(12.54%)	1.71
05/2018	26.1	2.59%	11.50%	(0.19%)	3.84%	(11.58%)	(13.28%)	1.77
06/2018	78.8	(2.53%)	10.58%	(0.15%)	2.81%	(9.80%)	(2.82%)	1.89
07/2018	74.2	(3.11%)	13.95%	0.28%	2.02%	(9.10%)	(8.18%)	1.92
08/2018	85.2	(1.60%)	15.76%	0.34%	5.74%	(8.33%)	0.23%	2.01
09/2018	80.6	1.63%	14.14%	(0.40%)	4.91%	(7.78%)	(1.09%)	2.17
10/2018	35.1	2.67%	13.85%	0.45%	3.25%	(7.07%)	11.85%	2.30

ANNEX III. VARIABLES MATRIX

11/2018	73.8	0.52%	12.53%	(0.49%)	2.60%	(8.38%)	35.20%	2.40
12/2018	31.4	2.59%	12.10%	0.26%	(0.05%)	(4.38%)	40.66%	2.53
01/2019	37.3	0.51%	12.81%	0.06%	(4.41%)	(2.11%)	(21.81%)	2.49
02/2019	43.2	(2.01%)	11.11%	0.09%	0.98%	(3.94%)	0.21%	2.46
03/2019	55.4	(1.03%)	10.22%	0.74%	1.64%	(3.45%)	7.05%	2.40
04/2019	14.7	(2.07%)	11.40%	(0.39%)	(1.61%)	2.59%	(6.08%)	2.41
05/2019	54.4	(2.65%)	10.50%	0.08%	(2.14%)	3.29%	(7.31%)	2.41
06/2019	44.3	(2.72%)	11.00%	(0.62%)	(4.83%)	4.37%	(19.82%)	2.19
07/2019	73.6	(3.91%)	8.21%	0.10%	0.23%	5.24%	(17.15%)	2.17
08/2019	73.5	(3.49%)	7.55%	(0.10%)	(2.23%)	5.77%	(26.12%)	2.02
09/2019	100	(8.43%)	7.69%	(0.15%)	0.93%	6.66%	(15.39%)	1.96
10/2019	37.3	(8.55%)	7.16%	(0.17%)	(0.29%)	7.22%	(30.34%)	1.66
11/2019	71.8	(6.47%)	7.65%	0.26%	(1.63%)	8.17%	(37.18%)	1.64
12/2019	28.6	(1.54%)	6.81%	0.43%	(0.67%)	7.06%	(44.54%)	1.61
01/2020	13.2	(7.81%)	6.52%	(4.51%)	(3.78%)	10.01%	(35.66%)	1.63
02/2020	24.7	(6.78%)	10.03%	0.36%	(0.10%)	11.53%	(31.09%)	1.41
03/2020	24.7	(3.64%)	5.89%	(0.13%)	(3.77%)	15.57%	(39.83%)	0.69
04/2020	53.2	(12.26%)	1.78%	(9.12%)	(4.10%)	13.32%	(34.52%)	0.50
05/2020	29.8	(15.05%)	(4.06%)	0.72%	(7.33%)	11.89%	(33.46%)	0.48
06/2020	55.3	(3.80%)	(2.49%)	5.02%	(0.55%)	10.05%	(33.19%)	0.40
07/2020	91	(5.26%)	(0.66%)	3.03%	0.69%	8.40%	(27.24%)	0.25
08/2020	88.1	(2.78%)	(2.85%)	1.30%	(1.15%)	7.29%	2.13%	0.26
09/2020	37.8	4.29%	(3.87%)	1.38%	(5.34%)	5.59%	(26.67%)	0.08
10/2020	38.4	0.00%	(5.30%)	0.80%	(3.80%)	2.22%	(4.74%)	0.17
11/2020	11.1	1.37%	(4.47%)	1.16%	(3.13%)	4.20%	(3.45%)	-0.23
12/2020	35.1	8.11%	(3.67%)	1.64%	1.59%	2.21%	11.76%	-0.29
01/2021	25.8	7.50%	(2.52%)	1.40%	1.53%	0.47%	29.50%	-0.42
02/2021	60.6	5.81%	(13.07%)	(0.27%)	0.72%	(3.29%)	160.03%	-0.48
03/2021	26.7	1.10%	(2.79%)	0.48%	0.77%	(3.43%)	39.38%	-1.56

04/2021	22.3	2.17%	1.03%	0.10%	3.23%	(5.39%)	44.13%	-1.80
05/2021	19	6.38%	6.86%	(1.18%)	4.99%	(5.49%)	57.08%	-2.00
06/2021	69.1	(3.00%)	5.40%	1.29%	5.28%	(6.70%)	90.25%	-1.83
07/2021	52.5	5.15%	3.67%	(0.12%)	(1.73%)	(6.45%)	109.19%	-1.89
08/2021	89.1	(1.96%)	6.18%	(0.44%)	3.19%	(7.05%)	68.45%	-1.80
09/2021	38.6	1.00%	0.00%	(0.82%)	4.03%	0.00%	155.75%	-1.81
10/2021	4.6	0.00%	0.00%	0.60%	0.00%	0.00%	134.30%	-1.70
11/2021	40	0.00%	0.00%	1.33%	0.00%	0.00%	(100.00%)	-1.85
12/2021	40	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	-1.85

# Annex IV

## Python Code

```
"""
Created on Mon Dec 26 12:35:33 2022

@author: paula
"""

### Importing packages and importing NG Price from Excel

import numpy as np
import pandas as pd
import openpyxl as xl

import os
os.getcwd()

from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.seasonal import seasonal_decompose
import matplotlib.pyplot as plt
from sklearn.metrics import mean_squared_error
from statsmodels.tools.eval_measures import rmse
import warnings
warnings.filterwarnings("ignore")
```

```
import time

from datetime import date
from datetime import datetime
today = date.today()

df=pd.read_excel('NGPrice4.xlsx')
df.head(10)

df.isnull().sum()
df=df.fillna(method='pad')
df.isnull().sum()

from os import listdir
listdir()

ax = df['Natural gas, US price'].plot(figsize = (16,5), title =
'Monthly Historical Natural Gas Price (USD) Sep 2000 - Dec
2021')
ax.set(xlabel='Dates', ylabel='(USD)');
print("Today's date:", today)

df.describe()

### Separating train and test data

train_data=df[0:200]
test_data=df[200:]
predict_data=df[120:]

print('Observations: %d' % (len(df)))
print('Train Dataset:', train_data.shape)
print('Test Dataset:', test_data.shape)
```

```
ax= train_data.plot(figsize=(10,5))
test_data.plot(ax=ax, color='r')
plt.legend(['train', 'test'])
```

### **### Normalize Data**

```
from sklearn.preprocessing import MinMaxScaler
```

```
scaler = MinMaxScaler(feature_range = (0,1))
train_data_scaled = scaler.fit_transform(train_data)
print(train_data_scaled); print(train_data_scaled.shape)
```

### **### Create a data structure with 6 time lags and 1 output**

```
X_train = []
y_train = []
for i in range(6, len(train_data_scaled)):
    X_train.append(train_data_scaled[i-6:i,0])
    y_train.append(train_data_scaled[i,0])
X_train, y_train = np.array(X_train), np.array(y_train)
print(X_train); print(); print(y_train)
```

### **### Reshape Data**

```
X_train = np.reshape(X_train, (X_train.shape[0],
X_train.shape[1], 1))
print(X_train.shape); print(); print(X_train)
```

### **### Recurrent Neural Network**

```
#from keras.preprocessing.sequence import TimeseriesGenerator
#from keras.models import Sequential
#from keras.layers import Dense
#from keras.layers import LSTM
```

```
from tensorflow.keras.callbacks import EarlyStopping
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers

model = tf.keras.Sequential()

# adding 1st LSTM layer and some dropout regularization
model.add(tf.keras.layers.LSTM(units=200,
input_shape=(X_train.shape[1], 1), return_sequences=True,
activation = 'relu'))

model.add(tf.keras.layers.Dropout(0.2))

# adding 2nd LSTM layer and some dropout regularization
model.add(tf.keras.layers.LSTM(units=200,
return_sequences=True))

model.add(tf.keras.layers.Dropout(0.2))

# adding 3rd LSTM layer and some dropout regularization
model.add(tf.keras.layers.LSTM(units=200,
return_sequences=True))

model.add(tf.keras.layers.Dropout(0.2))

# adding 4th LSTM layer and some dropout regularization
model.add(tf.keras.layers.LSTM(units=200))

model.add(tf.keras.layers.Dropout(0.2))

# adding output layer
model.add(tf.keras.layers.Dense(units=1))

#compiling RNN
model.compile(loss='mse', optimizer='adam')

early_stopping = EarlyStopping(monitor='loss', patience=16)

# fitting RNN on training set
model.fit(X_train, y_train, epochs= 50, batch_size=1,
        verbose=2, callbacks=[early_stopping])

### Predict Natural Gas Price Sep 2010-Dec 2021
dataset_total = pd.concat((train_data, test_data), axis=0)
print(dataset_total)
```

```
dataset_total = pd.concat((train_data, test_data), axis=0)

# inputs = dataset_total[len(dataset_total) - len(test_data) -
6:].values

inputs = dataset_total[len(dataset_total) - 136 - 6:].values
inputs = inputs.reshape(-1,1)
inputs = scaler.transform(inputs) # transforming input data

X_test = []
y_test = []

for i in range (6, 142):
    X_test.append(inputs[i-6:i, 0])
    y_test.append(train_data_scaled[i,0])

X_test, y_test = np.array(X_test), np.array(y_test)
X_test = np.reshape(X_test, (X_test.shape[0], X_test.shape[1],
1))
print (X_test)

pred_price = model.predict(X_test)
pred_price = scaler.inverse_transform(pred_price)
print(pred_price)

### Compare predicted values to real values
a = pd.DataFrame(pred_price)
a.rename(columns = {0: 'Predicted'}, inplace=True);
a.index = predict_data.index
compare = pd.concat([predict_data, a],1)
compare

### Graph predicted values vs real values

plt.figure(figsize= (15,5))
```



```
plt.plot(compare['Natural gas, US price'], color = 'red', label
="Actual Natural Gas Price")

plt.plot(compare.Predicted, color='blue', label = 'Predicted
Price')

plt.title("Natural Gas Price Prediction")

plt.xlabel('Time')

plt.ylabel('Natural gas price')

plt.legend(loc='best')

plt.show()
```

**### Compute performance metrics, R2 and MSE**

```
import math

from sklearn.metrics import r2_score

test_score = math.sqrt(mean_squared_error(compare['Natural gas,
US price'], compare.Predicted))

print('Test Score: %.2f RMSE' % (test_score))

# Explained variance score: 1 is perfect prediction

print('Variance score (test): %.2f' % r2_score(predict_data,
pred_price))
```

# Annex V

## Predicted vs. Actual Prices

	Henry Hub	Machine Learning	BM AR Model	ARX Model 1
Sep-10	3.89	4.51	4.47	4.40
Oct-10	3.43	4.16	4.10	3.98
Nov-10	3.71	3.75	3.66	3.46
Dec-10	4.25	3.73	3.93	3.78
Jan-11	4.49	4.05	4.39	4.31
Feb-11	4.09	4.34	4.63	4.50
Mar-11	3.97	4.13	4.24	4.08
Apr-11	4.24	4.05	4.15	4.00
May-11	4.31	4.23	4.39	4.28
Jun-11	4.54	4.31	4.45	4.34
Jul-11	4.42	4.55	4.67	4.60
Aug-11	4.06	4.50	4.54	4.46
Sep-11	3.90	4.25	4.21	4.09
Oct-11	3.57	4.07	4.07	3.92
Nov-11	3.24	3.75	3.76	3.60
Dec-11	3.17	3.43	3.46	3.30
Jan-12	2.67	3.24	3.39	3.17
Feb-12	2.51	2.87	2.93	2.69
Mar-12	2.17	2.65	2.78	2.56
Apr-12	1.95	2.39	2.44	2.18
May-12	2.43	2.19	2.23	1.98
Jun-12	2.46	2.28	2.68	2.51
Jul-12	2.95	2.30	2.70	2.50
Aug-12	2.84	2.58	3.15	3.01
Sep-12	2.85	2.60	3.05	2.87
Oct-12	3.32	2.67	3.05	2.86
Nov-12	3.54	2.96	3.49	3.34
Dec-12	3.34	3.20	3.69	3.56
Jan-13	3.33	3.18	3.50	3.34

*ANNEX V. PREDICTED VS. ACTUAL PRICES*

Feb-13	3.33	3.19	3.49	3.37
Mar-13	3.81	3.20	3.49	3.36
Apr-13	4.17	3.53	3.93	3.86
May-13	4.04	3.88	4.26	4.19
Jun-13	3.83	3.91	4.15	4.01
Jul-13	3.62	3.82	3.95	3.83
Aug-13	3.43	3.68	3.76	3.61
Sep-13	3.62	3.50	3.57	3.41
Oct-13	3.68	3.54	3.75	3.64
Nov-13	3.64	3.56	3.80	3.69
Dec-13	4.24	3.54	3.75	3.64
Jan-14	4.71	3.99	4.32	4.28
Feb-14	6.00	4.50	4.75	4.73
Mar-14	4.90	5.67	5.93	5.97
Apr-14	4.66	5.07	4.91	4.87
May-14	4.58	4.96	4.69	4.59
Jun-14	4.59	4.86	4.62	4.52
Jul-14	4.05	4.85	4.63	4.54
Aug-14	3.91	4.39	4.11	3.97
Sep-14	3.92	4.15	3.99	3.86
Oct-14	3.78	4.01	4.02	3.91
Nov-14	4.12	3.84	3.89	3.75
Dec-14	3.48	4.04	4.19	4.10
Jan-15	2.99	3.60	3.57	3.40
Feb-15	2.87	3.25	3.14	2.94
Mar-15	2.83	3.01	3.02	2.85
Apr-15	2.61	2.85	2.98	2.82
May-15	2.85	2.66	2.77	2.59
Jun-15	2.78	2.69	3.01	2.84
Jul-15	2.84	2.65	2.94	2.81
Aug-15	2.77	2.68	3.00	2.89
Sep-15	2.66	2.65	2.93	2.82
Oct-15	2.34	2.59	2.82	2.70
Nov-15	2.09	2.43	2.51	2.37
Dec-15	1.93	2.26	2.28	2.11
Jan-16	2.28	2.12	2.13	1.96
Feb-16	1.99	2.18	2.46	2.30
Mar-16	1.73	2.08	2.16	1.97
Apr-16	1.92	1.95	1.91	1.71
May-16	1.92	1.96	2.10	1.97
Jun-16	2.59	1.96	2.12	1.99
Jul-16	2.82	2.24	2.72	2.63
Aug-16	2.82	2.45	2.93	2.86
Sep-16	2.99	2.56	2.93	2.89
Oct-16	2.98	2.71	3.10	3.04
Nov-16	2.55	2.77	3.08	3.01

*ANNEX V. PREDICTED VS. ACTUAL PRICES*

---

Dec-16	3.59	2.59	2.65	2.58
Jan-17	3.30	3.11	3.67	3.66
Feb-17	2.85	3.05	3.37	3.31
Mar-17	2.88	2.87	2.96	2.83
Apr-17	3.10	2.82	3.02	2.96
May-17	3.15	2.91	3.20	3.14
Jun-17	2.98	2.96	3.24	3.19
Jul-17	2.98	2.88	3.06	2.96
Aug-17	2.90	2.86	3.08	2.98
Sep-17	2.98	2.81	3.00	2.85
Oct-17	2.88	2.83	3.08	2.93
Nov-17	3.01	2.78	2.99	2.89
Dec-17	2.82	2.83	3.11	3.02
Jan-18	3.87	2.74	2.88	2.76
Feb-18	2.67	3.35	3.94	3.94
Mar-18	2.69	2.80	2.80	2.72
Apr-18	2.80	2.74	2.82	2.71
May-18	2.80	2.70	2.89	2.78
Jun-18	2.97	2.69	2.91	2.81
Jul-18	2.83	2.77	3.05	2.95
Aug-18	2.96	2.70	2.92	2.79
Sep-18	3.00	2.77	3.07	2.97
Oct-18	3.28	2.81	3.10	3.01
Nov-18	4.09	3.00	3.38	3.28
Dec-18	4.04	3.60	4.17	4.11
Jan-19	3.11	3.75	4.00	3.92
Feb-19	2.69	3.28	3.18	3.02
Mar-19	2.95	2.98	2.84	2.73
Apr-19	2.65	2.93	3.04	2.95
May-19	2.64	2.72	2.77	2.63
Jun-19	2.40	2.62	2.75	2.61
Jul-19	2.37	2.44	2.53	2.35
Aug-19	2.22	2.37	2.50	2.38
Sep-19	2.56	2.27	2.37	2.23
Oct-19	2.33	2.38	2.70	2.60
Nov-19	2.65	2.30	2.47	2.35
Dec-19	2.22	2.44	2.77	2.64
Jan-20	2.02	2.27	2.38	2.27
Feb-20	1.91	2.16	2.18	2.04
Mar-20	1.79	2.06	2.07	1.94
Apr-20	1.74	1.97	1.95	1.80
May-20	1.75	1.91	1.90	1.77
Jun-20	1.63	1.88	1.91	1.79
Jul-20	1.77	1.82	1.77	1.71
Aug-20	2.30	1.85	1.89	1.82
Sep-20	1.92	2.05	2.43	2.37

*ANNEX V. PREDICTED VS. ACTUAL PRICES*

---

Oct-20	2.39	1.98	2.06	1.97
Nov-20	2.61	2.18	2.37	2.32
Dec-20	2.59	2.33	2.69	2.65
Jan-21	2.71	2.40	2.65	<u>2.64</u>
Feb-21	5.35	2.50	2.77	2.75
Mar-21	2.62	3.99	5.03	5.13
Apr-21	2.66	2.83	2.68	2.64
May-21	2.91	2.74	2.72	2.71
Jun-21	3.26	2.73	2.98	3.00
Jul-21	3.84	3.01	3.30	3.28
Aug-21	4.07	3.44	3.84	3.75
Sep-21	5.16	3.71	4.07	4.02
Oct-21	5.51	4.75	5.07	5.08
Nov-21	5.05	5.36	5.41	5.40
Dec-21	3.76	5.25	5.31	5.30

# Bibliography

1. **Diaz, Elena Maria and Perez-Quiros, Gabriel.** GEA tracker: A daily indicator of global economic activity. *Journal of International Money and Finance*. 2021.
2. **Lopez, Paula Albendea.** *Modelling the natural gas market*. Madrid : Universidad Pontificia de Comillas, 2022.
3. **Horner, Matthew, Pakzad, Shamim N. and Gulgec, Nur Sila.** *Parameter Estimation of Autoregressive-Exogenous and Autoregressive Models Subject to Missing Data Using Expectation Maximization*. Bethlehem, PA, United States : Department of Civil and Environmental Engineering, Lehigh University, 2019.
4. **Clark, T.E., West, K.D.** Approximately normal tests for equal predictive accuracy in nested models. *J. Econometrics*. 2007, Vol. 138, págs. 291–311.
5. **EIA.** *Monthly Energy Review*. Washington D.C. : U.S. Energy Information Administration, 2022.
6. **NOAA National Centers for Environmental Information.** Residential Energy Demand Temperature Index (REDTI) - Overview. *NOAA Website*. [Online] 2022. <https://www.ncdc.noaa.gov/societal-impacts/redti/overview>.
7. **Federal Reserve Bank of Atlanta.** Wu-Xia Federal Funds Rate. [Online] 2022. <https://www.atlantafed.org/cqer/research/wu-xia-shadow-federal-funds-rate>.
8. **Gill, Dee.** The ‘Shadow Rate’ Can Measure the Effects of QE. *Chicago Booth Review*. [Online] February 9, 2017. <https://www.chicagobooth.edu/review/shadow-rate-can-measure-effects-qe#:~:text=Chicago%20Booth's%20Jing%20Cynthia%20Wu,3%20percent%20thorough%20mid%2D2014..>
9. **Baumeister, C. and Hamilton, J.D.** *Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks*. s.l. : American Economic Review, 2019. pp. 109(5): 1873-1910.
10. **OECD.** Industrial Production (indicator). *OECD Data*. [Online] 2022. <https://data.oecd.org/industry/industrial-production.htm>.

11. **York, Stephen.** Cold weather brings near record-high natural gas spot prices. *U.S. Energy Information Administration Website*. [Online] March 5, 2021. <https://www.eia.gov/todayinenergy/detail.php?id=47016>.
12. **S&P Global.** UK NPB: Natural Gas Price Assessment. *S&P Global Commodity Insights*. [Online] 2022. <https://www.spglobal.com/commodityinsights/es/our-methodology/price-assessments/natural-gas/uk-nbp-natural-gas-price-assessments>.
13. **Cai, Yifei and Wu, Yanrui.** *UNDERSTANDING GAS PRICING MECHANISMS: IMPLICATIONS FOR THE ASIAN MARKET*. Tokyo : Asian Development Bank Institute, 2020.
14. **Chant, Tim de.** \$11 billion in 9 days—Texas’ natural gas sellers cashed in on deep freeze. *Ars Technica*. [Online] December 7, 2021. <https://arstechnica.com/tech-policy/2021/07/11-billion-in-9-days-texas-natural-gas-sellers-cashed-in-on-deep-freeze/>.
15. **Chen, James and Silberstein, Samantha.** Henry Hub. *Investopedia*. [Online] August 25, 2021. [https://www.investopedia.com/terms/h/henry\\_hub.asp](https://www.investopedia.com/terms/h/henry_hub.asp).
16. **Croft, Cameron P.** How do you process natural gas? *Croft Production Systems Website*. [Online] May 29, 2020. <https://www.croftsystems.net/oil-gas-blog/https-www-croftsystems-net-oil-gas-blog-how-do-you-process-natural-gas/>.
17. **Natural Gas Supply.** [Online] 2018. <http://naturalgas.org/business/supply/>.
18. **Campbell, J.Y., Thompson, S.B.** *Predicting excess stock returns out of sample: Can anything beat the historical average?* 2008. pp. 1509– 1531.
19. **BP.** *Statistical Review of World Energy - 70th Edition*. 2021.
20. **EIA.** Natural Gas Explained. *U.S. Energy Information Administration*. [Online] December 2, 2021. <https://www.eia.gov/energyexplained/natural-gas/>.
21. **MET Group.** Natural Gas Transportation: How Is Natural Gas Transported? *MET Group Website*. [Online] September 3, 2020. <https://group.met.com/en/media/energy-insight/natural-gas-transportation>.
22. **S&P Global.** Platts JKM LNG Price Assessment. *S&P Global Commodity Insights*. [Online] 2022. <https://www.spglobal.com/commodityinsights/en/our-methodology/price-assessments/lng/jkm-japan-korea-marker-gas-price-assessments>.

23. **Processing Natural Gas.** [Online] 2018.  
<http://naturalgas.org/naturalgas/processing-ng/>.
24. **TTF Gas - 4 things you always wanted to know but were too afraid to ask.** *Enkumo Website.* [Online] December 2, 2020.  
<https://enkumo.com/2020/12/02/ttf-gas-4-things-you-always-wanted-to-know-but-were-too-afraid-to-ask/>.
25. **Borunda, Alejandra.** Natural gas is a much 'dirtier' energy source than we thought. *National Geographic Website.* [Online] February 19, 2020.  
<https://www.nationalgeographic.com/science/article/super-potent-methane-in-atmosphere-oil-gas-drilling-ice-cores>.
26. **World Bank Group.** *Commodity Markets Outlook: Causes and Consequences of Metal Price Shocks.* Washington DC : World Bank Group, 2021.
27. **Energy Education.** **Compressed Natural Gas.** *Energy Education Website.* [Online] June 25, 2018.  
[https://energyeducation.ca/encyclopedia/Compressed\\_natural\\_gas](https://energyeducation.ca/encyclopedia/Compressed_natural_gas).
28. **Union of Concerned Scientists U.S.A.** Environmental Impacts of Natural Gas. *UCSUSA Website.* [Online] June 19, 2014.  
<https://www.ucsusa.org/resources/environmental-impacts-natural-gas>.
29. **MET Group.** **LNG: Definition, Storage, Transportation and Usage.** *MET Group Website.* [Online] August 5, 2020.  
<https://group.met.com/en/media/energy-insight/lng-liquefied-natural-gas>.
30. —. **Natural Gas Environmental Impact: Problems and Benefits.** *MET Group Website.* [Online] September 30, 2020. <https://group.met.com/en/mind-the-fyouture/mindthefyouture/natural-gas-environmental-impact>.
31. **IEA.** **Gas 2020. 2021-2025: Rebound and beyond.** *IEA.* [Online] 2021.  
<https://www.iea.org/reports/gas-2020/2021-2025-rebound-and-beyond>.
32. —. **Gas 2020. 2021-2025: Rebound and beyond.** *IEA.* [Online] 2021.  
<https://www.iea.org/reports/gas-2020/2021-2025-rebound-and-beyond>.
33. **Kelisto Redacción.** **Guerra entre Rusia y Ucrania: ¿cómo afecta el conflicto al precio del gas, la luz y los combustibles?** *Kelisto.* [Online] March 7, 2022.  
<https://www.kelisto.es/gas/consejos-y-analisis/subida-gas-rusia-ucrania#:~:text=En%20su%20caso%2>.



34. Kilic, Zeynep Beyza. Los precios del gas en Europa se incrementan un 60% en dos semanas después de que Rusia redujo suministro. *Agencia Anadolu*. [En línea] June 23, 2022. <https://www.aa.com.tr/es/econom%C3%ADa/los-precios-del-gas-en-europa-se-incrementan-un>.
35. ICE. Dutch TTF Gas Futures. *ICE*. [Online] 27 de June de 2022. <https://www.theice.com/products/27996665/Dutch-TTF-Gas-Futures/data?marketId=5396828&span=2..>
36. Pascual, Lorenzo y Ruiz, Esther. *Predicción de series temporales basada en Machine Learning: aplicaciones económicas y financieras*. 2020.
37. BBVA. 'Machine learning': ¿qué es y cómo funciona? *BBVA*. [Online] November 8, 2019. <https://www.bbva.com/es/machine-learning-que-es-y-como-funciona/>.
38. Salguero, Ricardo Alonzo Fernández. *Series Temporales Avanzadas: Aplicación de Redes Neuronales para el Pronóstico de Series de Tiempo*. 2021.
39. Pedrycz, Shyi-Ming and Chen, Witold. *Deep Learning: Algorithms and Applications*. s.l. : Springer, 2020.
40. Kelleher, John D. *Deep learning*. s.l. : MIT press, 2019.
41. Gulli, Sujit and Pal, Antonio. *Deep learning with Keras*. s.l. : Packt Publishing Ltd, 2017.
42. Shukla, Kenneth and Fricklas, Nishant. *Machine learning with TensorFlow*. s.l. : Manning Greenwich, 2018.
43. Li, Zhifeng, et al. *Deep learning with long short-term memory for time series prediction*. s.l. : IEEE Communications Magazine, 2019. pp. 57(6):114–119.
44. Courville, Yoshua, et al. *Deep learning*. s.l. : MIT press Cambridge, 2016. Vol. 1.
45. Sharp, Tom. An Introduction to Support Vector Regression (SVR). *Towards Data Science*. [Online] March 3, 2020. <https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2>.
46. Wei, Yujie. *Quantifying Variation of Subsampling-based Ensemble Methods Under a U-statistic Framework*. Wellesley College : s.n., 2021.
47. Kingma, Diederik P. and Ba, Jimmy Lei. *ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION*. 2015.

48. **Yathish, Vishal.** Loss Functions and Their Use In Neural Networks. *Towards Data Science*. [Online] August, 4 2022. <https://towardsdatascience.com/loss-functions-and-their-use-in-neural-networks-a470e703f1e9>.
49. **Brownlee, Jason.** Dropout Regularization in Deep Learning Models with Keras. *Machine Learning Mastery*. [Online] July, 6 2022. <https://machinelearningmastery.com/dropout-regularization-deep-learning-models-keras/>.
50. **Sim, Victor.** LSTMs or CNNs for predicting Stock Prices? *Medium*. [Online] January, 26 2021. <https://medium.com/hands-on-data-science/lstms-or-cnns-for-predicting-stock-prices-2974c0c8c4ef>.
51. **EIA.** Natural Gas Spot Price. *eia.gov*. [Online] 2022. <https://www.eia.gov/dnav/ng/hist/rngwhhdM.htm>.

