POLITECNICO DI MILANO

Scuola di Ingegneria Industriale e dell'Informazione

Dipartimento di Chimica, Materiali e Ingegneria Chimica "Giulio Natta"

Corso di Laurea Magistrale in Ingegneria Chimica

A hybrid economic/econometric model of crude oil prices for PSE/CAPE applications under historical fluctuations and market uncertainties

Relatore: Prof. Davide MANCA

Tesi di Laurea Magistrale di:

Valentina DEPETRI matr. 813268

Anno Accademico 2014 – 2015

"Don't never prophesy: If you prophesies right,

ain't nobody going to remember

and if you prophesies wrong,

ain't nobody going to let you forget"

Mark Twain (1835-1910)

Contents

Figures

Tables

Acronyms

APARCH	Asymmetric Power Autoregressive Conditional Heteroskedasticity model
AR	Autoregressive model
ARCH	Autoregressive Conditional Heteroskedasticity model
ARMA	Mixed Autoregressive Moving Average model
BRIC	Brazil, Russia, India, China
CAPE	Computer Aided Process Engineering
CFTC	Commodity Futures Trading Commission
CO	Crude Oil
DAISY	Differential Algebra for Identifiability of Systems
DCD	Dynamic Conceptual Design
EGARCH	Exponential Generalized Autoregressive Conditional Heteroskedastic model
EIA	Energy Information Administration
EROI	Energy Return on Energy Investment
EUR	Estimated Ultimate Recoverable quantity
FIM	Fisher Information Matrix
GARCH	Generalized Autoregressive Conditional Heteroskedastic model
GDP	Gross Domestic Product
GJR-GARCH	Glosten-Jagannathan-Rukle Generalized Autoregressive Conditional Heteroskedastic model
IEA	International Energy Agency

MA	Moving Average model
NYMEX	New York Mercantile Exchange
OECD	Organization for Economic Cooperation and Development
OPEC	Organization of the Petroleum Exporting Countries
OPEX	OPerating EXpeditures
PSE	Process Systems Engineering
UAE	United Arab Emirates
USC/bbl	US Cent/Barrel
USD/bbl	US Dollar/Barrel
WTI	West Texas Intermediate

Abstract (English)

The main driver for this thesis was to model and forecast the quotations of crude oil (CO), which is the reference component in most if not all chemical supply chains. In order to understand the forces that cause price fluctuations, the work analyzes some historical events that in last decade influenced CO market. After a presentation of the state-of-the-art, some existing models are simulated and evaluated in terms of time-granularity, forecast horizon, and type of explanation provided on quotation trend (*i.e.* econometric vs economic models). One of the most important remarks in the Sections dedicated to econometric and economic models is that the implementation of economic models over long-time horizons presents some problematic issues, such as the need for supply-and-demand variable forecast. Conversely, the econometric models do not care of the forces that cause price fluctuations. Hence, the thesis proposes a revised economic model (called "OPEC-based model") to forecast quarterly prices of CO over short- and medium-term horizons so to take into account recent variations of both Brent and WTI quotations. The new economic model is credited by its own power to include the contributions from both CO producers and consumers, which are clustered into OPEC and OECD organizations. In addition, this model considers also supply-and-demand variables. The obtained results are useful for Dynamic Conceptual Design problems (*e.g.* design of chemical plants under market uncertainty and similarly production/allocation planning/scheduling). These problems call for the creation of possible future scenarios according to stochastic variations of markets, country economic development, political/economic decisions at international level. The OPEC-based model can be manipulated in order to create future scenarios with an overall bullish or bearish trend. Finally, the thesis shows and simulates a hybrid model that combines the OPEC-based quarterly model with a monthly econometric model. The thesis focuses on the dynamic evolution of raw material prices as a function of the real market demand, global supply, geographical localization, market uncertainties, and historical background.

Abstract (Italian)

L’obiettivo della presente tesi è stato la modellazione e la previsione dei prezzi del greggio, precursore di riferimento dell’industria chimica. Al fine di individuare le cause delle sue fluttuazioni, la tesi analizza alcuni eventi che hanno influenzato il mercato petrolifero nell’ultimo decennio. Dopo aver presentato lo stato dell’arte, alcuni modelli sono stati simulati e valutati in termini di granularità e orizzonte temporale e del tipo di spiegazione fornita per l’andamento delle quotazioni (modelli econometrici ed economici). Una delle osservazioni più importanti fatte nelle Sezioni dedicate ai modelli economici ed econometrici è che i primi presentano alcune problematiche su lunghi orizzonti temporali, come la necessità di prevedere le variabili che si riferiscono alla domanda e all’offerta. Al contrario, i modelli econometrici non si preoccupano delle cause delle oscillazioni dei prezzi. La tesi propone quindi una rivisitazione di un modello economico (noto come modello OPEC) per la previsione dei prezzi trimestrali del greggio su orizzonti temporali medio-brevi, al fine di tenere in considerazione i recenti andamenti di Brent e WTI. Il nuovo modello economico è accreditato dalla sua capacità di comprendere sia i produttori sia i consumatori, raggruppati in modo semplificato nelle organizzazioni OPEC e OECD, e le variabili relative a domanda e offerta. I risultati ottenuti sono utili per il *Dynamic Conceptual Design* (*e.g.*, per valutazioni economiche soggette alle incertezze di mercato o *scheduling*/*planning*). Questi problemi hanno richiesto la creazione di possibili scenari futuri basati su variazioni stocastiche o su decisioni politiche/economiche internazionali. Il modello OPEC può essere manipolato al fine di creare scenari rialzisti o ribassisti. Infine, la tesi presenta e simula un modello ibrido, nato dalla combinazione del modello OPEC con un nuovo modello econometrico che lavora con quotazioni mensili e in media mobile. La tesi continua gli studi sull’evoluzione dinamica dei prezzi di materie prime quali funzioni della reale domanda di mercato, l’offerta globale, la posizione geografica, le incertezze di mercato e il panorama storico.

Introduction and aim of the work

Crude oil (CO) is one of the most basic and globally distributed raw materials that are usually taken to oil refineries and petrochemical plants to separate hydrocarbon fractions by distillation and produce derivatives by various chemical treatments. According to the most recent BP Statistical Review of World Energy (2015), in 2014 oil was the world's dominant fuel. Fluctuations in the CO prices have both direct and indirect impact on the global economy. Indeed, CO prices are observed and studied very closely not only by investors worldwide, but also by process designers/managers because CO is a reference component for both the Oil&Gas and petrochemical supply chains, and plays a role in a number of industrial utilities (*e.g.*, electric energy, hot water, steam). The quotations of CO are either directly (*i.e.* distillates) or indirectly (*i.e.* derived commodities) taken into account for the economic assessment and feasibility study of PSE problems such as the scheduling and planning of supply chains, and the design of chemical plants. As a function of the specific problem to be solved, the time interval chosen for the economic assessment can cover a short-, medium-, or long-term horizon (*i.e.* from hours/days to months/years). Manca (2013) showed how CO economics influences at a great extent also the quotations of commodities and utilities, which on their turn play a major role in the economic assessment of OPEX (operative expenditures) terms. Additionally, Mazzetto *et al.* (2013) used CO as the reference component for econometric models of bioprocesses and showed a functional dependency of both raw biomaterials and final bio-products from the CO market.

Due to a high degree of volatility, the real price of CO is difficult to model. As in case of other commodities, the CO price experiences significant price swings in times of shortage or oversupply. Both CO and distillate prices can be affected by exogenous events that have the potential to disrupt the flow of oil and products to market, including geopolitical and weather-related incidents (Hamilton, 2005; Zhang *et al.*, 2008). These events may lead to either actual disruptions or create uncertainty about future supply or demand, which on their turn can lead to higher volatility of prices. The aforementioned variations are driven by short-term imbalances on supply-and-demand terms and by uncertainties originated by political, economic, and financial contributions. This is the main problem of short-term horizon models, involved in scheduling problems, which cover time horizons spanning from days to few weeks. On the other hand, the medium- and long-term horizon problems are difficult to solve due to the need to forecast the different variables involved (*e.g.*, levels of supply, demand, production, and capacity storage) for a rather long period of time (from few months to some years), as far as Conceptual Design and Process Systems Engineering (PSE) are concerned. The feasibility study of chemical plants depends partially on the purchase costs of raw materials and selling prices of products. In recent years, the term Dynamic Conceptual Design (DCD) has been proposed by Grana *et al.* (2009), Manca and Grana (2010), Manca *et al.* (2011), and Manca (2013), to account for variable prices/costs over different time horizons. It results rather convenient that the price/cost evaluation of commodities and utilities should not rely on customized models specifically carried out for each of them. On the contrary, it is worth and recommended to identify a reference component and *measure* the price/cost of commodities respect to such a component. Indeed, this MSc thesis continues the past studies about the dynamic evolution of prices/costs of raw materials, and focused on the price of the reference component of the refining supply chain under financial and economic uncertainties. Since CO is the precursor of a number of commodities and utilities, its cost is well-known, largely available in several databanks such as Energy Information Administration (EIA), International Energy Agency (IEA), ICIS, and periodically updated, there is the need to collect, revise, and create a new forecast model that takes into account the physical variables that affect petroleum market trend and possible future scenarios. The contribution of this thesis allowed finalizing the paper: Manca, D., Depetri, V., Boisard. C., *A crude oil economic model for PSE applications*, Computer Aided Chemical Engineering, 37, 491-496 (2015).

Motivation and structure of the work

Aim of this thesis is to study and develop a new crude oil (CO) model for PSE/CAPE applications, as its quotations play a central role in the definition of prices of distillated products, derivatives and utilities, such as electric energy. Most of the published manuscript lacks of a clear model classification and deals with a time horizon that is well below the time horizon involved in assessing the dynamics of OPEX terms (*i.e.* at least few years).

Furthermore, an important point to note from past studies is their preoccupation with one-step-ahead models, which estimate the variable of interest for the time-step immediately following the latest one. The call for a new model comes from the need of combining the supply-and-demand forces that cause crude-oil-price trend with market stochastic fluctuations to create a common thread with real market demand, global supply, and market uncertainties. The paper presents a number of econometric and economic models proposed in the literature and compares their features with a new model specifically designed to cover the specifications of the chemical Supply Chain in terms of scheduling, planning, and economic assessment of chemical plants. Indeed, the proposed economic model results to be used to forecast the price of CO over short-, and medium-term horizons, which are the time intervals intrinsic to PSE problems such as scheduling and planning. The model that appears to fit better for PSE purposes is the revised OPEC-based model (Cooper, 2003; Kaufmann *et al.*, 2004; Dees *et al.*, 2007). Unlike previous works (*i.e.* papers and thesis), the proposed economic model that takes into account the reality by means of the supply-and-demand term provides pseudo-real quotation values to the econometric models, which can simulate price fluctuations. Indeed, OPEC-based model comprises demand, inventories, production, and other variables that take into account the supply-and-demand level of actual CO market. These variables constitute the so called model *input variables*. As the input variables provide a link between model and market reality, forecast results do not come from relative movements of prices respect to previous quotations (*i.e.* from econometric model), but from economic considerations, which investigates CO market forces more deeply. Then, a suitable econometric model may exploit these results (pseudo-real point) and create several price scenarios with the background noise that characterizes market prices. In addition, the thesis discusses also the geographical localization of CO quotations that since 2011 have shown the divergence of Brent from West Texas Intermediate (WTI) prices. This point calls for a customization of model parameters according to the geographical region of influence where the economic assessment is carried out, and to the historical background on which the recent market trend lies.

Chapter 1 provides an overview of recent historical events that affected the trend of CO quotations. The world historical ferment calls for a continuous update of the forecasting model to produce reliable price scenarios and a consistent economic assessment for PSE/CAPE applications.

In order to present the state of the art on crude-oil-quotation modeling, **Chapter 2** proposes a short review of crude-oil-price forecast in the scientific literature, by highlighting and classifying the available models. Among possible classifications, this thesis differentiates between economic and econometric models, in order to eventually propose a hybrid model capable of taking into account the pros and cons of both of them. The description of the hybrid model is postponed to Chapter 6.

Several empirical studies show evidence that time series of CO prices, likewise other financial time series, are characterized by a fat tail distribution and volatility clustering, where these features can create problems when dealing with quotations dynamics. The volatility is often regarded as a feature of economic time series and this characteristic forced the scientific literature to switch from a deterministic description of the problem to its stochastic modeling and solution (Manca and Rasello, 2014). Based on that analysis of historical prices, **Chapter 3** describes econometric models, implements an autoregressive model, and offers a statistical analysis of those data that will be used also for the economic analysis and creation of future scenarios.

Chapter 4 deals with the main issues related to a physical characterization of crude-oil-price variations, by means of the descriptions and simulations of two economic models (Ye *et al.*, 2009, and Chevallier, 2014). The call for a reliable and consistent forecast, and the need for simplicity and a reduced number of forecasting parameters in Conceptual Design and more in general in PSE applications, brought to study and revise in **Chapter 5** the so-called OPEC model (Cooper, 2003; Kaufmann *et al.*, 2004; Hamilton, 2005; Dees *et al.*, 2007). The involved variables are not easy to model and forecast because they depend on the economic activities carried out in the involved countries (either producers or consumers). In addition, the abovementioned papers are not clear on how to forecast the inventories, production capacity, and on the role of US shale oil spread. These issues call for the need of developing new models of input variables and create future scenarios of CO price, and of the variables that take into account supply-and-demand issues.

Chapter 6 proposes a new hybrid model that merges the econometric model with the economic one, in order to simulate the trend proposed by supply-and-demand law, but in combination with the stochastic fluctuations of CO quotations. The last two chapters are accompanied by figures that propose the validation and simulation of the economic and hybrid models over suitable time horizons.

Chapter 1 Historical background

This Chapter proposes the analysis of most relevant events that influenced crude oil (CO) prices in recent past, such as the 2008 financial crisis, US shale oil spread, and current European and Chinese upsets, to understand the driving forces that cause price fluctuations. In addition, we discuss the geographical localization of CO quotations that since 2011 showed the divergence of Brent and WTI benchmarks.

Chapter 1 discusses also the evolution of CO markets, with an overview of demand-and-supply levels of both producer and consumer countries, which for the sake of simplicity are clustered into OPEC, OECD, and BRIC countries. Based on this historical background, Chapter 5 proposes a new economic model that is credited by its own power to include those CO producers and consumers, together with supply-and-demand variables.

1.1 Introduction to recent historical trends of crude oil quotations

The interaction between oil price, oil supply, and oil demand is eccentric and responds to different exogenous events that may occur in the market at a specific historical period (*e.g.*, Hamilton (2005) illustrates as exogenous such events as military conflicts, economic recession, and monetary policies). This point is particularly crucial, as prices do not respond explicitly to real events, but rather to their perception. Prices rise because there *might* be a shortage of oil, not because there *is* actually one. Prices fall only when that perception changes. This issue is extensively discussed by Hamilton (2005), Chevallier (2014), Davis and Fleming (2014), Dowling *et al.* (2014), with the list of pivotal events becoming more and more extended.

Table 1 reports a qualitative list of events (*e.g.*, global tensions, local conflicts, and rumors) that played a crucial role in determining the recent fluctuations of oil markets. A CO price model should at least account for and possibly forecast the contributions introduced by those conflicts, tenses, and events that may occur all over the world. For instance, in the first months of 2011, the conflicts in Libya and the tsunami in Japan, combined with the following Fukushima nuclear disaster, played a significant role in increasing the CO prices. Similar comments can also be made for the political situation of Iraq and Iran, which impacted significantly the quotations of CO in recent years.

Table 1 - List of events from 2008 to 2013 that affected crude oil prices (data from EIA).

Time	*Event*	*Δprice [%]*	*Δtime [quarter]*	*Absolute values [USD/bbl]*
July-December 2008	Financial crisis	-69.3	3	From 133.37 to 41.12
Since 2011	Shale gas / decrease of CO imports in the USA, and too many stocks in Cushing	shift WTI/Brent	13	-
15/02/2011 and 11/03/2011	War in Libya and tsunami in Japan / Fukushima	16	1	From 88.58 to 102.76
November 2011-March 2012	Political tensions with Iran/ strikes of oil workers in Nigeria	9.3	2	From 97.13 to 106.16
May-July 2012	End of the tensions / slow growth in China	-14.6	2	From 94.65 to 87.9
June-August 2013	Threat of an American attack to Syria	11.3	1	From 95.77 to 106.57

Shale gas, shale oil, international crises, embargos, available infrastructures, industrial and transport accidents, natural calamities, and weather variability are some examples of exogenous variables that may play a major role in the fluctuations of quotations even over short-time periods with a further influence produced by complex geopolitical backgrounds (Manca and Rasello, 2014).

It was estimated that a drop of 10 USD/bbl transfers roughly half point of global Gross Domestic Product (GDP) from producer countries to consumer ones (Agnoli, 2014). According to Goldman Sachs (2014), the recent highly variable trend of CO prices produced not only a global impact, but also made oil companies afraid of breaking their own neutrality threshold between incomes and outcomes. Indeed, oil producing countries and companies count on a certain price level to cover operative expenses and financial commitments. Even if European oil majors cut their spending in 2015 in response to the plummeting oil price (with the average cut in capital expenditures estimated at around 10% by six of the region's biggest firms), most of these operators expect their oil and gas production to rise in 2015 (Figure 1). BG Group (*i.e.* a British multinational oil and gas company) had the biggest reduction among the six at 30%, Total and Eni said they would spend 30% less on exploration in 2015, Shell reported they would cut spending by USD 15 billion over the next three years, but bucked the trend by keeping exploration expenses steady in 2015 (Figure 2).

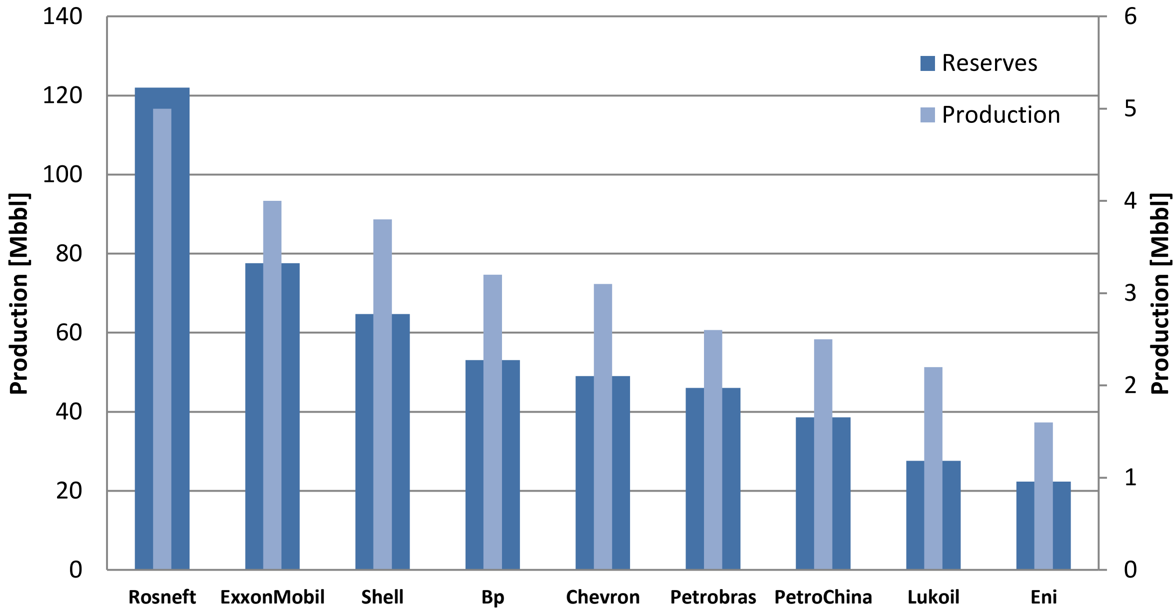


Figure 1 - Crude oil reserves and production of major refining companies (data taken from Agnoli, 2014).

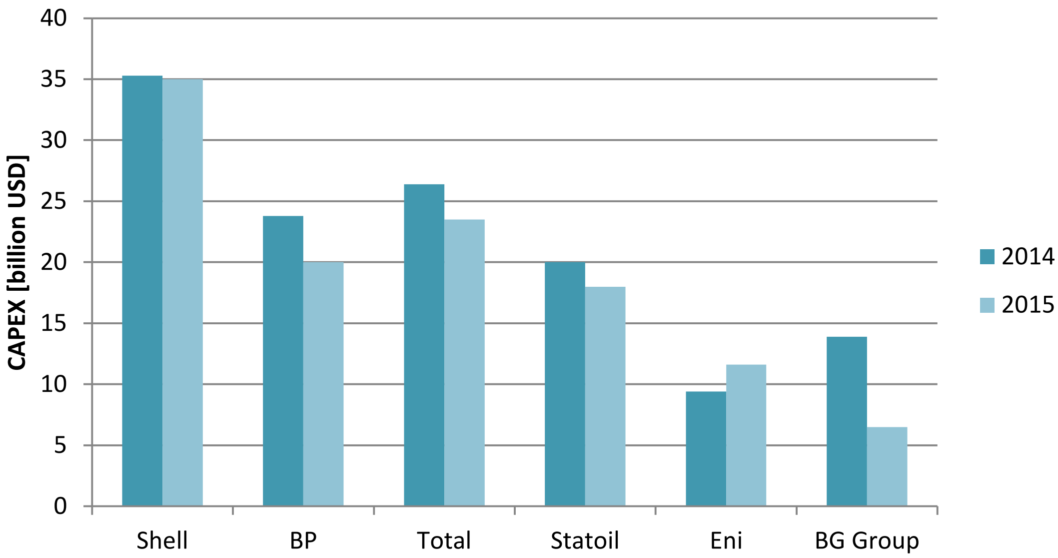


Figure 2 - Capital expenditures of major refining companies (data taken from Elliott, 2015).

Nowadays, besides this economic instability, new scenarios are emerging thanks to the production of shale oil in the USA, that has recently approached the production capacity of one of the biggest petroleum producers, *i.e.* Saudi Arabia, has already overtaken Russian gas production, and issued the first CO exporting licenses that are required for the export of CO to all destinations. For instance, the trend in US oil production is the key variable in the oil market this year and, according to a Deloitte survey (2014), several oil and natural gas industry executives believe that the United States will achieve energy independence in the next five to ten years (Bassett, 2014). In 2014 the USA exported a record of 3.8 Mbbl/d of petroleum products, up 347,000 bbl/d from the previous year, according to EIA (2015). Oil and Gas Journal (2015) reported that the production increase was driven by record-high refinery runs, which averaged 16.1 Mbbl/d in 2014, higher global demand for petroleum products, and export of motor gasoline, propane, and butane especially towards Central and South America, followed by exports to Canada and Mexico. Exports of distillate, meanwhile, declined for the first time since 2004. Most of that decrease can be ascribed to declines in exports to Western Europe and Africa, where distillate exports fell by 61,000 bbl/d and 8,700 bbl/d, respectively, in 2014 (EIA, 2015). In addition, during the second half of 2014, increased European refinery runs, exports from recently upgraded Russian refineries, and enhanced refinery capacity in the Middle East increased supply to European distillate markets, thus reducing the need for distillate from the USA.

1.1.1 The 2008 crude oil price crash

The first noteworthy event of recent CO quotation history is the financial crisis of 2008, when CO prices fell down all of a sudden due to the presence of excessive speculation (Chevallier, 2014). Indeed, it is possible to observe that the CO quotation curve (Figure 3) in second semester of 2008 saw a tremendous financial and economic calamity that was triggered by the US subprime mortgage crisis. After having trespassed the 145 USD/bbl value in July 2008, West Texas Intermediate (WTI) CO price crashed to 36 USD/bbl in December of that same year before and eventually bounced back to 76 USD/bbl in November 2009. For these reasons, the studies carried out in Chapter 3, Chapter 4, and Chapter 5 of this work start from January 2010 to avoid the impact that the financial crisis of 2008 had on petroleum markets. This anomalous trend of the last twenty years deals with the problem that CO prices are usually traded on futures market and that financial fundamentals (*e.g.*, the role of exchange and interest rates, or the commodity indexes) contribute to the petroleum market (Chevallier, 2014). This point will be clarified in Chapter 4 with the analysis and discussion on the main contributions of Chevallier's model.

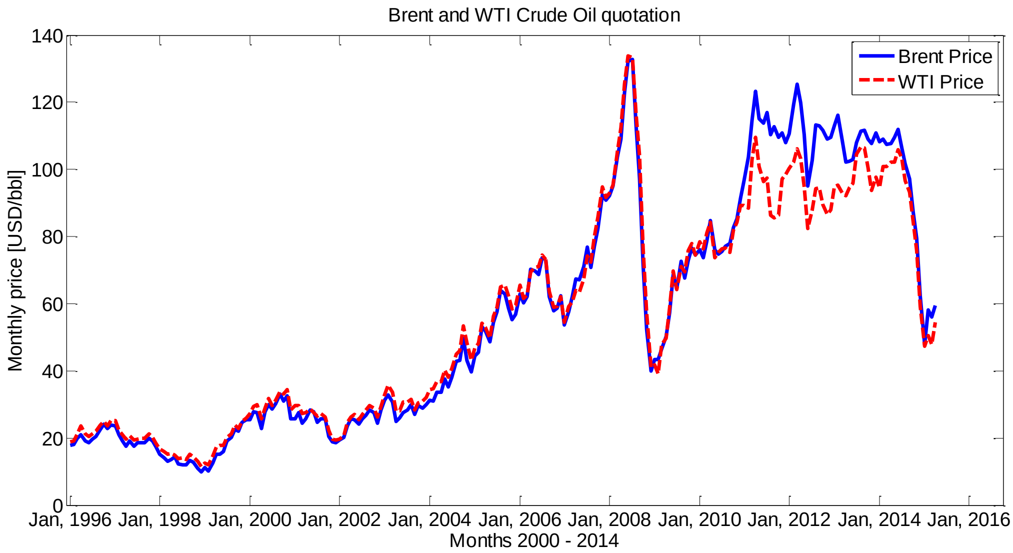


Figure 3 - Brent and WTI monthly quotations from Jan, 1996 to Apr, 2015 (data from EIA).

1.1.2 WTI and Brent divergence and shale oil "revolution"

According to Kao and Wan (2012), a CO benchmark is defined as *"the market from which the price changes first appear, and toward which the prices of other crude oils equilibrate"*. The most important global CO benchmarks are WTI and Brent. Liu *et al.* (2014) provided some details on those benchmarks. The quotations of WTI and Brent specialize in the US and Europe markets respectively. Indeed, to describe consistent scenarios according to different markets, it is reasonable to choose WTI quotations for North America refineries and Brent quotations for European refineries (Manca and Rasello, 2014).

Brent is the original name for the oil extracted from specific fields and collected through a pipeline that arrives to the Sullom Voe terminal in the Shetland Islands of Scotland. However, declining supplies from the original Brent fields led to blending with oil from the Ninian fields which took to widening the benchmark definition to include oil from Forties, Oseberg and Ekofiskls fields (hence the acronym BFOE).

WTI is a light and sweet CO that is collected from wells in Texas, New Mexico, Kansas and Oklahoma states to the storage facilities in Cushing, Oklahoma. The larger trading volumes of the New York Mercantile Exchange (NYMEX, *i.e.* the main commodity futures exchange for energy products), the CO contracts, and the fact that WTI contracts can efficiently incorporate London's information (with Brent data and thanks to 5-6 time zones between London and New York) into their dynamics, allowed WTI benchmark to become more influential than Brent regarding the quotations of other oils (Kao and Wan, 2012).

As shown in Figure 3 and Figure 4 from 2011 on, WTI and Brent quotations lost their mutual consistency (Kao and Wan, 2012; Sen, 2012; Dowling *et al.*, 2014; Liu *et al.*, 2014). As suggested by Liu *et al.* (2014), current pipeline constraints on the USA side (in particular at Cushing, Oklahoma) and shale oil spread have resulted in the divergence between Brent and WTI quotations. Indeed, Cushing has been the pricing point for WTI contracts since 1983 and now is spread over 9 square miles and has CO storage capacity around 65 kbbl (Sen, 2012). Even if the WTI – Brent price differential should be around 8-12 USD/bbl, which is the price that makes rail movement to US Gulf Coast economic, by observing Figure 4, one can conclude that recent differential values between Brent and WTI monthly quotations have been substantially larger. Hence, there are fewer opportunities to redirect oil flows out of Cushing.

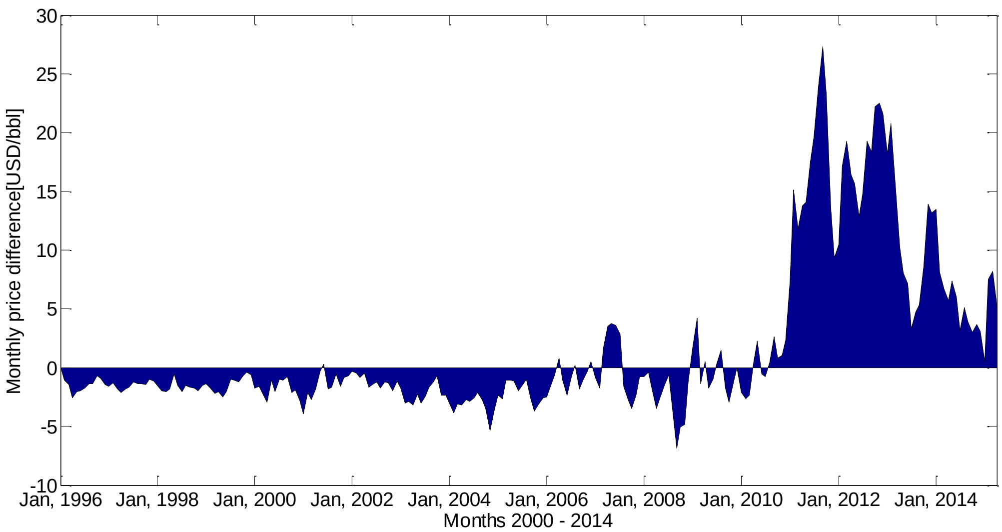


Figure 4 - Brent-WTI differential values between Brent and WTI monthly quotations (data from EIA).

According to BP Statistical Review of World Energy (2014), Brent averaged 108.66 USD/bbl in 2013 with a decline of 3.01 USD/bbl from the 2012 level. WTI continued to trade at a discount to Brent of 10.67 USD/bbl, driven by growing US production. Since 2011, the WTI discount has averaged 14.81 USD/bbl respect to Brent, compared with an average premium of 1.39 USD/bbl for the previous decade. The WTI – Brent differential narrowed to 5.66 USD/bbl in 2014 despite continued robust US production growth (BP Statistical Review of World Energy, 2015). The reason for that coming apart of their trends consists in a number of distinct but correlated reasons.

Since 2005, production of CO from conventional extraction means has not grown concurrently with demand growth, so the oil market switched to a new and different state, which can be coined as “phase transition”. Indeed, current manufacturing is “inelastic”, which means that it is unable to follow the demand fluctuations, and this pushes prices to oscillate significantly because the resources of other fossil fuels (*e.g.* tar sands, oil sands, shale oil, unconventional natural gas) do not seem to be able to fill the gap in the supply chain. The capacity to maintain and grow global supplies is attracting an increasing concern. In particular, the USA CO production curve shows a trend reversal: the curves of CO production and of net imports hit a new cross point in 2013 (Figure 7) and the USA CO production is close to Saudi Arabia one (Figure 8). As it happened twenty years ago, currently an increase in CO production can be observed because the capacity to maintain and grow a global supply attracts increasing investments in the discovery of unconventional oil reservoirs. Indeed, shale oil is an unconventional oil produced from rocks that hold deposits of organic compounds (*i.e.* kerogen), but that has not undergone enough geologic pressure, heat, and time to become conventional oil. Shale oil can be processed in two ways, *i.e. ex situ* (internal combustion and hot recycled solid technologies) and *in situ* technologies (wall conduction, ExxonMobil Electrofrac, and volumetric heating). The main difference consists in the presence of retorts used in the first ones, while *in situ* processes heat shale oil underground by injecting hot fluids into the rock formation. In the first method, shale oil is mined and brought to surface to be retorted to temperatures above 800 K (*i.e.* decomposition temperature of kerogen) in a vertical shaft retort by air or in a rotating kiln (Figure 5). Gases are removed from the top or recycled, while condensed shale oil is collected.

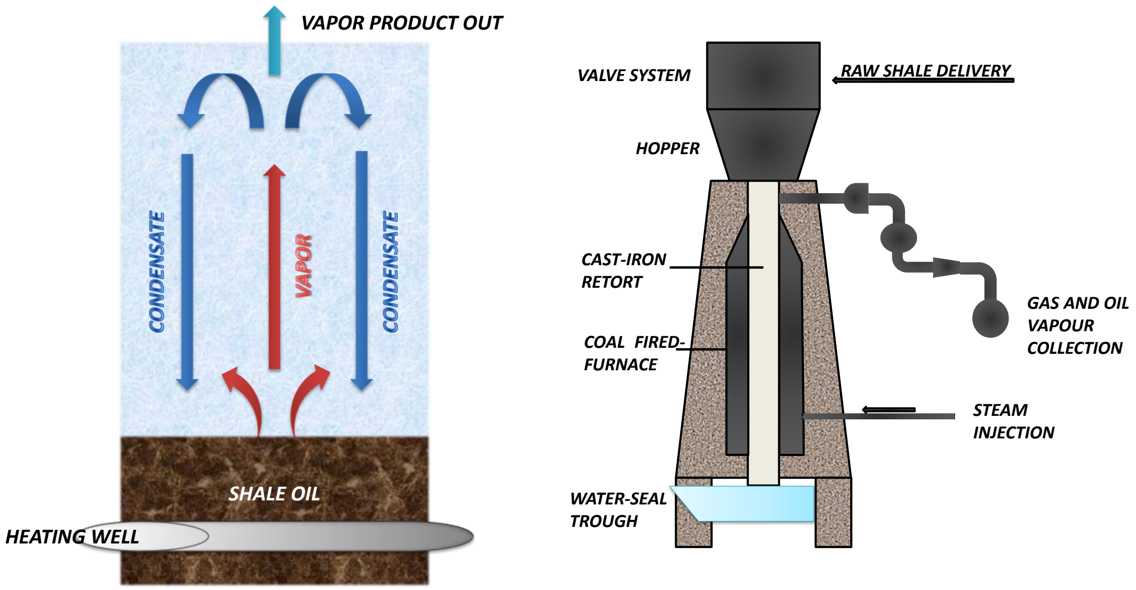


Figure 5 - Internal combustion technology for shale oil extraction and scheme of a vertical shaft retort.

Instead, *in situ* conversion process involves placing heating elements or heating pipes throughout the shale for up to three years until it reaches 600-650 K and releases the oil (Figure 6). According to Shell technology, the hot elements heat the shale in a cylindrical area of about 30 m in diameter. The liquefied oil seeps through cracks in the shale and pools in an area where it can be pumped to the surface. Meanwhile, the pressurized aqueous ammonia creates a barrier that protects surrounding ground water from contamination.

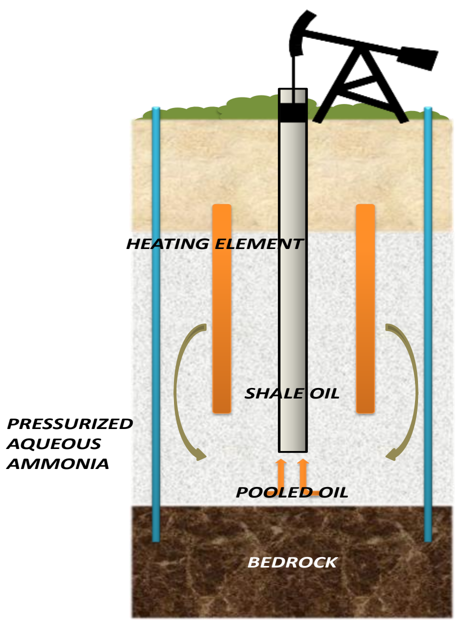


Figure 6 - In situ extraction of shale oil.

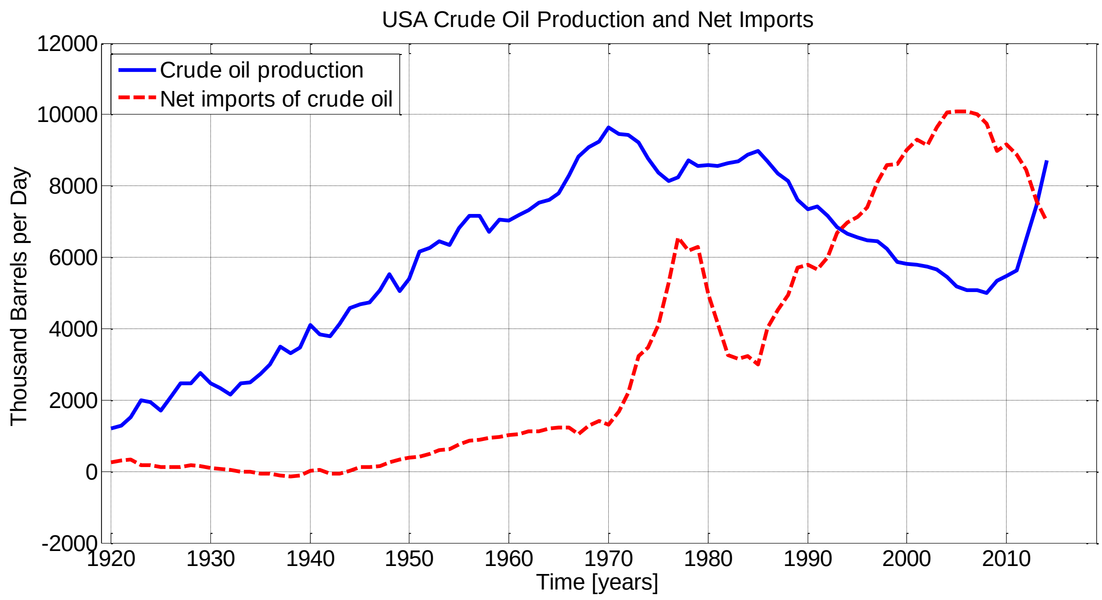


Figure 7 - US crude oil annual production and net imports from 1920 to 2010 (data from EIA).

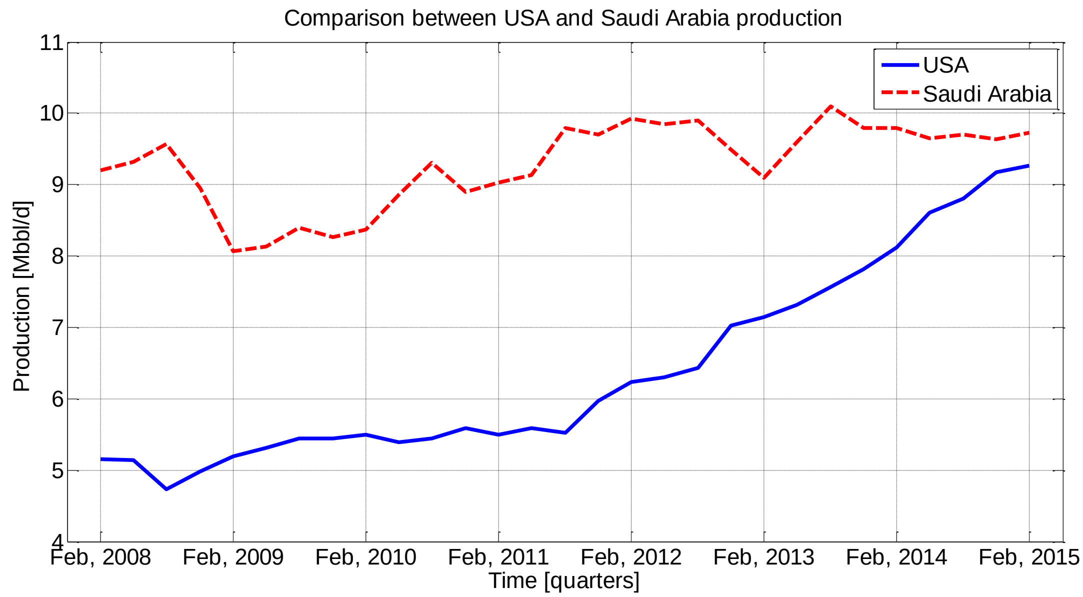


Figure 8 - Comparison between the USA crude oil production and Saudi Arabia production from Feb, 2008 to Feb, 2015 (quarterly data from EIA).

Shale oil is diffused especially in the USA, Australia, and Brazil. The extraction of shale oil in the USA, which counts for 2 trillion barrels of these unconventional reserves, started at the beginning of the last century, but since 2005 resulted in a substantial growth of total CO production, which increased from 5.1 Mbbl/d in 2008 to 9.6 Mbbl/d in 2015 (EIA, 2015). For instance, shale oil output reached a record of 5.47 Mbbl/d in March 2015 thanks to technology improvements although the number of rigs exploring for oil was the lowest since 2013. The expansion in both shale oil production in US North Dakota's Bakken fields and tar sands in Canada resulted in the reduction of WTI outward capacity from Cushing's storage facilities to the refineries on the US Gulf Coast. According to Liu *et al.* (2014), the sudden abundance of shale oil produced a substantial discount of WTI respect to Brent quotations.

1.1.3 The last price crash in December 2014

As for the recent trend of CO prices (Figure 9), second semester of 2014 saw a sharp drop from 106.7 USD/bbl to 62.34 USD/bbl in benchmark Brent crude prices, and from 103.59 USD/bbl to 59.29 USD/bbl for WTI. The fourth quarter of 2014 saw a reduction of 25% from 101.82 to 76.4 USD/bbl for Brent and from 97.8 to 73.2 USD/bbl for WTI, while in the first quarter of 2015 the price went down of 30% from 76.4 to 54 USD/bbl for Brent and from 73.2 to 48.5 USD/bbl for WTI.

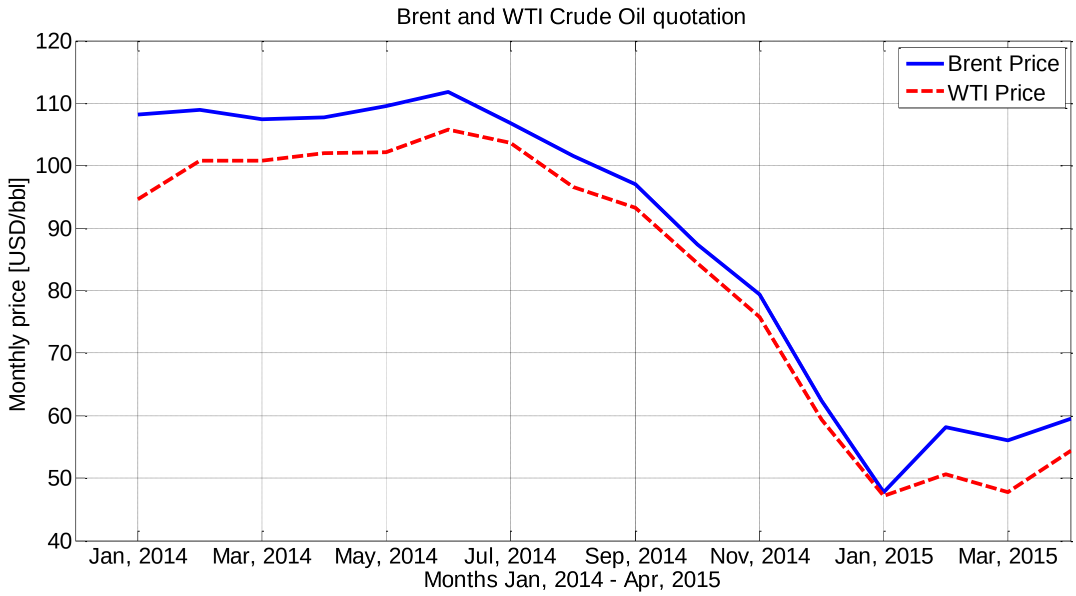


Figure 9 - Brent and WTI crude oil monthly quotations from Jan, 2014 to Apr, 2015 (data from EIA).

The roots of the last quarter crash lie on different factors, that can be distinguished in supply-and-demand factors, macroeconomic and financial factors, and political considerations: the massive Chinese reduction of import growth (from 16% in 2010 to 5% in 2014, even if China could become the world's largest importer by 2017 to outpace its domestic production according to ICIS, 2014), the stagnation of oil demand in western world and Japan, the oversupply due to the growth of Canadian oil sands and the USA fracking boom, which was already discussed, an imbalance in price ratios between oil and natural gas, the role of speculative investors, and a higher dollar respect to other currencies were responsible of the 50% drop in the price of crude from July, 2014 to December, 2014.

Brent technical floors of 70 USD/bbl and 54 USD/bbl for the price, dating back to 2010 and 2006 respectively (Davis and Fleming, 2014), were shown to be inconsistent with the recent market trend. This points out that no real support level exists in the current oil marketplace.

In front of these emerging scenarios, Middle East producers decided not to cut their quotas at the end of 2014 to maintain their production competitive with the global CO spot market. The highest authorities of Saudi Arabia declared acceptable that prices remained low for long periods if that would reduce investments in shale oil and rebalance global markets. The recent high volatility of CO prices made oil companies and some traditional producers (*i.e.* Iran, Venezuela, and Russia) afraid about breaking their own neutrality threshold between incomes and outcomes (*i.e.* their breakeven point). Indeed, the collapse of CO quotations can hurt the national budget of these producer and exporter countries, which traditionally pay public administration subsidies by means of their CO business revenues.

1.1.4 Very recent events: Greek crisis, withdrawal of Iranian embargo, and Chinese crisis of stock exchanges

The first event that is noteworthy is the Greek depression, which started in late 2009 because of the turmoil of global recession, structural weaknesses in the country economy (*e.g.*, government spending, tax evasion, lack of budget compliance), and a sudden crisis in confidence among lenders due to a misreporting of the government debt levels. By 2012, Greek debt-to-GDP ratio rose to 175%, nearly three times the EU’s limit of 60%. In last months, after three bailout programs from (i) European Commission, (ii) European Central Bank, and (iii) International Monetary Fund, Greece incurred in a governmental solvency problem and a national bankruptcy crisis that dragged down global equity and commodity markets. Greece is not a major oil producer or consumer, so it does not have a direct impact on CO markets, but this situation had two immediate effects on oil apart from supply-and-demand fundamentals. The first effect came in the form of currency fluctuations. The rising prospect that Greece would leave the Eurozone damaged the Euro currency to benefit the USD. Secondly, the calamity scared investors who feared a broader contagion with a consequent decrease of oil prices.

At the present time, the event that potentially could turn out to be a much bigger threat to the global economy than the Greek debt crisis is the stock market rout in China, with all the political, social, and economic risks it entails. Unlike most other stock markets, where investors are mostly institutional, more than 80% of investors in China are small retail investors. Although China's economy lost steam in past years (*e.g.*, according to Yao, 2015 its GDP growth rate halved from 14% in 2007 to 7.4% in 2014), the last six months have been seeing a record number of businesses listed on the Shanghai exchanges. This countercurrent event was fueled by a switch away from property investment following a clampdown by the government on excessive lending by banks, and liberalization laws that made easier for funds to invest, and for firms to offer shares to the public for the first time. At the center of the dramatic stock market slide are individual investors borrowing from a broker to buy securities, and the explosion in the so called margin lending that is a stock system by means of which the broker can make a demand for more cash or other collateral if the price of the securities has fallen. Consequently, shares plunged 30% in three weeks since mid-June, hundreds of firms suspended dealings, and fears that the slump will spill over into other markets have grown until the present time. Not surprisingly Brent quotations collapsed below the 60 USD/bbl for the first time since mid-April 2015 and closed at their lowest level in nearly three months on July 1st, 2015 (*i.e.* at 51.56 USD/bbl). This fact is particularly interesting to better understand the historical background, as China is the second largest oil consumer in the world. Indeed, commodities were sucked into this market turmoil and Chinese stock market plunged the world economy and fuel demand.

Another noteworthy event is the Iranian nuclear deal, which could bring the market to CO and petroleum products oversupply. On July 14th, 2015 Iran and the six world powers (*i.e.* United States, United Kingdom, France, China, Russia, and Germany) struck a landmark agreement to curb Tehran's nuclear activities in exchange for lifting the crippling economic and financial sanctions imposed on the country in 2012. After this decision, Iran may look to sell large volumes of polymers to Europe, while China, which became Iran's major market when the sanctions were imposed, will likely see a 30% decrease of supply from that country, according to Fadhil (2015). According to a Dubai-based petrochemical trader, CO prices are set for more volatility with Iran back, since its re-entry in European market raises oversupply concerns. For instance, on the deal day the opening price of Brent was 56.86 USD/bbl, down by 0.99 USD/bbl from the previous close, while WTI was trading at 51.24 USD/bbl, down by 0.96 USD/bbl. Earlier, the European benchmark fell to 56.61 USD/bbl and WTI went down to 50.41 USD/bbl, with a 1.79 USD/bbl contraction (Fadhil and Dennis, 2015). These data highlight that CO prices are affected by Iran production capacity and the sudden increase of its outputs must be taken into account when forecasting possible scenarios of CO production and quotations.

1.2 The crude oil market

In order to clarify the data and the terms that are used in the following Sections, this Chapter discusses the distinction between crude oil and petroleum. The *crude oil* is a mixture of unrefined hydrocarbons deposits that exists as a liquid in natural underground reservoirs, remains a liquid when brought to the surface, and can produce usable products, such as gasoline, diesel, kerosene, asphaltenes, and various forms of petrochemicals (*e.g.*, ethylene, propylene, aromatics). *Petroleum products* are produced from the processing of CO and other liquids at petroleum refineries, from the extraction of liquid hydrocarbons at natural gas processing plants, and from the production of finished petroleum products at blending facilities. Furthermore, *petroleum* is a broad category that includes both CO and petroleum products.

CO markets are characterized by the existence of a cartel together with the presence of independent producers. The main participants can be clustered into OPEC (Organization of Petroleum Exporting Countries), OECD (Organization for Economic Cooperation and Development), and BRIC countries (*i.e.* Brazil, Russia, India, and China).

OPEC is an intergovernmental organization that was created at the Baghdad Conference on September 10-14th, 1960, by Iraq, Kuwait, Iran, Saudi Arabia, and Venezuela. Later nine more governments joined OPEC: Libya, United Arab Emirates, Qatar, Indonesia, Algeria, Nigeria, Ecuador, Angola, and Gabon. OPEC decisions about quota and capacity utilization have a significant and immediate impact on oil price and some economic models were developed to comply with this hypothesis (Dees *et al.*, 2007; Cooper, 2003; Kaufmann *et al.*, 2004; Hamilton, 2005). Figure 10 shows the main importers of OPEC CO. Oil consumption comes especially from western world and Asia, where Chinese and Japanese flat economic growths have seen a stagnation of global CO demand.

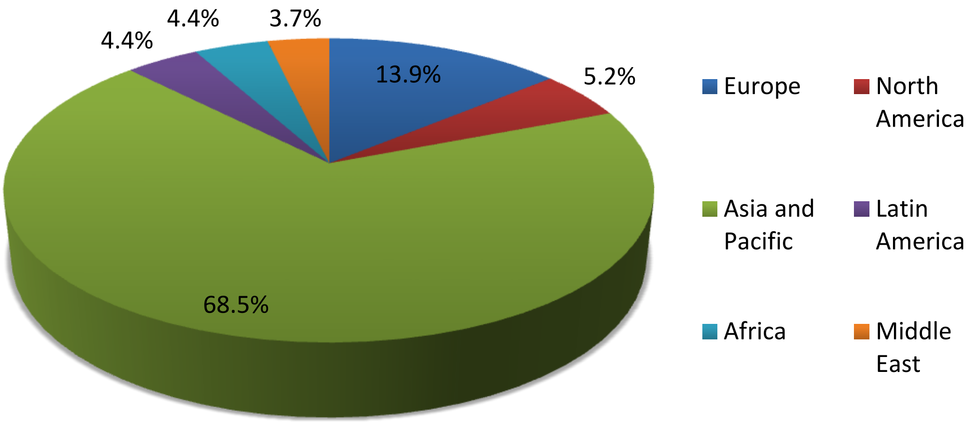


Figure 10 - OPEC exports by destination in 2013 (data taken from BP, 2014).

On the consumers' side, the present work distinguishes between OECD and BRIC countries. The first body is an international economic organization of 34 countries that was founded in 1961 to stimulate economic progress and world trade. Since OECD gathers the world most developed countries, it appears to be the most obvious consumer of CO and other petroleum products. Furthermore, the United States are obtaining a rising power on the producer side: the increasing concern about oil sands in both the USA and Canada has given a new power to the US economy that could become independent of OPEC decisions about quotas and production. Equally, the economic background is changed and is more complex than the one of 5-10 years ago, as OPEC and OECD do not include the so called BRIC countries, which are considered to be at a similar stage of newly advanced economic development, and emerging countries such as Indonesia, which abandoned OPEC from 2008 to June, 2015. Indeed, in June 2015 the Southeast Asian country re-joined OPEC, because it increased its business relations with Saudi Arabia oil companies (Suratman, 2015).

The following graphs contain information about CO proved reserves, production, consumptions, and refining capacity, with a glance at the global situation and at the individual realities of both producers and consumers organizations. From 1998 to 2013 (and previously from 1980s according to BP Statistical Report, 2014), there was an increase of OPEC and BRIC proved reserves, especially to the detriment of OECD slice of reserves (Figure 11 and Figure 12). As it is depicted in Figure 13 and Figure 14, the distribution of OPEC reserves became more homogeneous, and Saudi Arabia lost its record for the benefit of Venezuela. However, Saudi Arabia remains the leading oil producer within OPEC and is the world's largest oil exporter.

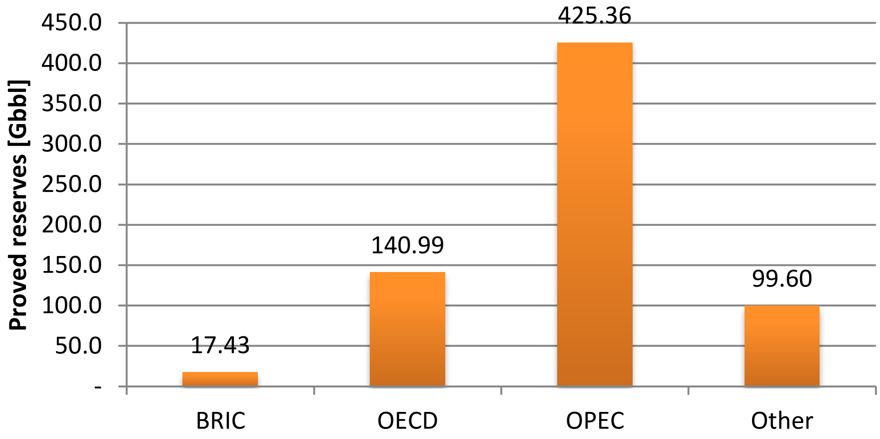


Figure 11 - Global proved reserves in 1998 (data taken from BP, 2014).

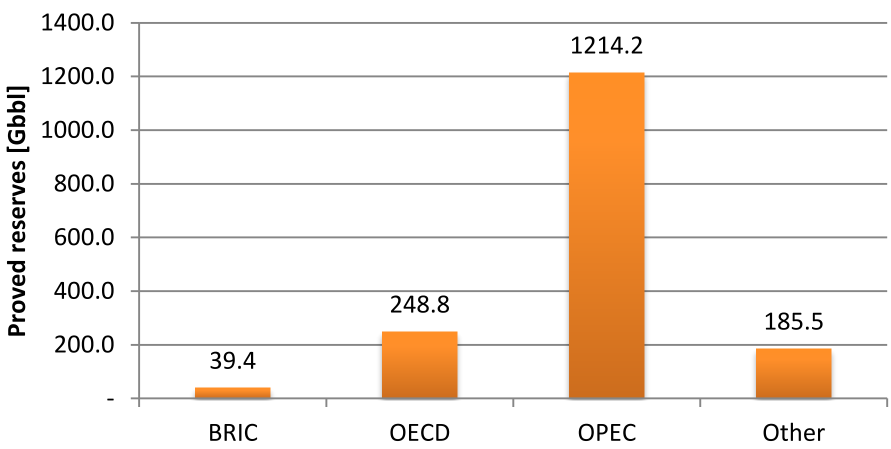


Figure 12 - Global proved reserves in 2013 (data taken from BP, 2014).

Among OECD, the country that rose its proved reserves is Canada (Figure 15 and Figure 16), while BRIC countries saw the increasing contribution of Russian and Brazilian reserves (Figure 17 and Figure 18).

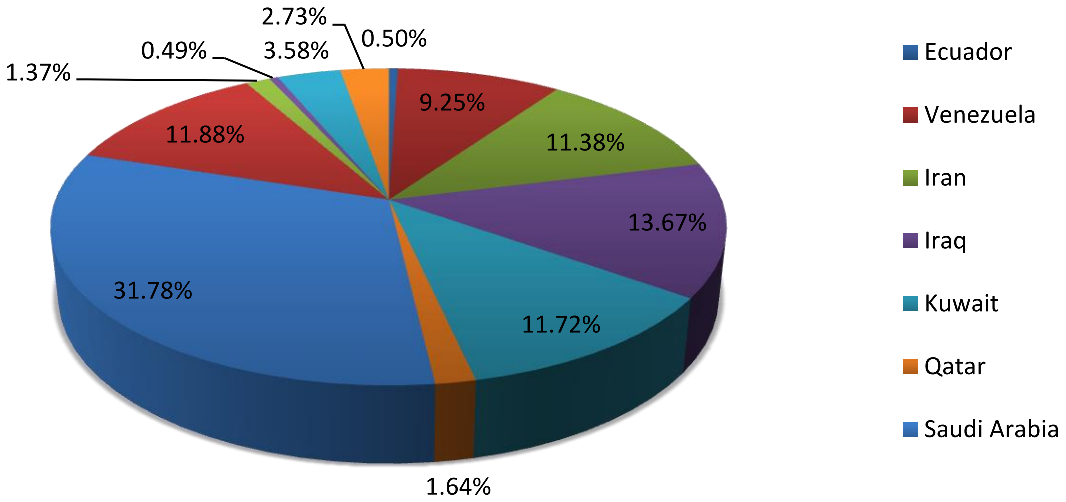


Figure 13 - OPEC proved reserves in 1998 (data taken from BP, 2014).

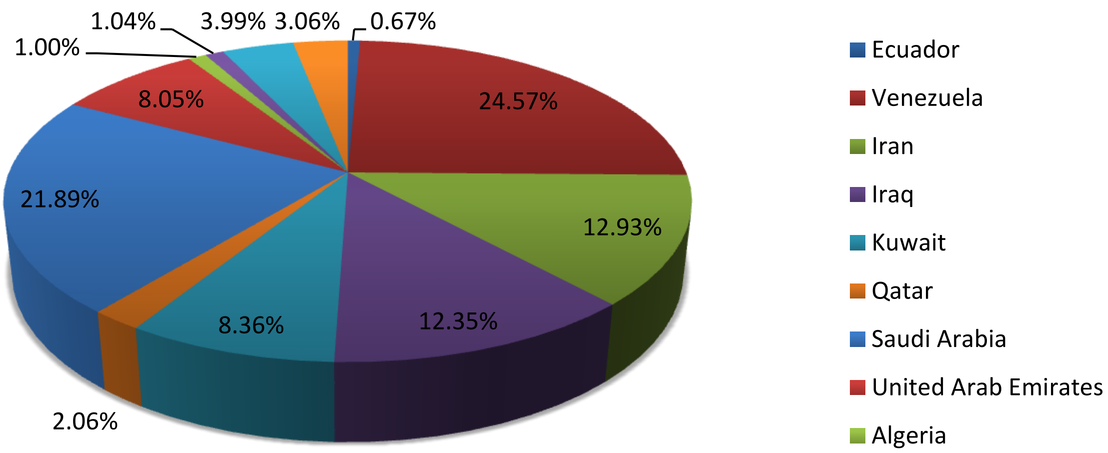


Figure 14 - OPEC proved reserves in 2013 (data taken from BP, 2014).

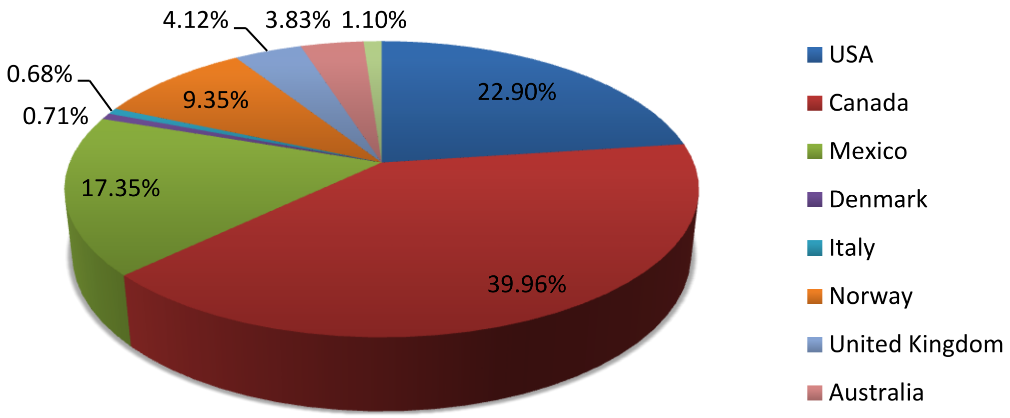


Figure 15 - OECD proved reserves in 1998 (data taken from BP, 2014).

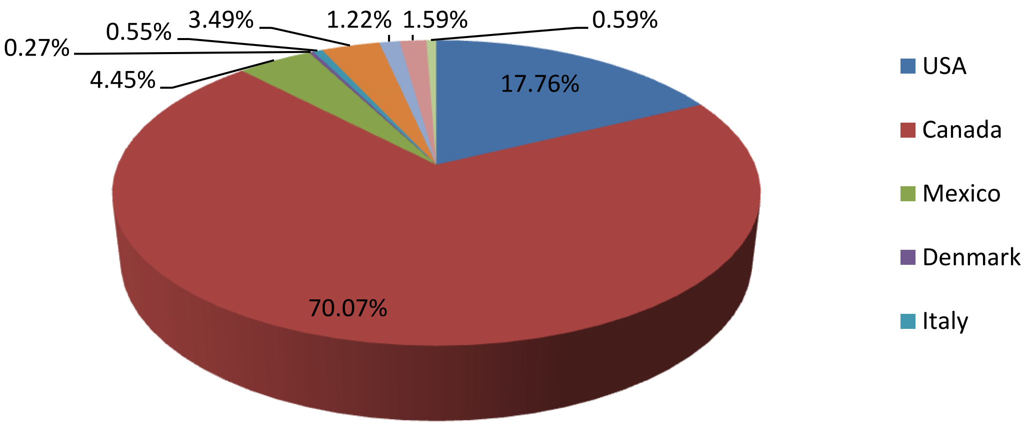


Figure 16 - OECD proved reserves in 2013 (data taken from BP, 2014).

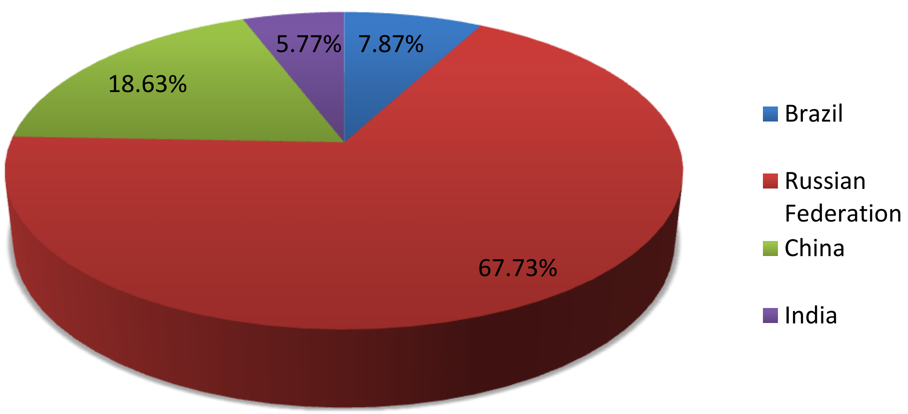


Figure 17 - BRIC proved reserves in 1998 (data taken from BP, 2014).

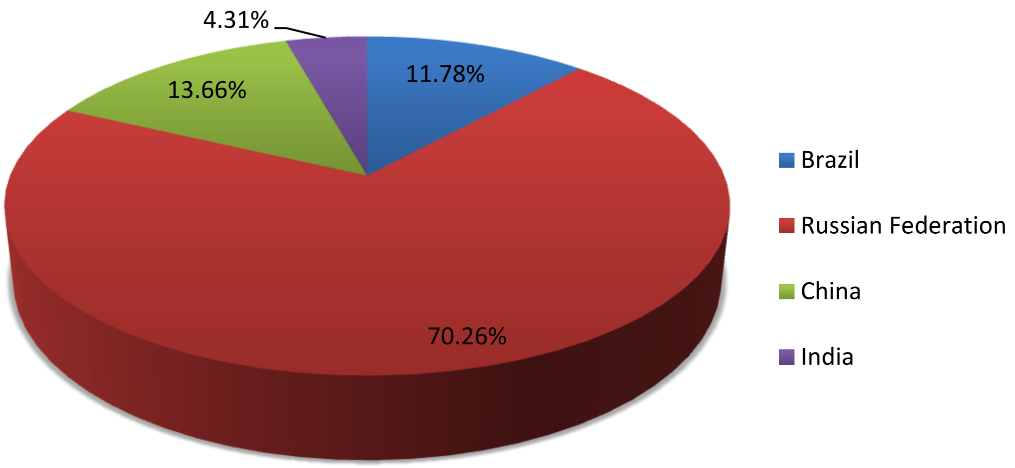


Figure 18 - BRIC proved reserves in 2013 (data taken from BP, 2014).

As for CO production, the global leading position is occupied by OPEC countries (Figure 19 and Figure 20), even if the last twenty years have seen the increase of BRIC CO production, to benefit Chinese, Russian, and Brazilian output percentage (Figure 21 and Figure 22). Among OECD countries, the USA saw an increasing production from 37% in 1998 to 49% in 2013, and in the meanwhile Canadian oil sands conferred a new power to North America CO output at the expense of Mexican, Danish, and British percentages of production (Figure 23 and Figure 24).

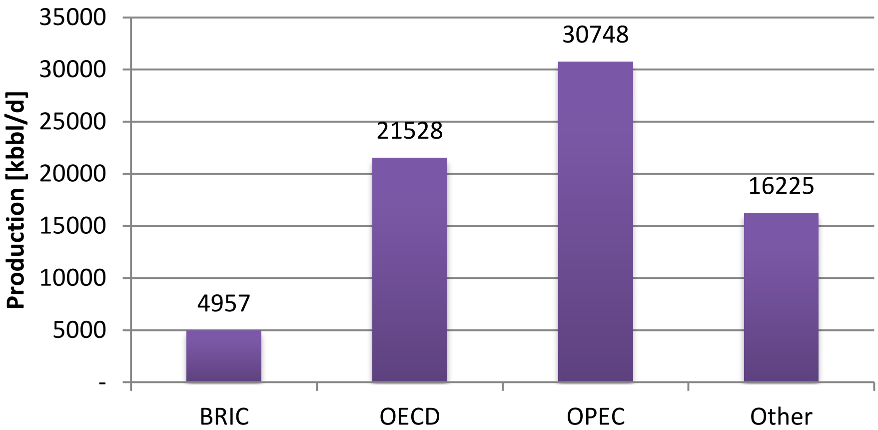


Figure 19 - Global crude oil production in 1998 (data taken from BP, 2014).

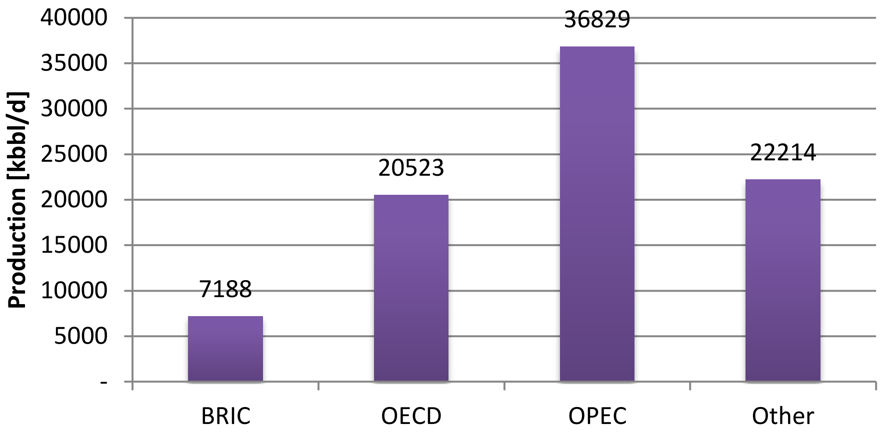


Figure 20 - Global crude oil production in 2013 (data taken from BP, 2014).

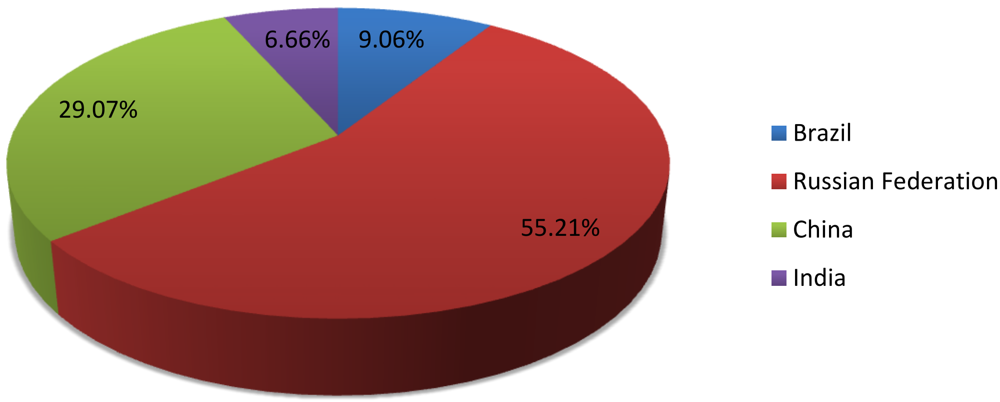


Figure 21 - BRIC crude oil production in 1998 (data taken from BP, 2014).

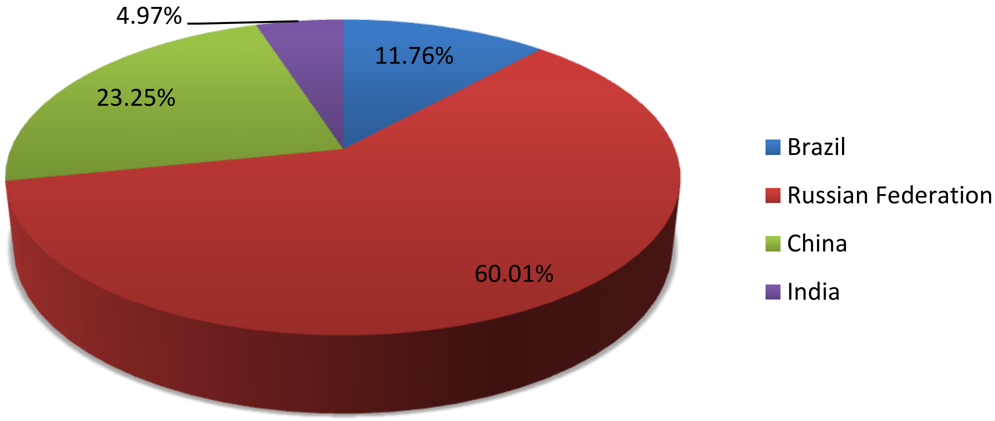


Figure 22 - BRIC crude oil production in 2013 (data taken from BP, 2014).

A different situation characterizes OPEC countries, which redeployed their own CO production for the benefit of Saudi Arabia, while Venezuela, Iran, and Nigeria cut their output fraction (Figure 25 and Figure 26).

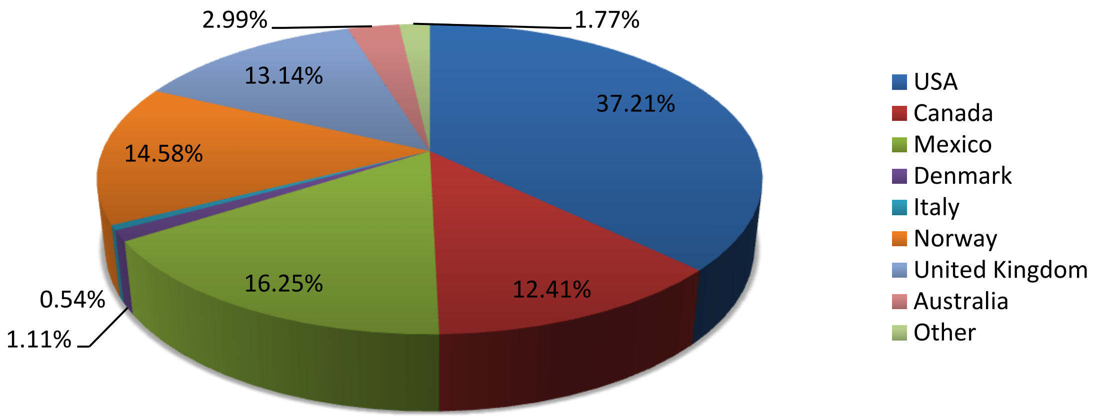


Figure 23 - OECD crude oil production in 1998 (data taken from BP, 2014).

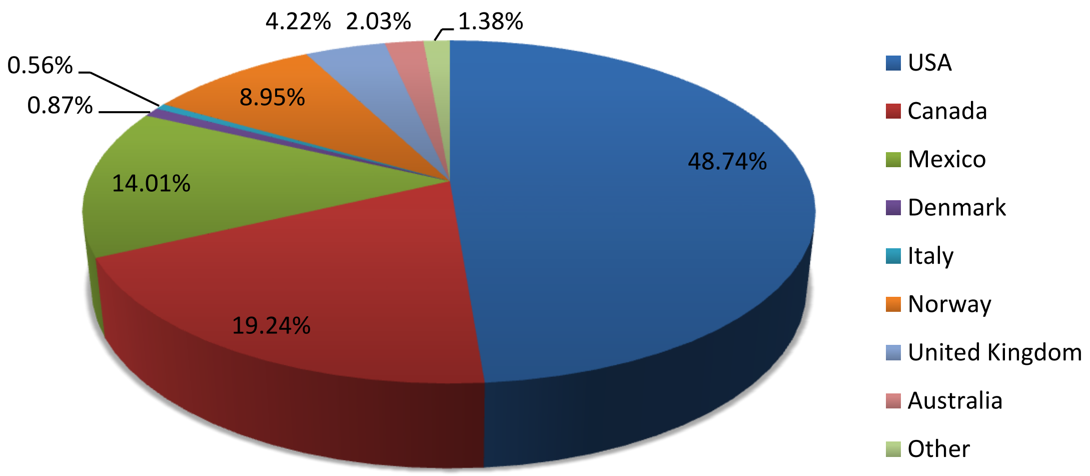


Figure 24 - OECD crude oil production in 2013 (data taken from BP, 2014).

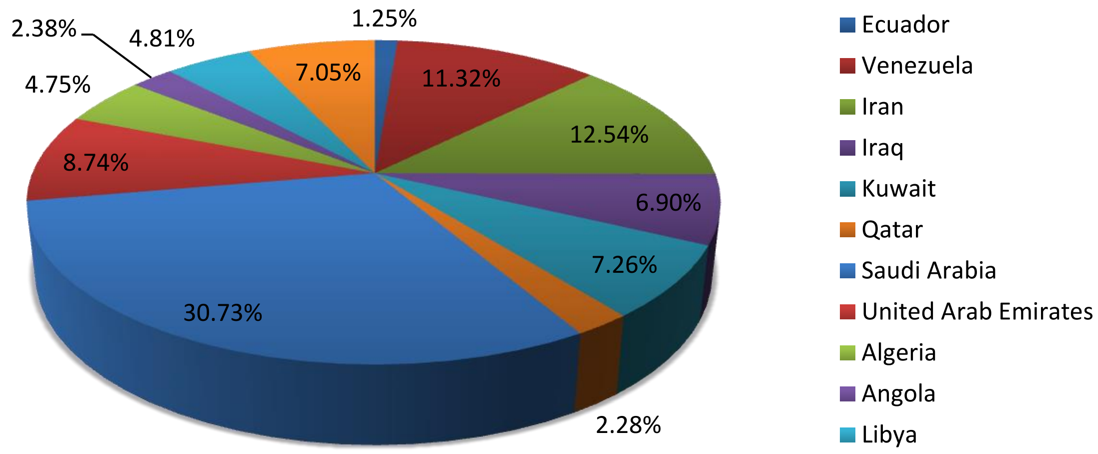


Figure 25 - OPEC crude oil production in 1998 (data taken from BP, 2014).

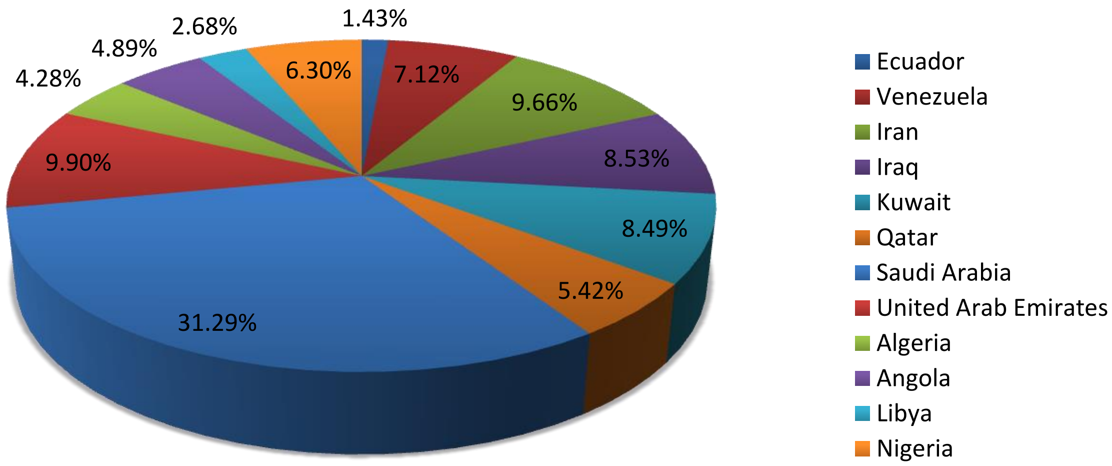


Figure 26 - OPEC crude oil production in 2013 (data taken from BP, 2014).

Figure 27 and Figure 28 show that CO consumption of OECD countries increased from 1998 to 2013, and BRIC countries enhanced their slice of demand because of the increasing growth of China, whose consumptions had passed from 20% to 37% in 1996 according to BP Statistical Report, 2014 (Figure 29 and Figure 30). As BP Statistical Report (2014) does not contain data about all OPEC countries, Figure 27 reports the consumption of some OPEC members. Figure 30 shows how the Chinese consumptions slowed in 2013, while India and Russia kept on improving respect to twenty years before.

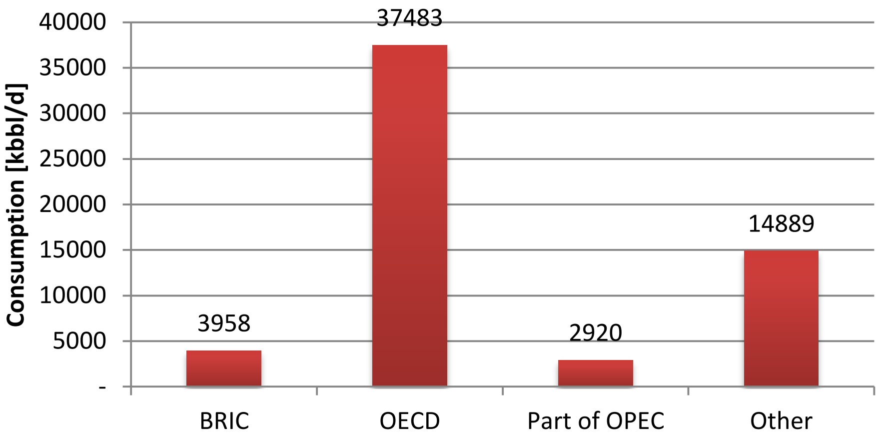


Figure 27 - Global crude oil consumption in 1998 (data taken from BP, 2014).

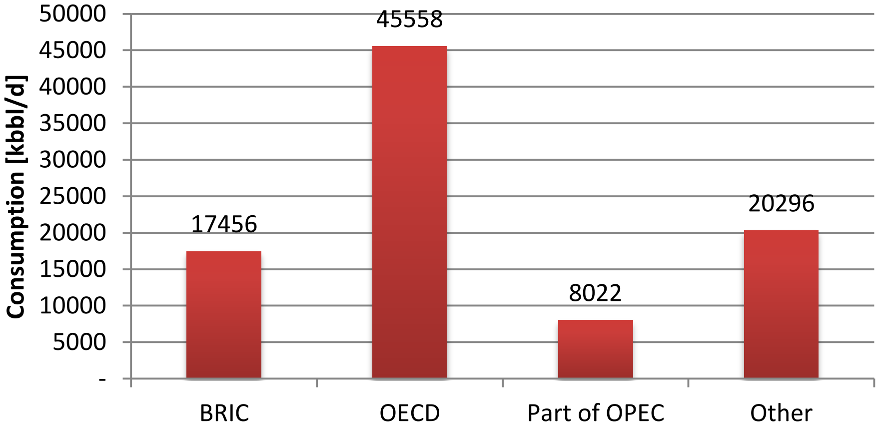


Figure 28 - Global crude oil consumption in 2013 (data taken from BP, 2014).

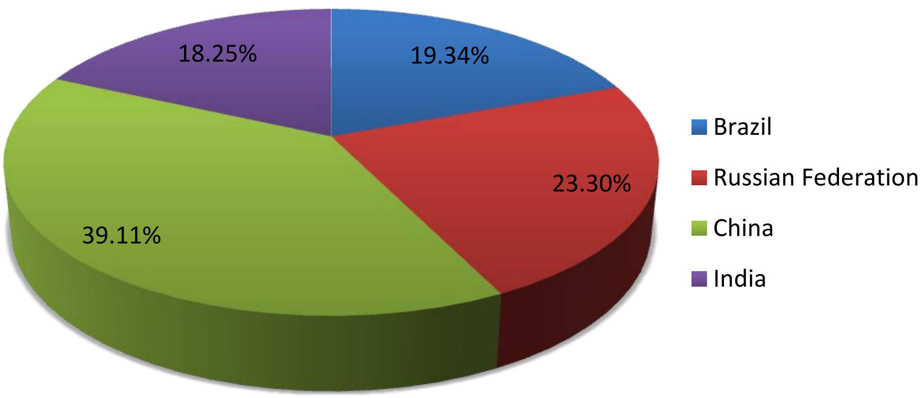


Figure 29 - BRIC crude oil consumption in 1998 (data taken from BP, 2014).

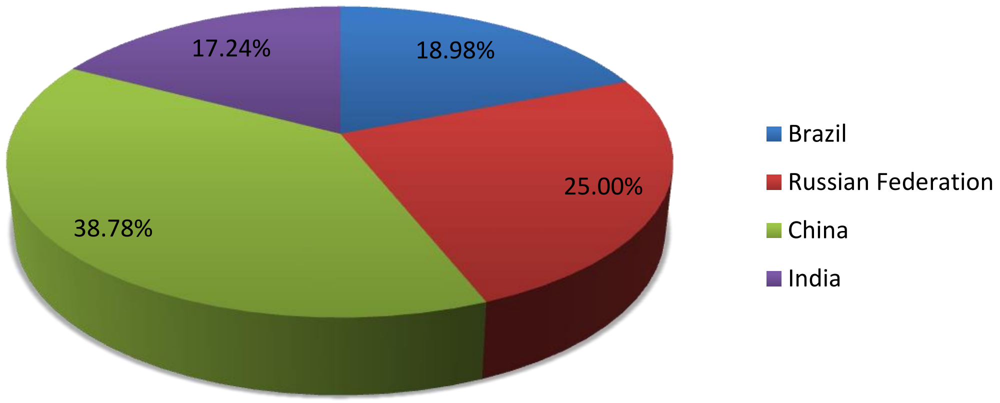


Figure 30 - BRIC crude oil consumption in 2013 (data taken from BP, 2014).

According to BP (2014), the consumptions from 1965 to 2013 distributed better for most OECD countries than the USA, which since 1998 account for 40% of OECD crude consumptions (Figure 31 and Figure 32).

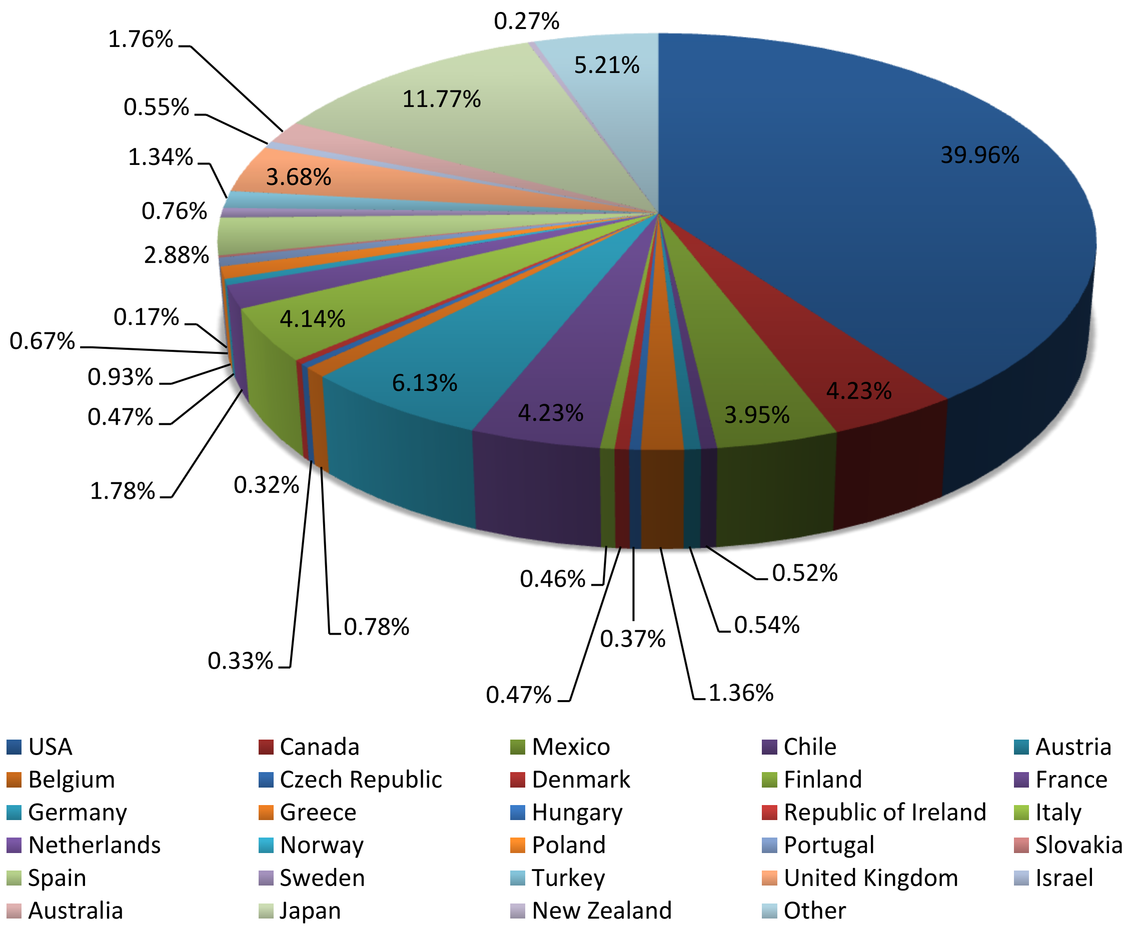


Figure 31 - OECD crude oil consumption in 1998 (data taken from BP, 2014).

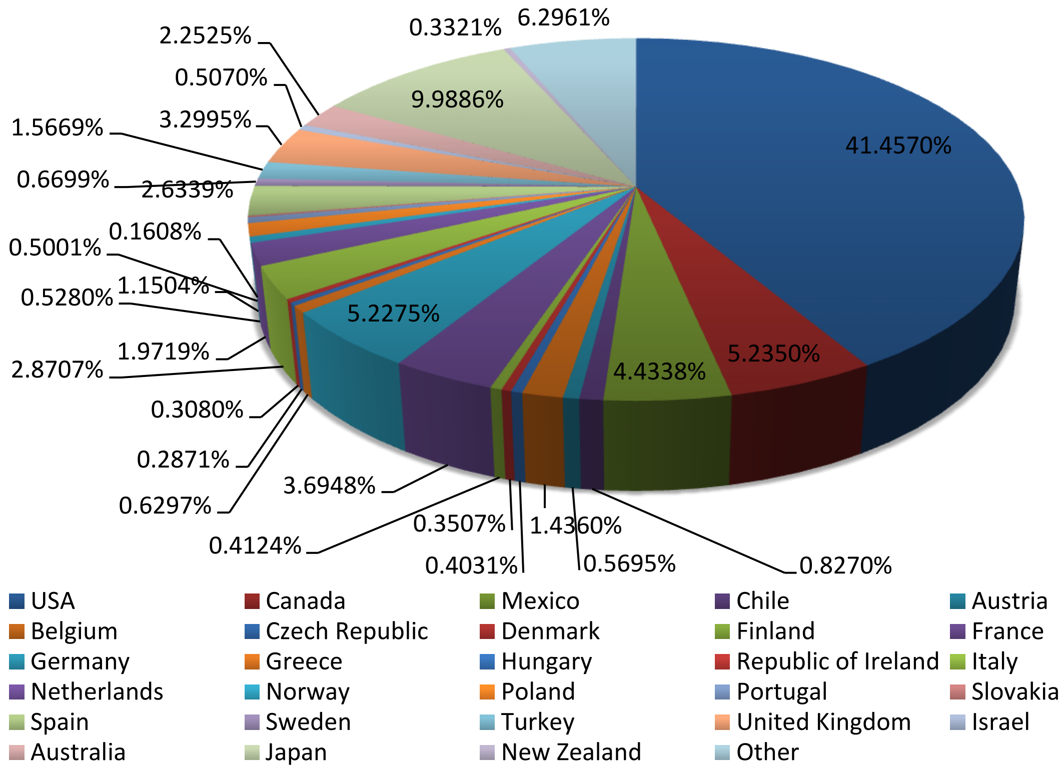


Figure 32 - OECD crude oil consumption in 2013 (data taken from BP, 2014).

The reported refinery capacity data signify the sum of reported atmospheric crude distillation and condensate splitting capacity. The capacity term refers to the amount of raw materials that a distillation facility can process under usual operating conditions, taking into account scheduled downtimes. As for production and consumption, BRIC countries increased their refining capacity respect to OECD nations (Figure 33 and Figure 34), which saw since 1965 a reduction of US capacity for the benefit of Mexico, South Korea, and Japan.

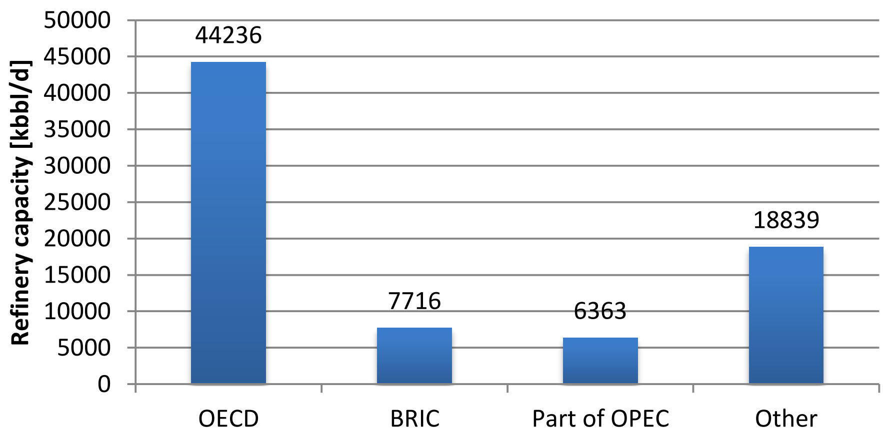


Figure 33 - Global refinery capacity in 1998 (data taken from BP, 2014).

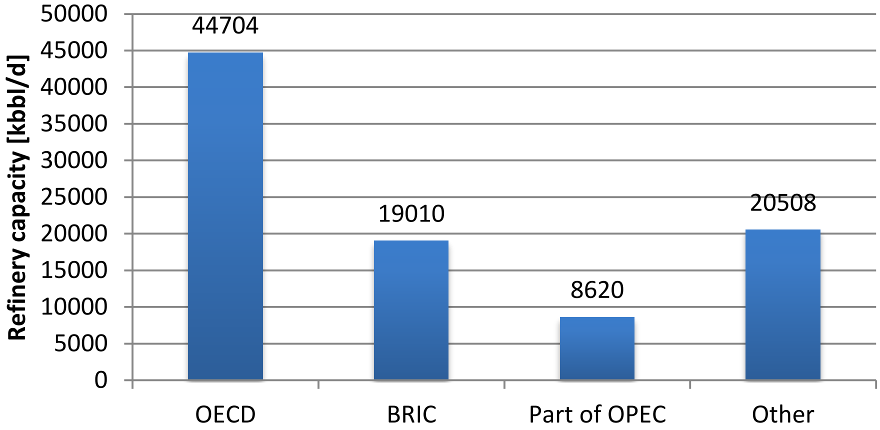


Figure 34 - Global refinery capacity in 2013 (data taken from BP, 2014).

Since 1998, the distribution of OECD refinery capacity has not seen significant variations (Figure 35 and Figure 36).

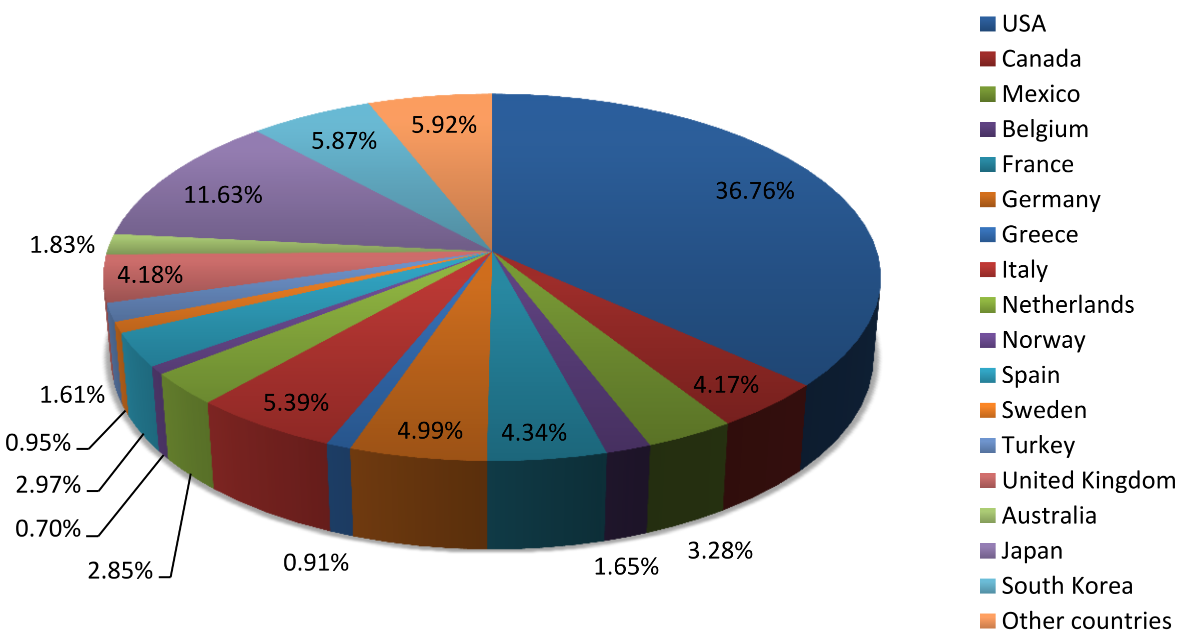


Figure 35 - OECD refinery capacity in 1998 (data taken from BP, 2014).

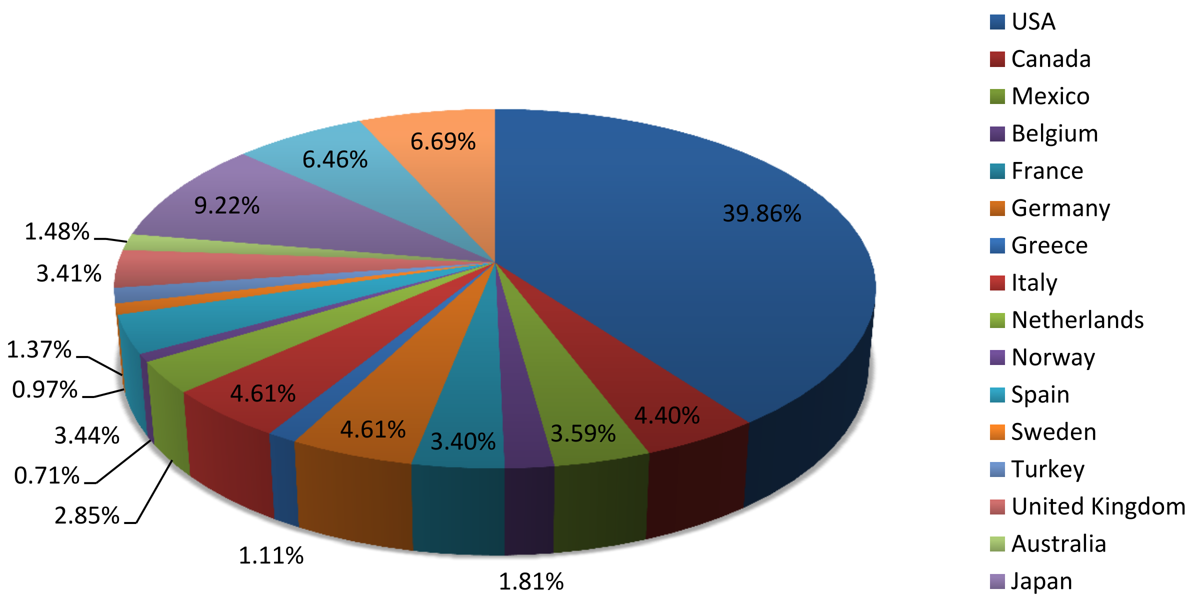


Figure 36 - OECD refinery capacity in 2013 (data taken from BP, 2014).

Chapter 2 Literature overview

As the scientific literature has shown a significant attention to modeling the quotation of CO for forecasting purposes, this Chapter proposes a brief review of the main papers that contributed to this research field. It is worth underlining that most of the published manuscript have a financial background and a time horizon that is well below the year span and shorter than the time horizon involved in assessing the dynamics of OPEX terms.

In order to classify the information found in dedicated literature papers, this thesis proposes three separate classifications for CO price models, according to the time-horizon of the forecast, the time granularity of data, and the mechanism used to generate the simulated prices and explain their trend.

2.1 An insight into market, price, and volatility of crude oil

A number of scientific papers have investigated the dynamics of oil prices and their volatilities over different time horizons, by means of different econometric and economic models described in detail in Chapters 3 and 4.

Recent studies of CO market cover a number of different areas and issues and examine the characteristic of these markets according to various aspects. As far as the most recent time period is concerned, several articles focus on the oil price dynamics in the 2002-2014 period that was characterized by high volatility, high intensity jump, strong upward drift and was concurrent with underlying fundamentals of oil markets and world economy. Most of the literature manuscripts are based on a financial background that focuses on the forecasting capability to trade the CO by means of futures, selling/buying options, and other financial operations that have a time horizon which is much shorter than the one typically used in PSE/CAPE applications (Manca, 2013). For instance, the modeling activity tries often to determine possible interconnections between price volatility and stock price fluctuations. It distinguishes between futures and spot prices: the former is a contract between two parties to buy or sell an asset for a price agreed upon today with delivery and payment occurring at the delivery date; the latter specifies the settlement price of a contract of buying or selling a commodity or currency, which is normally two business days after the trade date. Chen (2014) forecasted CO price movements by means of oil-sensitive stock indices with one-month ahead predictions. The novelty of its study consists in suggesting a new and valuable predictor (*i.e.* AMEX oil index) that both reflects timely market information and is readily available, since it is a price-weighted index of the leading companies involved in the exploration, production, and development of petroleum. Other studies focused on the growing presence of other financial operators in the oil markets and led to the diffusion of trading techniques based on extrapolative expectations (Kaufmann and Ulmann, 2009; Cifarelli and Paladino, 2010). Zagaglia (2010) studied the dynamics of oil futures in NYMEX by using a large dataset that included global macroeconomic indicators, financial market indices, quantities and prices of energy products, and carried out a Factor-Augmented Autoregressive model for oil futures prices. In addition, Zagaglia found that a factor correlated to purely financial developments contributed to the model performance in conjunction with elements related to energy amounts and prices.

Another relevant number of papers suggest that oil prices are mainly driven by supply-and-demand considerations. For instance, Kilian (2008) decomposed the real price of CO into supply shocks, shocks to the global demand for industrial commodities, and demand shocks. Similarly, Dvir and Rogoff (2014) presented evidence of connections among four variables in the US and global oil markets: oil production, stocks of CO, the real price of oil, and measures of incomes.

Kaufmann *et al.* (2004) applied statistical models to estimate the causal relationship between CO prices and several factors, such as capacity utilization, production quotas, and production levels. Kaufmann *et al.* (2008) investigated the factor that might have contributed to CO price increase in the first quarter of 2008, by expanding a model for CO prices to include refinery utilization rates, a non-linear effect of OPEC capacity utilization and conditions in futures markets as explanatory variables and found that their model performed relatively well when used for forecasting purposes. The contribution of Ye *et al.* (2009) and Chevallier (2014) is postponed to Chapter 4, where a description of such models is extensively discussed to analyze and identify fundamental physical and economic factors.

Other studies tribute a great importance to oil price volatility, because of a number of correlated reasons. Indeed, fluctuations of oil prices have often had great impacts on the economy. For example, most of the US post World War II recession was preceded by sharp increases in CO prices (Hamilton, 1983). Several papers used the standard deviation of price differences as a measure of volatility of commodity prices (Kang and Yoon, 2013; Salisu and Fasanya, 2013). These papers considered different plausible models for measuring volatility in the oil price and consequently compared the performance of such models to Brent and WTI oil markets. The fluctuations cannot be modeled by a simple average trend-line (Manca, 2013). Actually, high volatility, high intensity jumps, and bull-and-bear periods make difficult for investors to track the CO price trend, and for process designers/managers to perform a sound economic assessment in feasibility studies. Kang and Yoon (2013) dealt with more sophisticated econometric techniques that are widely used today. The general concept that was proven to work better over high-frequency time series in financial markets is the generalized autoregressive conditional heteroskedastic model (GARCH) and its modifications (*e.g.*, IGARCH, TGARCH, EGARCH).

The scientific literature questioned also on the role of CO to forecast the price/cost of further raw materials (*e.g.*, Manca, 2013; Manca and Rasello, 2014). Mazzetto *et al.* (2013) used CO as the reference component for econometric models in bioprocesses and showed a functional dependency of both raw biomaterials and final bio-products from the CO market. Mazzetto and coauthors evaluated also Autoregressive Distributed Lag, Fully Stochastic, and Time Series Decomposition models in order to remove the limiting assumption of keeping constant, for long-term horizons, the price/cost of commodities and utilities (as it happens in conventional conceptual design techniques, Douglas, 1988).

Along with other issues, the relationship between supply-and-demand variables, and the availability of models to forecast the input variables are of great interest. Indeed, another area of research that is noteworthy, even if it is only indirectly linked to CO price modeling, is the trend of CO production. According to Hubbert (1956), CO production follows the so called “Peak Oil Theory”, because it adheres to the production of most finite resources in a market economy (*i.e.* that resource depletes faster than it can be replaced), where the observed quantity initially grows in production, then reaches a maximum peak and eventually declines gradually to zero (Bardi, 2009). This phenomenon derives from the intrinsic features of the oil resource. Actually, the extraction of an abundant and cheap resource leads first to an economic growth and to increasing investments in further extraction. Gradually, the cheap resource gets depleted and extraction costs increase due to the need of extracting lower quality deposits. As a consequence, investments cannot keep pace with these rising costs. The key factors that generate Hubbert's curve are the positive feedback that derives from the reinvestment of the profits generated by the resources and the negative feedback that derives from the gradual depletion of the low cost resources (Bardi and Lavacchi, 2009). What finally generates the typical bell curve (Figure 37) for an energy resource is not the monetary cost but an energy cost, known as EROI (Energy Return on Energy Investment), which quantifies the variation with time of the net energy of extraction (Bardi and Lavacchi, 2009; Miller and Sorrell, 2013). A higher exploration rate leads to a higher and earlier production peak, whilst a lower exploration rate has the opposite effect.

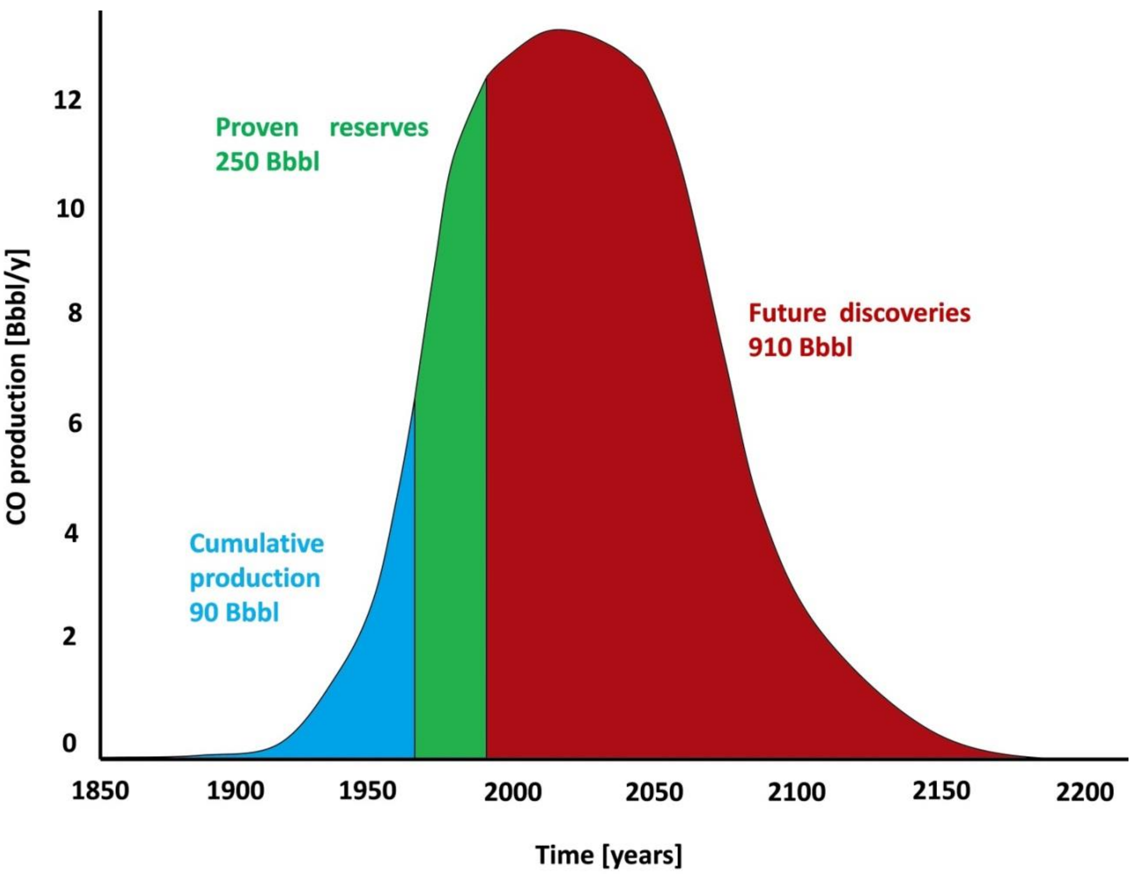


Figure 37 - Hubbert global peak oil plot (adapted from Hubbert, 1962).

At last, Dees *et al.* (2007) proposed not only a rule based on changes in market conditions, but also two models for oil demand and oil supply. Dees and coworkers were not clear on how to forecast the inventories and production capacity. For this reason, the model proposed in Chapter 5 tries to overcome this lack of detail and clarity in the scientific literature.

2.2 Crude oil models classification

In order to arrange the notions that have been presented in the previous paragraph, we now propose and discuss three classifications of CO price models (see also Table 2 for further details).

Table 2 - Summary of the mentioned models according to different classifications.

Reference	***Time-horizon***	***Time-granularity***	***Type of model***
Chen (2014)	Short	Monthly	Economic
Chevallier (2014)	Long	Weekly	Economic
Cifarelli and Paladino (2010)	Long	Weekly	Econometric
Dees *et al.* (2007)	Long	Quarterly	Economic
Dvir and Rogoff (2014)	Medium	Monthly	Econometric
Ghaffari and Zare (2009)	Short	Daily	Econometric
Kang and Yoon (2013)	Medium	Daily	Econometric
Kauffmann *et al.* (2004)	Medium	Quarterly	Economic
Kaufmann and Ullman (2009)	Short	Daily	Econometric
Kilian (2008)	Medium	Monthly	Economic
Manca and Rasello (2013)	Long	Monthly	Econometric
Salisu and Fasanya (2013)	Medium	Daily	Econometric
Ye *et al.* (2009)	Short	Monthly	Economic
Zagaglia (2010)	Medium	Monthly	Econometric

The first classification criterion is the length of the time horizon that can be used reliably to forecast the CO prices. Depending on the specific problem, the time horizon taken for the economic assessment can cover a short- (*e.g.*, Ghaffari and Zare, 2009; Ye *et al.*, 2009), medium- (*e.g.*, Chevallier, 2014), and long-term period (*e.g.*, Manca and Rasello, 2014) (*i.e.* from hours/days to months/years). On one hand, short-term horizon models, usually related to financial activities, cover periods of few days/weeks, and deal with the quantification of variations driven by short-term imbalances between supply and demand, which are triggered by sudden and therefore unexpected political, economic, and financial uncertainties (Hamilton, 2005, Zhang *et al.*, 2007). On the other hand, medium- and long-term horizon problems are challenging due to the need for forecasting the different variables involved in the market background (*e.g.* the levels of supply, demand, production and capacity storage) for rather long time intervals (*i.e.* from few months to some years) as required by PSE/CAPE targets related to either Conceptual Design or planning of chemical processes/plants.

The second classification criterion is based on the time-granularity (*i.e.* time discretization) of the dataset of input variables (*i.e.* days, weeks, months, years). This concept is closely related to the extent of the forecasting horizon, according to the number of steps that the model is able to forecast consistently. For instance, Ghaffari and Zare (2009) chose a one-day horizon for their forecasts in order to predict whether the CO price is just likely to rise or decline (they are indeed interested to model the sign of price variations for sell-and-buy purposes on quite short time horizons), while Ye *et al.* (2009) proposed a forecasting model based on one-step-ahead monthly predictions.

The third classification criterion is based on the intrinsic nature of the forecasting models that, according to Aprea *et al.* (2005), can be divided into economic and econometric models. These two categories are conceptually different for both the mechanism used to generate future prices and the type of explanation provided about their trend. The first category involves purely economic variables and simulates the fluctuations in spot and futures market with the supply-and-demand law. The second category does not take into account the forces that cause price fluctuations, but is only focused on the trend followed by the prices and on possible future evolutions. The reliability of economic models over rather long-term horizons presents some problematic issues, which at the time appear insurmountable, such as the need for long-term forecasts of the supply levels, and demand and capacity storage, together with the inability to predict the variables that affect the price trend over long periods (*e.g.*, technological developments, political dynamics, changes in collective behaviors, and both national and international backgrounds). Chapter 3 and Chapter 4 present and discuss the main features of econometric and economic models and provide some model examples.

Chapter 3 Econometric models

This Chapter investigates the dynamics of WTI and Brent prices and their volatilities, by applying different econometric and statistical tools. After a brief introduction to time series analysis, we discuss the main models used to shape the financial time series, with their main properties and possible modifications. The identification and validation of econometric models calls for the acquisition of historical price data, so this Chapter shows the statistical data of both CO price series and CO shock series, with the shock series being normally distributed. Eventually, the Chapter presents the so called random walk model (i.e. based on a Markovian process) that can simulate the future quotations of the two benchmarks and provide a probabilistic distribution of prices, which defines the likely domain where prices can move. As Chapter 6 proposes another econometric model in moving average that has been developed in Barzaghi and Conte (2015), the last Section of this Chapter shows the limitations of Markovian model, too.

3.1 Introduction to time series analysis

A time series is a collection of observations made sequentially through time. Several examples occur in a variety of fields, ranging from economics (*e.g.*, CO quotations, commodity prices, utility costs) to engineering (*e.g.*, failure temperatures, yield of a batch process, rate of equipment rupture). Methods for the time series analysis constitute an important area of statistics (Bezruchko and Smirnov, 2010; Chatfield, 2000; Chatfield, 2003; Commandeur and Koopman, 2007; Massarotto, 2005). The study of time series allows to both interpret a phenomenon, by identifying possible trends, seasonal variations, other periodic variations, and stochasticity, and predict its future performance. If future values can be predicted exactly from past values, then a series is said to be deterministic. However, most series are stochastic, or random, as the future can partially be determined by past values. Time series modeling and forecasting include several applications, such as production and capacity planning, sales forecasting, inventory or stock control, and the evaluation of alternative economic strategies to run a plant (Chatfield, 2004). The following Sections show the theoretical procedures of time series analysis that are typical of statistical mathematics.

Traditional methods of time series analysis are mainly concerned with decomposing the variation of a series into components representing trend, seasonal variation, and other incidental cyclic changes. Any remaining variations are then attributed to irregular fluctuations. This approach is not always the optimal one but is particularly valuable when the variation is dominated by trend and seasonality (Chatfield, 2003). The trend is the long-term change in the variable mean level, while seasonality consists of one or more periodic components that are repeated with regularity, *e.g.* with annual, seasonal, monthly, weekly, or daily period. For instance, gasoline demand is typically higher in summer and lower in winter. This yearly variation is easy to understand, and can readily be estimated by plotting time series, whenever seasonality is of direct interest. In addition, economic time series, such as company revenues, exhibit oscillations, which are predictable to some extent with a period varying from 3-4 years to more than 10 years (Sterman, 2000; Chatfield, 2003), depending on the measured variable and on the economic growth in that period (*i.e.* the Economy usually behaves differently when it goes into recession rather than when it emerges again).

The stochastic variability is the difference between the true value and trend with seasonality: it can be treated as a stationary stochastic process, *i.e.* a random series with mean zero and a suitable distribution. Real-world economic processes (*e.g.* CO prices) look non-stationary because their characteristics are not constant over an observation interval. In the theory of random processes, the non-stationarity of a process implies temporal changes in its multidimensional distribution (Bezruchko and Smirnov, 2010). If one does not pretend to get a precise and unique forecast of future states, then a probabilistic approach is traditionally used. This is the feature of the stochastic approach to time series that is used in this Chapter.

3.1.1 Modeling financial time series by means of econometric models

The essential point in the analysis of financial time series is the need to deploy a model capable of taking into account previous observations to extract significant features of the data. Based on such time series, the forecasting models allow investors and process designers/managers to forecast possible future trends based on known past observations. In case of financial/economic time series, three steps should be undertaken to deploy the model (Chatfield, 2000). The first one consists in the model *identification*, whose parameters are estimated at a later stage by either maximum likelihood or non-linear least squared methods (*i.e.* model *estimation*). Eventually, the model should conform to the specifications of a stationary univariate process (*i.e.* a process with mean and variance constant over time) and the residuals should be independent and constant in mean and variance over time (*i.e.* model *checking*). The following Sections show a list of econometric models and their main characteristics that are wide used in the forecast of price time series, which for the sake of simplicity is called in the general form $\{Y_t\}$.

3.1.1.1 ARMA model

Mixed Autoregressive Moving Average (ARMA) models are a general class of models that allow examining the dynamics of individual time series. A general ARMA model consists of two parts: an autoregressive part and a moving average one. The qth-order Moving Average (MA) process is defined as:

$$Y_t = \mu + \varepsilon_t + \theta_1 \cdot \varepsilon_{t-1} + \theta_2 \cdot \epsilon_{t-2} + \cdots + \theta_q \cdot \varepsilon_{t-q} \tag{1}$$

while the pth-order Autoregressive (AR) process can be written as:

$$Y_t = \varphi_0 + \varphi_1 \cdot Y_{t-1} + \varphi_2 \cdot Y_{t-2} + \cdots + \varphi_p \cdot Y_{t-p} \tag{2}$$

Hence, ARMA processes are defined as follows:

$$Y_t = \varphi_0 + \varphi_1 \cdot Y_{t-1} + \varphi_2 \cdot Y_{t-2} + \cdots + \varphi_p \cdot Y_{t-p} - \varepsilon_t - \theta_1 \cdot \varepsilon_{t-1} - \theta_2 \cdot \epsilon_{t-2} - \cdots - \theta_q \cdot \varepsilon_{t-q} \tag{3}$$

where φ_0, φ_1,…, φ_p, θ_1,…, θ_q are real coefficients, and ε_t is a white noise process, *i.e.* a sequence of elements, which are uncorrelated across time, with mean zero and variance σ^2. This representation implies that Y_t is modeled as a weighted average of past observations with the addition of a white noise error. The identification of optimal values for p and q in the ARMA(p,q) model is usually achieved by plotting the partial autocorrelation functions for an estimate of p and likewise by using the autocorrelation function for an estimate of q (see Section 3.1.2.4). Further information can be obtained by considering these same functions for the residuals of a model fitted with an initial selection of p and q. The usual way to estimate parameters in ARMA models, after choosing p and q, is based on the least-squares-regression method where the values of the adaptive parameters minimize the error term. The main advantages of this type of model lie on its mathematical tractability to approximate general stationary processes, and the relatively simple procedure to compute the parameters. However, as it was shown in the literature (Chatfield, 2000; Chatfield, 2003), ARMA models are not suitable to evaluate the entire distribution of nonlinear processes because they are linear, very sensitive to outliers, and, moreover, they do not take into consideration conditional heteroskedasticity, since the variance is constant over time.

3.1.1.2 ARCH model

In financial time series there is also the interest to forecast not only the *level* of the series, but also its variance. In order to clarify the terms that are used in this Chapter, the distinction between conditional and unconditional variance is presented.

In probability theory and statistics, a conditional variance is the variance of a conditional probability distribution, *i.e.* the variance of a random variable given values of one or more other variables, while the unconditional variance is just the standard measure of the variance.

While the AR models imply the unconditional variance being constant, changes in the variance are very important to understand financial markets and it is worth considering that conditional variance may demonstrate a different behavior and significant changes over time. The main idea behind the ARCH models, which were proposed by Engle (1982), is the following: the forecast based on the past information is presented as a conditional expectation depending upon the values of past observations. Therefore, the variance of such a forecast depends on past information and may be a random variable. Accordingly, the ARCH(p) model has the following form:

$$Y_t = \beta_0 + \sum_{i=1}^{p}(\beta_i \cdot Y_{t-i}) + \varepsilon_t \tag{4}$$

$$\varepsilon_t = \nu_t \sqrt{h_t} \tag{5}$$

$$h_t = \alpha_0 + \sum_{i=1}^{p} (\alpha_i \cdot \varepsilon_{t-i}^2) \tag{6}$$

where ν_t is a random variable distribution or white noise with mean zero and variance 1, independent of the past error term ε_t, and β_0, β_i, α_0, and α_i are model coefficients evaluated with a maximum likelihood method as, according to Engle (1982), an ordinary least squares estimation does not produce as good results as the maximum likelihood one. The main advantage of ARCH model is that it takes into account that conditional variance is substantially affected by the squared residual term in any of the previous periods. It can be seen from the model that large past squared shocks ε_t^2 imply a large conditional variance h_t. Therefore, under the ARCH framework, large shocks tend to be followed by further large shocks. This feature is similar to the volatility clustering phenomena observed in CO quotation series. However, the model assumes that positive and negative shocks have the same effects on volatility, because the volatility depends on the square of the previous shocks. In practice, price of a financial asset such as CO responds differently to positive and negative shocks, as it is partly explained in Chapter 6 in order to present the new hybrid model and in Section 3.2 of this Chapter (Figure 41).

Moreover, the ARCH model provides only a mechanical way to describe the behavior of conditional variance and, like other econometric models, gives no indication about what causes such trend to occur (Abledu *et al.*, 2013).

3.1.1.3 GARCH model

The GARCH model, which was elaborated by Bollerslev (1986) in order to extend Engle’s framework by developing a technique that allows the conditional variance to be an ARMA process, has the following form:

$$Y_t = \alpha_0 + \sum_{i=1}^{p} (\alpha_i \cdot Y_{t-i}) + \varepsilon_t \tag{7}$$

$$\varepsilon_t = \nu_t \sqrt{h_t} \tag{8}$$

$$h_t = \alpha_0 + \sum_{i=1}^{p} (\alpha_i \cdot \varepsilon_{t-i}^2) + \sum_{j=1}^{q} (\beta_j \cdot h_{t-i}) \tag{9}$$

where ν_t is a random variable with mean zero and variance 1, h_t is the conditional variance, ε_t is the error term, and α_0, α_i, and β_j are adaptive parameters. GARCH models encounter the same weaknesses as ARCH ones, because they respond without distinction to positive and negative shocks.

3.1.1.4 EGARCH model

In order to overcome the cons of GARCH model, in particular to model asymmetric effects for positive and negative returns, Nelson (1991) proposed the Exponential GARCH (EGARCH) model. Conditional variance is described according to the following formulation:

$$\varepsilon_t = \nu_t \sqrt{h_t} \tag{10}$$

$$log(h_t) = \alpha_0 + \sum_{j=1}^{q} \left[\beta_j \cdot log(h_{t-i})\right] + \sum_{i=1}^{p} \left[\theta_k \cdot \left(\frac{\epsilon_{t-k}}{\sqrt{h_{t-k}}}\right) + \gamma_k \cdot \left(\frac{\epsilon_{t-k}}{\sqrt{h_{t-k}}}\right) - \sqrt{\frac{2}{\pi}}\right] \tag{11}$$

where logged conditional variance $log(h_{t-i})$ is used in order to relax the positive constraint of model coefficients.

3.1.1.5 GJR-GARCH model

Glosten, Jagannathan and Rukle (1993) modeled GARCH processes as follows:

$$\varepsilon_t = \nu_t \sqrt{h_t} \tag{12}$$

$$h_t = \alpha_0 + \sum_{i=1}^{p} (\alpha_i \cdot \varepsilon_{t-i}^2) + \sum_{j=1}^{q} \left(\beta_j \cdot h_{t-i}\right) + \sum_{i=1}^{p} \left(\gamma_i \cdot \varepsilon_{t-i}^2 \cdot I_{\{\varepsilon_{t-i} \geq 0\}}\right) \tag{13}$$

where $I_{\{\varepsilon_{t-i} \geq 0\}}$ assumes value 1 if the shock is non-negative, value zero elsewhere. The non-negativity condition models asymmetric consequences of positive and negative returns.

3.1.1.6 APARCH model

The Asymmetric Power ARCH (APARCH) model of Ding, Granger, and Engle (1993) is one the most general GARCH models. Indeed, the specific feature of APARCH model is the capability to estimate a power coefficient δ that was assumed to be equal to 2 in all previous models thus providing a possible higher flexibility in terms of identification potential. The general form of APARCH model is as follows:

$$\varepsilon_t = \nu_t \sqrt{h_t} \tag{14}$$

$$h_t^{\delta} = \alpha_0 + \sum_{i=1}^{p} \alpha_i (\epsilon_{t-i} - \gamma_i \cdot \varepsilon_{t-i})^{\delta} + \sum_{j=1}^{q} \left(\beta_j \cdot h_{t-i}^{\delta}\right) \tag{15}$$

3.1.2 General methodology and statistical tools

In economics, different measures of economic activity are typically recorded at regular intervals. This thesis concentrates on the discrete time series of CO quotations, which is a series measured at discrete time intervals. Some discrete variables, with large means, can be treated as if they were approximately normally distributed and modeled according to a continuous distribution. Hence, this Section shows the statistical tools that are used in the distribution analysis of CO prices.

3.1.2.1 Time plot

The first step in any time series analysis or forecasting exercise is to plot the observations against time, to obtain what is called a time plot of the data. The graph should show important data features such as trend, seasonality, outliers (*i.e.* an observation point that is significantly distant from other observations and probably affected by a gross-error), smooth changes in structure, turning points, and sudden discontinuities. Another well-known kind of graph is the scatter plot that allows exploring the relationship between variables. The time plot of a single variable can be regarded as a form of scatter plot with time being treated as the second variable.

3.1.2.2 Measures of skewness and kurtosis

As suggested in Mardia (1970), the measures of skewness and kurtosis have proved useful in developing a test of normality and in investigating the robustness of the standard normal theory procedure. Skewness is a measure of symmetry, or more precisely, the lack of symmetry. For univariate data Y_1, Y_2,..., Y_N, the formula for skewness is:

$$S = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})^3}{(N-1) \cdot s^3} \tag{16}$$

where $\bar{Y}$ is the mean, s is the standard deviation, and N is the number of data points. The skewness of a normal distribution is zero, and any symmetric data should have a skewness near zero. Negative values of skewness indicate data that are skewed left and positive values indicate data that are skewed right. By "skewed left", we mean that the left tail is longer than the right tail. Similarly, "skewed right" means that the right tail is longer than the left one. Some measurements (*e.g.*, failure times in reliability studies) have a lower bound and are skewed right, because they physically cannot be negative.

Kurtosis is a measure of whether data are peaked or flat relative to a normal distribution, *i.e.* data sets with high kurtosis tend to have a distinct peak near the mean, while data sets with low kurtosis tend to have a flat top near the mean. For univariate data Y_1, Y_2,…, Y_N, the formula for kurtosis is:

$$K = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^4}{(N-1)\cdot s^4} \quad (17)$$

where $\bar{Y}$ is the mean, s is the standard deviation, and N is the number of data points. Since the kurtosis of a normal distribution is three, some sources (DeCarlo, 1997) use the following definition of excess kurtosis:

$$K = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^4}{(N-1)\cdot s^4} - 3 \quad (18)$$

According to the second definition, a positive kurtosis indicates a peaked distribution (*i.e.* leptokurtic) and negative kurtosis indicates a flat distribution (*i.e.* platykurtic). For instance Dirac's delta function has an infinite kurtosis. Figure 38, Figure 39, and Figure 40 show histograms for 10000 random numbers that are generated from a LogNormal, a Normal and a Weibull distribution.

Table 3 - Skewness and kurtosis of the sample distributions.

	Normal	Weibull	Lognormal
skewness	0.0077	1.1679	4.5034
kurtosis	2.9796	4.684	41.6333

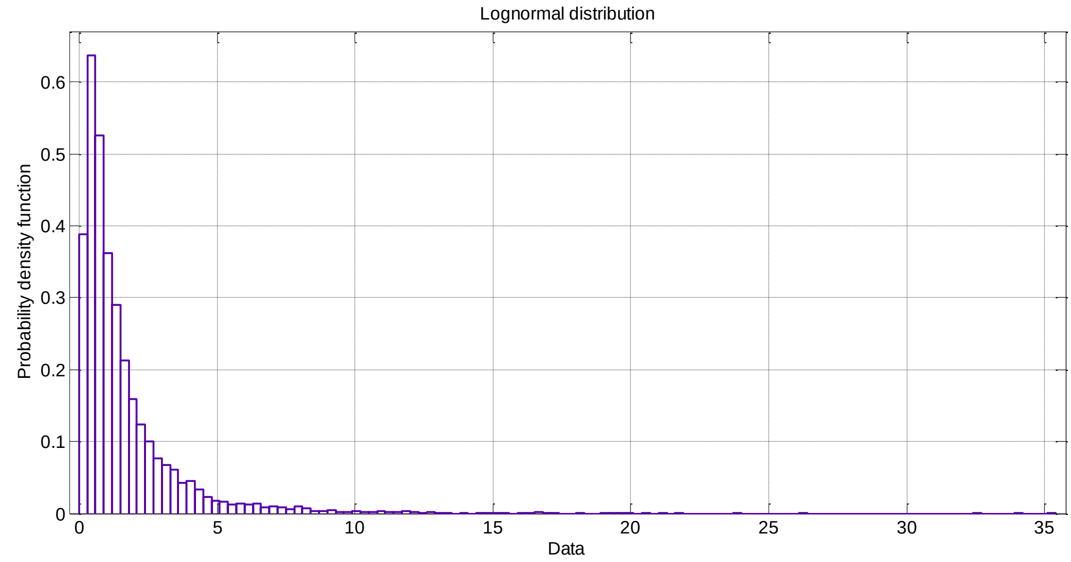


Figure 38 - Histogram for 10000 lognormal random numbers.

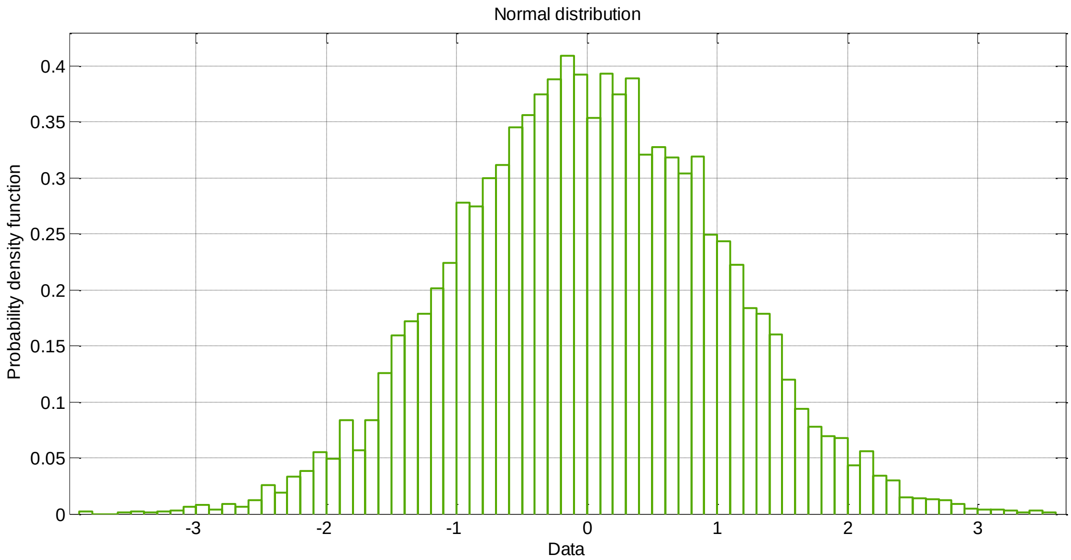


Figure 39 - Histogram for 10000 normal random numbers.

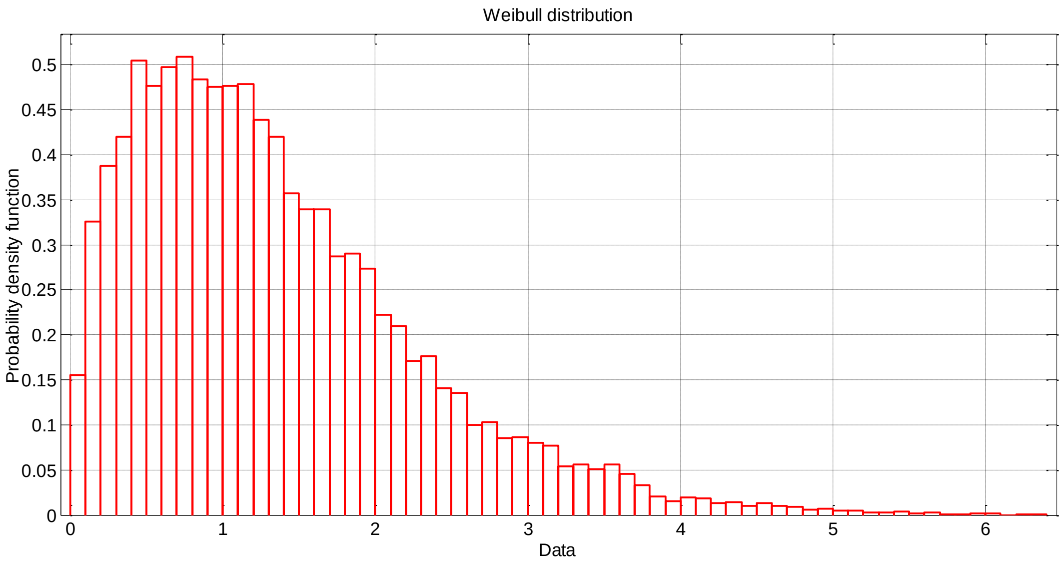


Figure 40 - Histogram for 10000 Weibull random numbers.

The first histogram is the sample from a LogNormal distribution that shows heavy tails and a single peak that is unbalanced to the left of the distribution. Indeed, the skewness is 4.5034 and the kurtosis is 41.633 (Table 3).

The second histogram is a sample from a normal distribution, which is a symmetric distribution with *well-behaved* tails. Indeed the histogram verifies the symmetry: the skewness is 0.0077, and the kurtosis is 2.9796, *i.e.* near to the theoretical value of 3.

The third histogram is a sample from a Weibull distribution with the shape parameter equal to 1.5. The Weibull distribution is a skewed distribution with the amount of skewness and the degree of decay depending on the value of the shape parameter. For this data set, the skewness is 1.1679 and the kurtosis is 4.684, which indicates moderate skewness and kurtosis.

3.1.2.3 Jarque-Bera test

There are several tests that can be used for the validation of normality hypothesis, as a function of the sample size (Gilbert, 1987):

- Shapiro and Wilk test ("W test"): can evaluate the hypothesis of the existence of normal or lognormal distribution, if the number of available data is lower than 50;

- D'Agostino test: can assess the same hypothesis and is used if the number of data is equal or greater than 50;
- Normal Quantile-Quantile (QQ) plot: it is a graphical test whose reliability is rather poor if not associated with any other more comprehensive test;
- Lilliefors test: is used in case of large data set, usually bigger than 1000;

Jarque-Bera test is used to test whether a given distribution is normal and is defined as follows:

$$JB = \frac{N}{6}\left(S^2 + \frac{1}{4}K^2\right) \tag{19}$$

where N is the sample size, S is the skewness, and K is the kurtosis. Null hypothesis (*i.e.* data are normally distributed) is rejected at $\alpha\%$ significance level if JB is larger than the α-quantile of the chi-square distribution with 2 degrees of freedom.

3.1.2.4 Correlation and autocorrelation function

A useful diagnostic tool to investigate the randomness of a set of observations is the correlogram, which is a graph that represents the correlation of two historical series, *i.e.* an observed time series and another time series shifted *k* time points into the past. Since *k* equals the distance in time between the observations, it is called a lag. The value pairs are represented in a Cartesian diagram with the delays on the abscissa axis and the corresponding correlation indexes on the ordinate axis. The correlation index, which measures in a non-dimensional way the interaction between two variables (Manca, 2013), is defined as follows:

$$corr(X,Y) = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} \tag{20}$$

where:

$$cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)] \tag{21}$$

$$var(X) = E[(X - \mu_X)^2] \tag{22}$$

Moreover, when the correlations are computed between a variable shifted in time and itself, the correlogram becomes an autocorrelogram.

3.2 Statistical analysis and results of price and volatility of crude oil

A statistical analysis of the data sets was performed with *Matlab*©, to better quantify the stochastic contribution in the model of CO quotations, and extrapolate the dynamic behavior of price series and of shock series. In general, relative shocks (also known in literature as *returns*) are calculated in the following way (for the *i*-th step):

$$R_t = \frac{P_i - P_{i-1}}{P_{i-1}} \tag{23}$$

where P_i and P_{i-1} are two subsequent quotations, and their difference is called shock (Figure 41 and Figure 42).

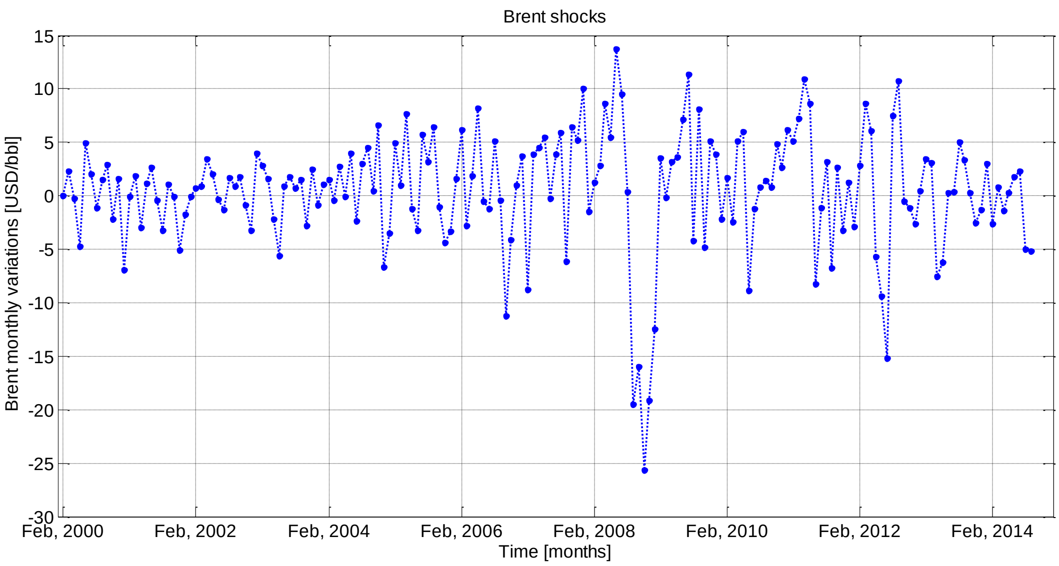


Figure 41 - Brent monthly shocks from 2000 to 2014.

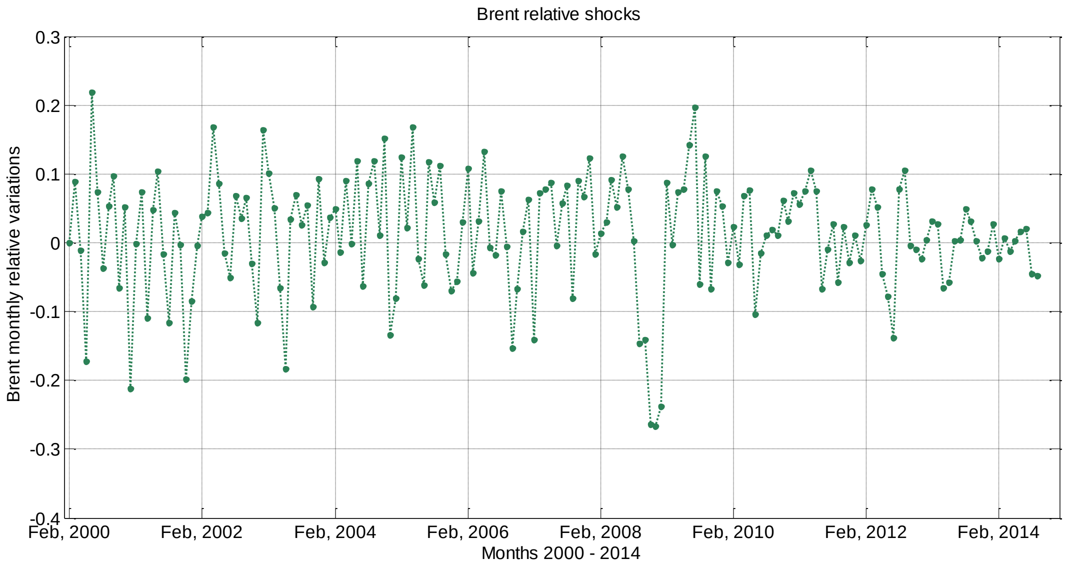


Figure 42 - Brent monthly relative shocks from 2000 to 2014.

As the distribution of prices quantifies the range within which future prices are likely to move, and the frequency or probability that these data belong to certain range intervals, Table 4 and Table 5 collect some statistical features of the time series of CO prices. These tables contain a set of values that is representative of the volatility of prices reported as the variation of quotations of CO from the previous to the following sampling time (*i.e.* shock and relative shock) from January, 2000 to August, 2014. The monthly data of Brent and WTI CO spot prices are used in our analysis and come from the US Energy Administration databank (www.eia.gov/forecasts/steo/query), because of their completeness.

Table 4 - Descriptive statistics results of crude oil prices [USD/bbl].

	2009-2014		2005-2014		2000- 2014	
	BRENT	WTI	BRENT	WTI	BRENT	WTI
Mean	96.0682	87.3774	86.1997	81.6763	66.7031	64.3951
Median	107.62	90.755	83.745	83.275	63.485	64.065
Mode	43.32	94.51	39.95	94.51	25.62	94.51
Min	43.32	39.09	39.95	39.09	18.71	19.39
Max	125.45	109.53	132.72	133.88	132.72	133.88
Standard deviation	21.1076	15.7918	24.579	20.0744	33.8915	29.3389
Skewness	-0.9125	-1.161	-0.0749	0.0422	0.1986	0.1257
Kurtosis	2.776	4.1511	1.6913	2.5265	1.6274	1.8112

Table 5 - Descriptive statistics results of crude oil shocks [USD/bbl].

	2009-2014		2005-2014		2000-2014	
	BRENT	WTI	BRENT	WTI	BRENT	WTI
Mean	0.8554	0.8068	0.4922	0.4284	0.4324	0.3936
Median	0.81	1.02	1.095	1.075	0.93	0.995
Mode	2.61	-12.35	-1.22	1.4	-2.78	1.4
Min	-15.18	-12.35	-25.65	-27.5	-25.65	-27.5
Max	11.31	14.18	13.73	14.18	13.73	14.18
Standard deviation	5.245	5.3651	6.5798	6.533	5.5891	5.5151
Skewness	-0.3628	-0.147	-1.1581	-1.1345	-1.2394	-1.2365
Kurtosis	3.2484	3.1093	5.2889	5.731	6.6776	7.419

If one compares the data of WTI quotations with gasoline prices (U.S. Average Gasoline Regular Grade Retail Price, taxes included, by EIA, 2015) it is possible to observe that the skewness of these series is different as well as the mean values and volatilities (Table 6).

Table 6 - Descriptive statistics results of monthly gasoline prices [USC/gal].

	2009-2014	2005-2014	2000-2014
Mean	320.3478	300.5923	250.6266
Median	338	301.8	255.5
Mode	361.1	280.3	139.7
Min	178.8	168.7	108.6
Max	390.6	406.2	406.2
Std deviation	53.1717	59.3855	85.8584
Skewness	-0.8304	-0.2426	0.0628
Kurtosis	2.7996	2.0114	1.6236

The autocorrelograms of price variations in Figure 43 and Figure 44 represent the correlation index of CO absolute shocks and CO relative shocks respect to the same series subject to progressively longer time shifts. They show the absence of periodic phenomena and confirm the influence of stochastic contributions. The horizontal lines identify the confidence bounds, which are equal to $\left[-z_{1-\alpha/2}/\sqrt{n};\ +z_{1-\alpha/2}/\sqrt{n}\,\right]$, where $1-\alpha$ is the confidence probability, and z_ζ is the ζ-th quantile of a normal distribution (Massarotto, 2005). The autocorrelograms do not show any significant trends of the correlation terms (*i.e.* monotonically increasing, decreasing, or periodic behavior). Therefore it is possible to draw the conclusion that the shocks of CO quotations are originated by rather stochastic phenomena. As reported by Salisu and Fasanya (2013), Brent is more volatile than WTI in terms of standard deviation of quotations. Conversely, the same volatility dominance of Brent over WTI (Salisu and Fasanya, 2013) cannot be observed when both absolute and relative oil price shocks are considered (see also Table 4 and Table 5 for quantitative details).

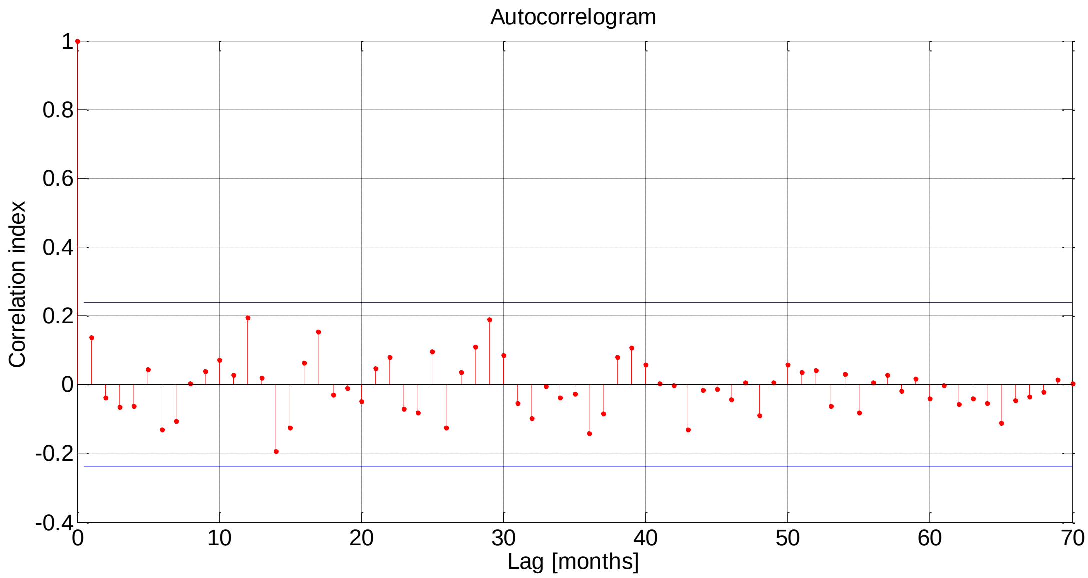


Figure 43 - Autocorrelogram of WTI shocks on monthly basis (data from Jan, 2009 to Aug, 2014).

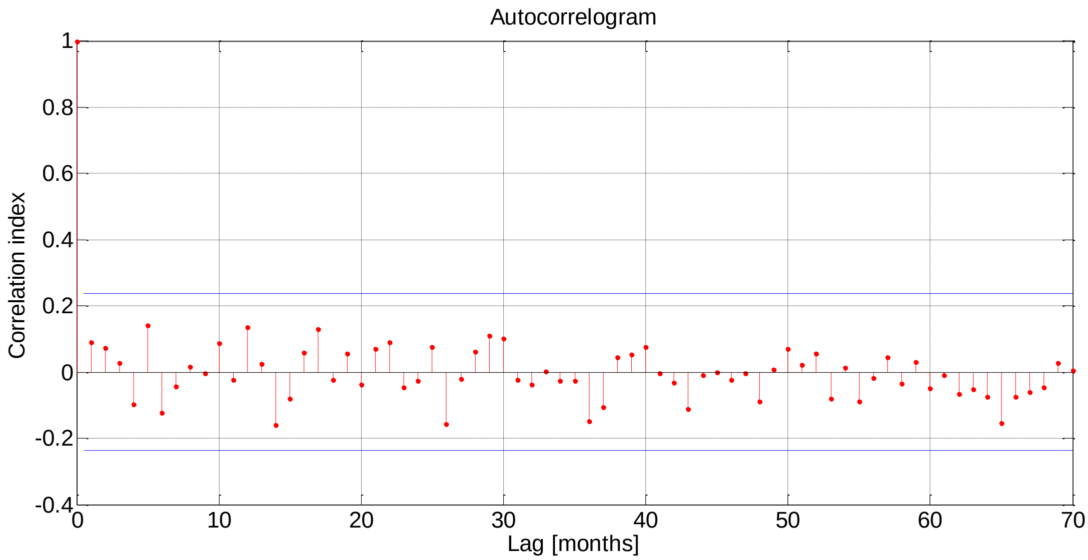


Figure 44 - Autocorrelogram of WTI relative shocks on monthly basis (data from Jan, 2009 to Aug, 2014).

As occurs frequently in financial markets, kurtosis of the shocks is greater than 3 in all cases, thus the density function is characterized by a fatter tail when compared to the standard Gaussian distribution. Conversely, the coefficient of skewness is either positive or negative but rather near zero in different cases for both Brent and WTI prices and returns. This point shows that there is an acceptable symmetry of the probability distribution of prices and shocks. According to Jarque-Bera's test, the CO price distribution does not result to be normally distributed, while the distributions of absolute shocks and relative shocks are normal at the 5% significance level. This experimental evidence is not in agreement with Kang and Yoon (2013) and Salisu and Fasanya (2013) results. Indeed, Salisu and Fasanya said that relative shock distribution is not normal for both CO benchmarks, and Kang and Yoon supported this hypothesis. In order to reinforce these results, Figure 45, Figure 46, Figure 47, and Figure 48, represent the histograms and cumulative density functions of shock and relative shock data, respectively. There is a rather good agreement between the shock distributions and the Gaussian curve with the same mean and standard deviation than the shock data. In addition to the improved results respect to Salisu and Fasanya (2013) and Kang and Yoon (2013), histogram and cumulative density function information is used to create the variation term in the random model that is described deeper in Section 3.3.

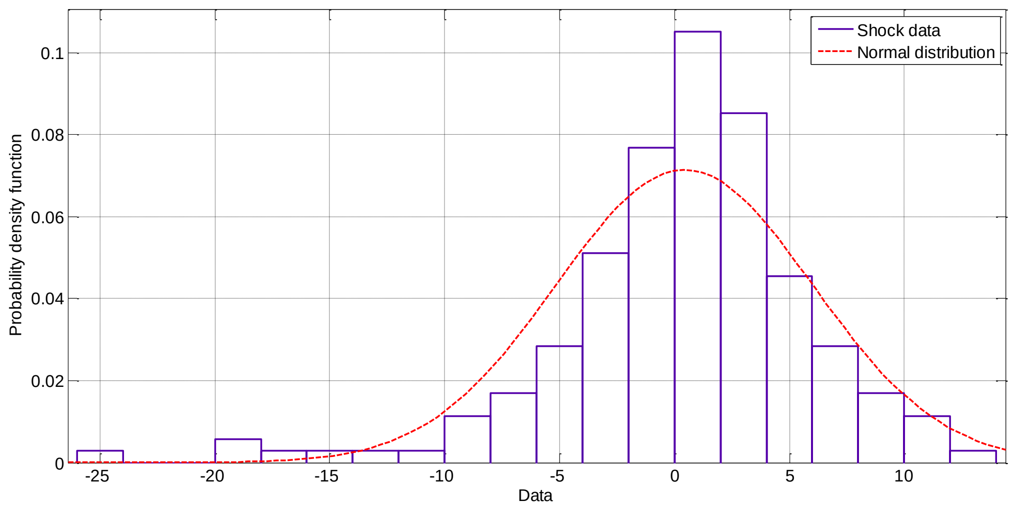


Figure 45 - Comparison between histogram of shock data and the normal distribution.

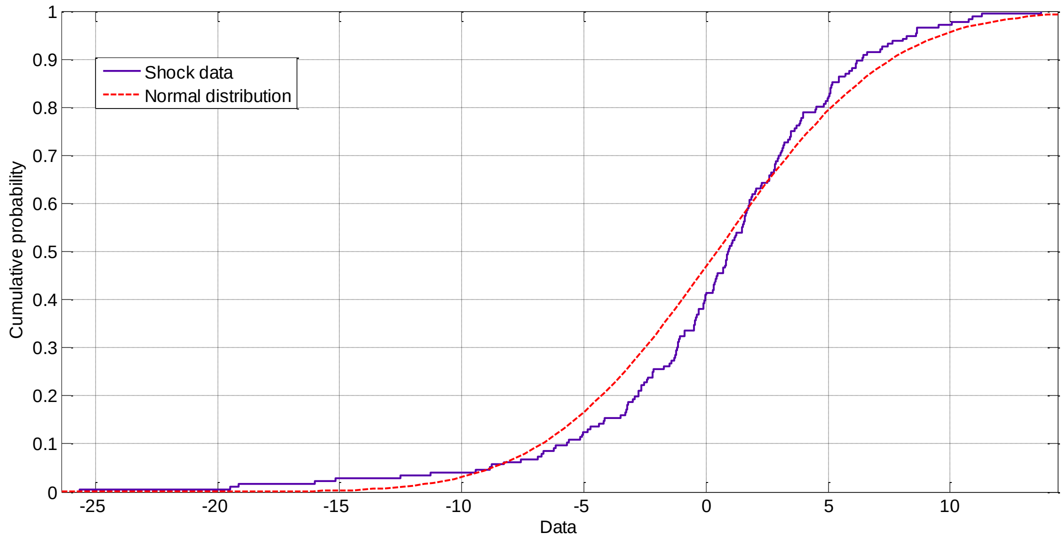


Figure 46 - Comparison between the cumulative frequency functions of shock data and normal distribution.

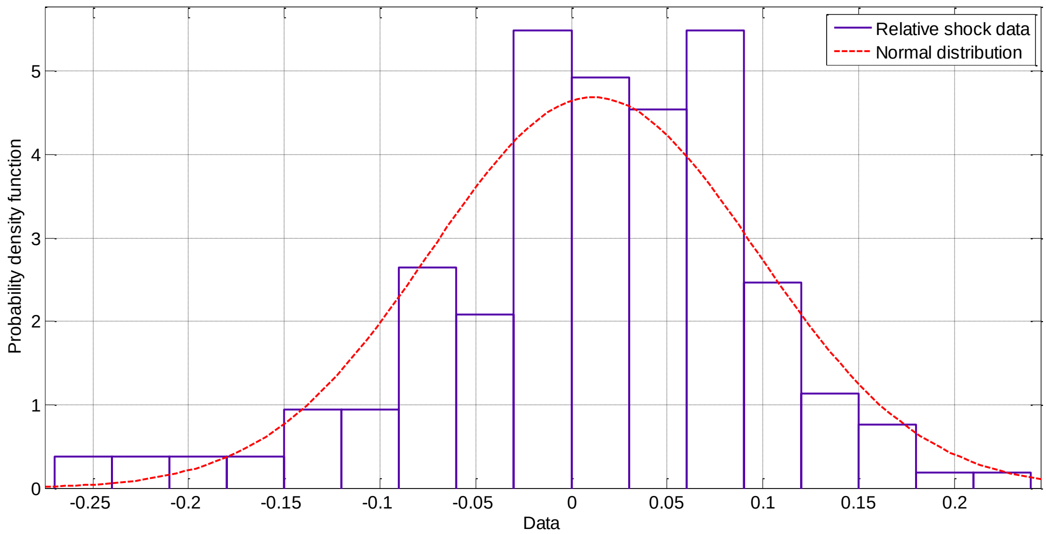


Figure 47 - Comparison between histogram of relative shock data and the normal distribution.

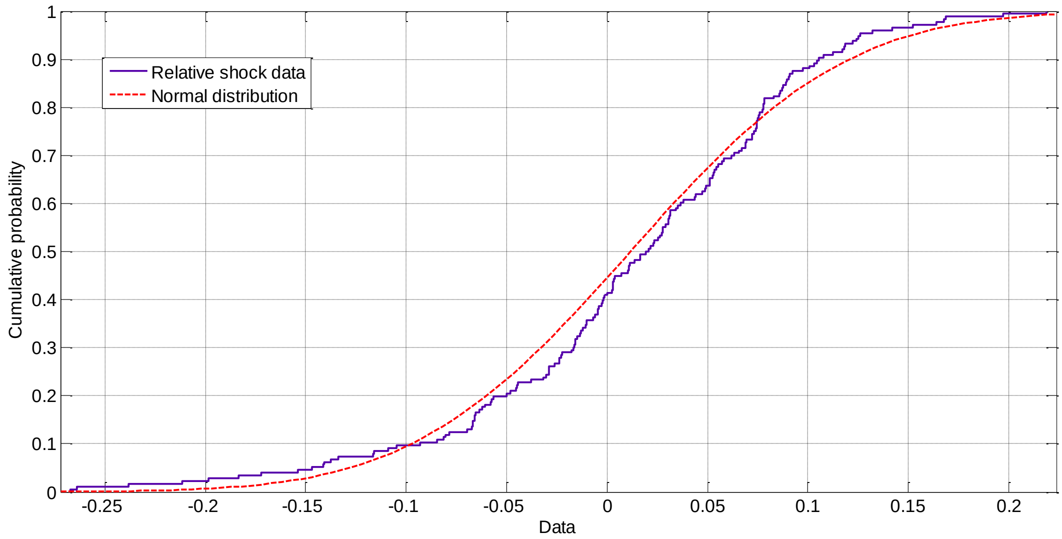


Figure 48 - Comparison between the cumulative frequency functions of relative shock data and normal distribution.

In particular, the normal distribution of shocks allows estimating the local drift of random walk (*i.e.* the short-term trend) by choosing a random number belonging to a normal distribution with mean and standard deviation equal to the mean and standard deviation of the shock distribution (Figure 45).

3.3 Random walk model

Morana (2001), Fong and See (2002), Alquist and Kilian (2010), and Alquist *et al.* (2011) are some examples of the literature interest in financial models (*e.g.*, Benchmark model, Hotelling model, other parsimonious econometric forecasts) that will not be deepened in the present work of thesis, because they have a financial background and a too short forecast horizon. The econometric model analyzed in this Section is called random walk model and can be thought of as the mathematical description of a Markov process, which is a stochastic and time-discrete process where the transition probability to a new state depends only on the previous state and not on the history that brought the system to that point (Mazzetto *et al.*, 2013). The new state is *memoryless* and derives stochastically from the previous one. Indeed, the simplest Markov process models the state of a system with a random variable that changes with time. Since the distribution of prices defines the range within which it is likely that prices will move, some statistical features of the time series of CO prices are already collected and discussed in Table 4 and Table 5. Since shocks are normally distributed (Figure 45) and the autocorrelograms do not show any significant trends of the correlation terms (Figure 43), the random walk model can be identified by the following equations:

$$P_{co,i} = P_{co,i-1} + \Delta P_{co} \quad (24)$$

$$\Delta P_{co} = X_v + randn \cdot \sigma_v \quad (25)$$

with: $X_v = 0.5070$, $\sigma_v = 5.4146$ for Brent, $X_v = 0.4804$, $\sigma_v = 5.5025$ for WTI, and $randn$ a function that returns a random number belonging to a normal distribution with average value 0 and standard deviation 1. These equations are different from the model proposed by Manca (2013) because they consider the absolute shocks of CO prices, as the *log-likelihood* function respect to the normal distribution is slightly higher for the absolute shock distribution, even if it assumes different values according to the time period analyzed. It is important to underline immediately that the random walk model has not been used in the hybrid model creation. Indeed, Barzaghi and Conte (2015) overcame the evidence of crude oil price as a Markovian process by using the moving average. In other words, if you remove the background noise from CO quotations, the prices result to depend on the quotations of the two previous months. For further information about the moving average econometric model the reader can see Chapter 6.

Despite the detachment between these analyses, the random walk model results are described here, as it is a significant example of the econometric model features. Furthermore, the present Chapter considered the background noise in the statistical analysis of CO shocks (*i.e.* the drift ΔP_{co} contains the background noise). Figure 49 and Figure 50 represent the diagrams of possible cumulative forecast scenarios for CO prices from November, 2015 to November, 2019. A price limit is an assigned threshold beyond which the quotation is expected not to go (*i.e.* either increase or decrease) on any single trading day from the previous day's closing price. In the literature the minimum and maximum WTI price fluctuations are restricted to 0.01 and 15 USD/bbl respectively, according to the expiration of future contracts (Fong and See, 2002). To avoid excessive and unrealistic changes, it should be introduced a limitation for monthly and weekly variations according to the statistical results of previous periods. Indeed, shocks belong to the range between $-\beta_1 \cdot \sigma_v$ and $\beta_2 \cdot \sigma_v$, where β_1 and β_2 are 3 and 2 respectively. These values are extracted from the detailed statistical analysis carried out in this Chapter on standard deviation values of the shocks distribution, compared with the tails dimension of the shocks histogram. This model performs better for either short- (*i.e.* from hours/days to weeks) or medium-term (*i.e.* from months to maximum two year) simulations, because the stochastic contribution (*i.e.* the random function in Equation (25) spreads possible variation and consequently increases the uncertainty. As the time advances, the random function returns a price variation that can increase or decrease the previous quotation, either stabilizing the prices or creating scenarios that reach very high and low quotations. By observing the fan charts in Figure 51 and Figure 52 that represent the probability that future values belong to a certain price range, it is possible to observe that the scenarios of CO prices are concentrated in the range between 50 and 160 USD/bbl and follow the bullish trend of the data series analyzed. However, rather rare curves highlight the potential to significantly exceed the historical threshold values. In particular, WTI has 11 scenarios above 150 USD/bbl at the end of the prediction horizon, but no one price goes below 30 USD/bbl, while Brent has respectively 14 and 3 end quotations that are higher than 150 USD/bbl and lower than 30 USD/bbl, respectively.

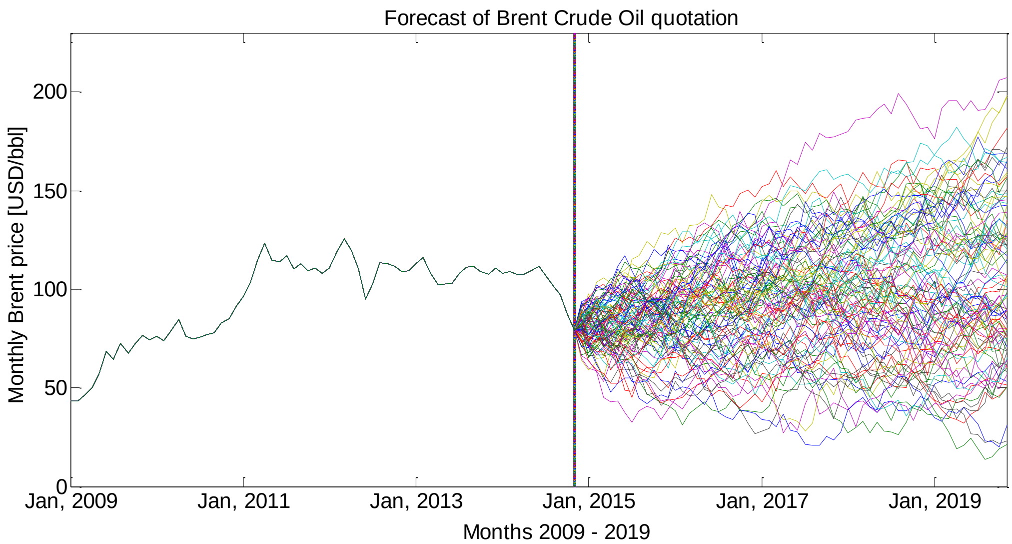


Figure 49 - Brent prices from 2009 to 2019: the period from November, 2014 to November, 2019 represents the forecast range (100 different simulations for 60 months).

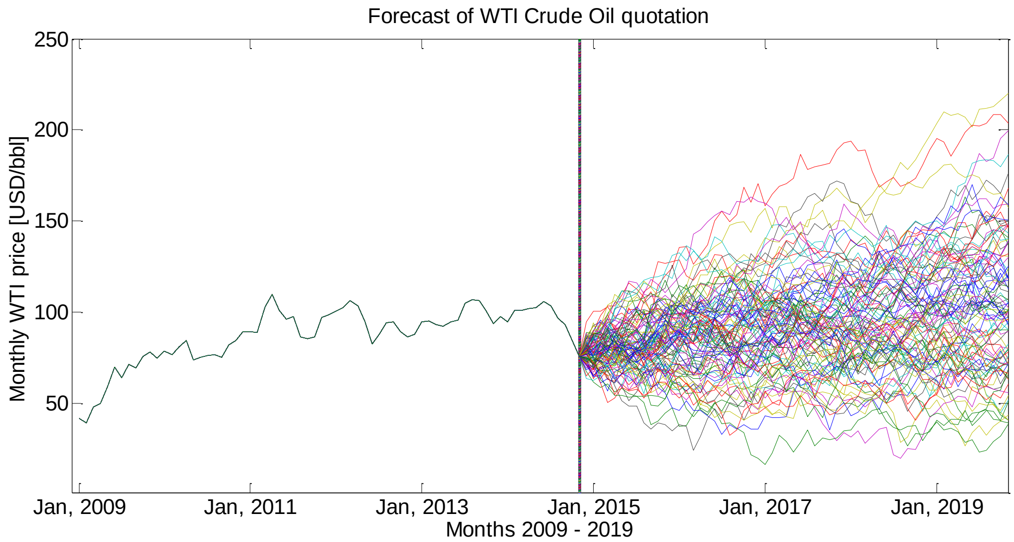


Figure 50 - WTI prices from 2009 to 2019: the period from November, 2014 to November, 2019 represents the forecast range (100 different simulations for 60 months).

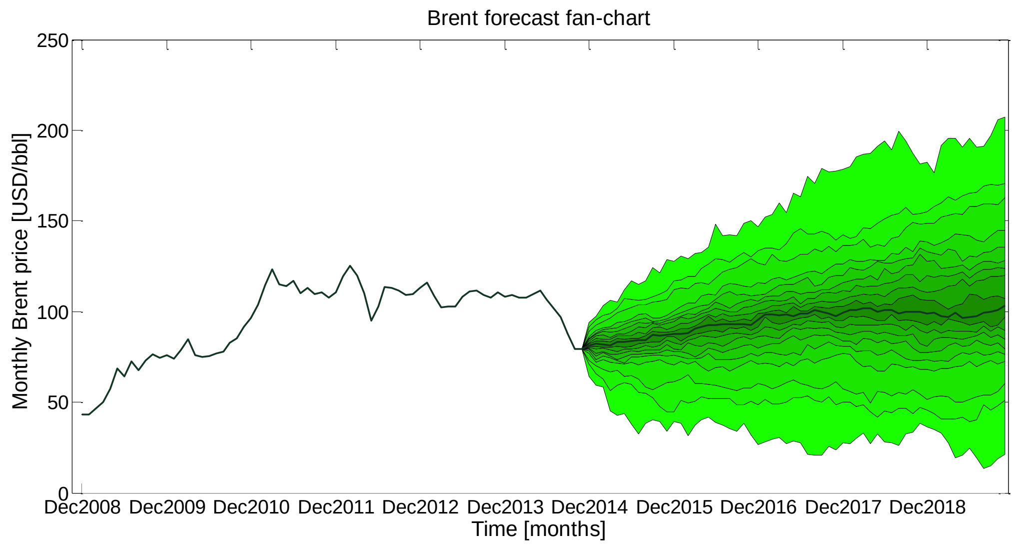


Figure 51 - Fan-chart of Brent prices from 2009 to 2019: the period from November, 2014 to November, 2019 represents the forecast range (100 different simulations) with probability from 0.1% (darker green) to 99.9% (lighter green).

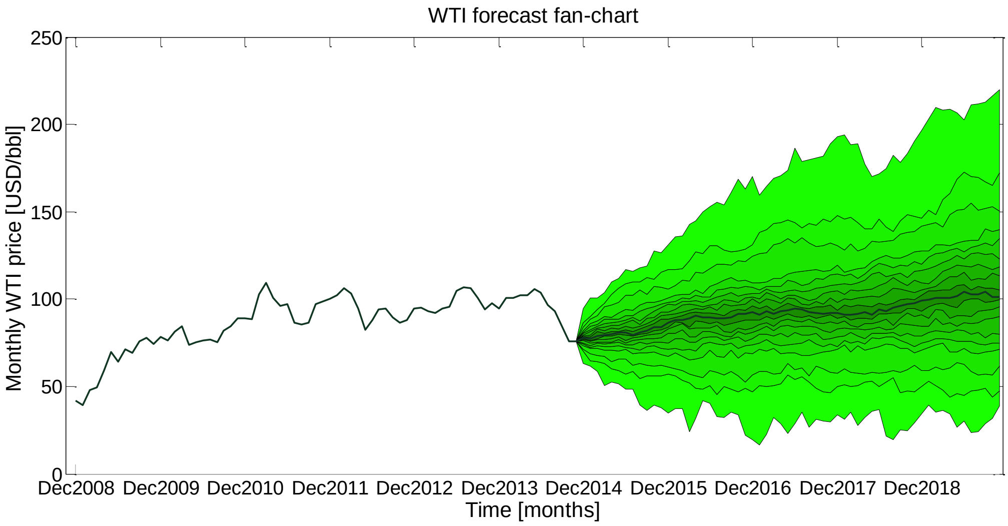


Figure 52 - Fan-chart of WTI prices from 2009 to 2019: the period from November, 2014 to November, 2019 represents the forecast range (100 different simulations) with probability from 0.1% (darker green) to 99.9% (lighter green).

The main power of this econometric model is that it catches the trend followed by the prices, studies possible future developments based on historical shock volatility, and allows getting probabilistic forecast bounds thank to fan-chart representation.

Chapter 4 Economic models

After the discussion on the economic models of Ye et al. (2009) and Chevallier (2010), this Chapter examines the fundamental factors that contribute to CO markets and quotations. Particular interest is devoted to inventories, excess production capacity, and distinction among physical, economic, and financial variables that influenced CO prices. At the same time, model simulations reveal the drawback of economic models that are based on future values of all the involved variables.

According to my present knowledge of the literature, economic models have not been used in the PSE/CAPE field even if they propose a valid contribution to planning and allocation purposes. Chapter 5 takes inspiration from these issues to overcome the problems that are traditionally conceived in the forecast of physical variables connected to CO market.

4.1 Introduction to the general features and issues of economic models

According to Aprea *et al.* (2005), the economic models involve purely economic variables (*e.g.*, GDP, USD exchange rate, inventories) and simulate the fluctuations of spot and futures market with supply-and-demand law. This is the case of the so called OPEC behavior model (Cooper, 2003; Kaufmann *et al.*, 2004; Hamilton, 2005; Dees *et al.*, 2007), Ye's model (Ye *et al.*, 2005, Ye *et al.*, 2006, Ye *et al.*, 2009), and Chevallier's model (Chevallier, 2014).

The complexity of economic models can be attributed to the diversity of factors that determine an economic process such as the quotation of CO. These issues include resource limitations (*e.g.*, non-renewable oil resources), environmental and geographical constraints (*e.g.*, stocking capacity of Cushing reservoir in the USA), institutional and legal requirements (*e.g.*, Russian sanctions, Iranian embargos), and purely random fluctuations (often due to the mood of investors originated by rumors and other subjective performance indexes). In finance these predictive models have been used since the '80s for trading and investment purposes, but at my knowledge they have not been used in the PSE/CAPE field, even if economic models can provide a valid contribution to scheduling and planning. The use of economic models on long-term horizons is somehow problematic and difficult to carry out as there is need for long-term forecasts of the levels of supply, demand, and capacity storage. The failure to consistently predict all the variables that affect the performance of prices over the long-run, such as technical advances in exploration and production of hydrocarbons, the political dynamics or changes in collective behaviors, and both national and international legislation makes the implementation and exploitation of economic models rather impractical as far as PSE/CAPE applications are concerned, because process designers/managers look for reliable models characterized by a simple structure and reduced number of parameters.

However, the importance of the economic literature lies on the fact that it acknowledges the impact of physical variables on CO quotations. Indeed, it often discusses the role of excess production capacity, capacity utilization rate (Ye *et al.*, 2009), and spare capacity (Chevallier, 2014) on market in general and on petroleum quotations in particular.

4.2 Ye's model

4.2.1 A brief literature review

Between 2005 and 2009 Michael Ye (St. Mary's College of Maryland, USA) developed a short-term model (Ye *et al.*, 2005; Ye *et al.*, 2006; Ye *et al.*, 2009) to monthly forecast the spot prices of WTI under normal market circumstances, *i.e.* his model did not take into account transitory geopolitical situations that can rise to different risk premiums, except for the effect that OPEC quota tightening had on petroleum market since April 1999 and the price disequilibrium following the terrorist attacks in the United States on September 11, 2001, when the price went down from 26.2 USD/bbl (September, 2001) to 19.71 USD/bbl (January, 2002). These market disequilibria are taken into account by means of dummy (*i.e.* Boolean) variables that are discussed in the model presentation (Section 4.2.3).

Ye's model deals with the fundamental relationship between inventories and prices and, even if the results are not relevant for forecasting purposes in PSE/CAPE applications, it investigates the impact on CO prices of fluctuating market fundamentals, such as inventories, production, imports, and demand.

Ye *et al.* (2005) focused on the concept of normal and relative levels. Ye and coauthors decomposed the observed level of petroleum market variables into two components: the normal level, which is determined by historical seasonal trends and reflects the normal operating requirements and market demand, and the relative level that is the difference between the observed level and the normal one, and represents short-horizon market fluctuations. As change in inventory equals the difference between demand and supply, which is the sum of field production and net imports, petroleum inventory demonstrates seasonal oscillations as well. Total petroleum inventory levels are a measure of the balance, or imbalance, between CO production and demand.

Ye *et al.* (2006) included two nonlinear inventory variables in the model: the first one is for the low-inventory state and the second one represents the high-inventory state. These variables were found to improve the ability to forecast the short-run CO price, because they capture at least some of the market psychology and traders' expectations in the short and middle-term horizon. Indeed, since inventory has a zero lower bound or some minimum operating requirements, short-term CO prices are expected to behave differently when the inventory level nears its lower bound than when it varies around its mid-range value.

At last, Ye *et al.* (2009) abandoned the aforementioned variables and studied the relationship between CO prices, inventories, and excess production. Furthermore a variable that is derived from cumulative excess production capacity was incorporated into the forecasting model to reflect the changing behavior on both supply-and-demand sides.

4.2.2 Determining fundamental factors: the role of inventory level and excess production capacity in economic models

Ye *et al.* (2006) investigated four potential inventory variables that also influenced the choice of the inventory variable adopted in the present work (see Chapter 5): (i) total industrial and governmental petroleum inventories, (ii) total industrial and governmental CO inventories, (iii) industrial petroleum inventories, and (iv) industrial CO inventories. The first two variables were ruled out since government inventories are generally of a strategic nature, do not vary significantly in the short-term, and are not determined by market forces. CO and product inventories are somewhat interchangeable, since CO inventories located near refining centers can be converted to product inventories relatively quickly. OECD end-of-period commercial inventory was ultimately chosen because it gives the best statistical results, according to Ye and coauthors. Long-term inventory trends exist due to the trends in government inventories, increases in long-term product demand, and the larger distribution and storage infrastructure needed to meet demand growth, and product differentiation. Among the relative level variables, it is the relative level of demand, field production, net imports, and inventory that matters most in the short-horizon forecast of CO price. Explicitly, the normal inventory level is calculated by:

$$IN_t^* = a_0 + b_1 \cdot T + \sum_{k=2}^{12}(b_k \cdot D_k) \tag{26}$$

where a_0, b_1, and b_k are estimated coefficient from de-trending and de-seasonalizing the observed total petroleum inventory, D_k are 11 seasonal dummy variables over a three year-long time horizon, and T is a linear trend. Instead, relative inventory level, which is denoted by RIN_t, is defined as the deviation of actual inventories from a historically determined normal level:

$$RIN_t = IN_t - IN_t^* \tag{27}$$

where IN_t is the actual industrial OECD petroleum inventory level in month t, and IN_t^* is the normal level given by Equation (26).

Ye *et al.* (2009) showed the relationship between OECD relative inventories and OPEC excess production capacity, reflecting the changing connection between supply-and-demand sides by three predictor variables: inventory, excess production capacity, and cumulative excess production capacity. The excess production capacity (CAP) is defined as the deviation from a critical level that alleviates market fears of shortage. Indeed, when it drops below an apparent critical level of 2 Mbbl/d market participants seem to become nervous about supply availability. So CAP is defined by:

$$CAP_t = max(EXCAP_t - 2\,, 0) \tag{28}$$

where $EXCAP_t$ is the excess production capacity at time t. This definition has the effect of ignoring its impacts on price changes when excess capacity is larger than 2 Mbbl/d.

Moreover, there is a cumulative impact of low excess production capacity, both in terms of market psychology and OPEC's responses to market acceptance of higher prices. Monthly cumulative excess production capacity over the critical level, denoted by $CUMCAP$, is defined as:

$$CUMCAP_t = \sum_{t=D}^{T} CAP_t \tag{29}$$

where D and T are respectively the initial and final time step of observation. This impact is acute at this time when the world CO demand is high and inelastic. Both the described variables play a central role in the main economic model that is described in Chapter 5, even if with different definitions and formulas.

4.2.3 Model description and simulation

Ye *et al.* (2005) proposed three comparative models, but the simulation of them is not performed in the present work. For the sake of completeness, the first proposed CO price model is the Relative Stock model that uses petroleum inventory as the only independent variable:

$$P_t = a + \sum_{i=0}^{3}(b_i \cdot RIN_{t-1}) + \sum_{j=0}^{5}\left(c_j \cdot D_j 911\right) + d \cdot LAPR99 + e \cdot P_{t-1} + \varepsilon_t \tag{30}$$

where subscript t represents the month, i is the subscript for the i-th month prior to the t-th month, j refers to the six months from October 2001 to March 2002, a, b_i, c_j, d, and e are adaptive coefficients to be estimated, $D_j 911$ is a set of single monthly variables that account for market disequilibrium following the 9-11 terrorist attack of 2001, and $LAPR99$ is a level-shifting variable that corresponds to the effect of OPEC quota on the petroleum market since 1999, when OPEC countries decided to cut their outputs. The last two variables, which take into account the market disequilibria determined by the abovementioned events, improve the model fit (Ye *et al.*, 2005). $D_j 911$ is a Boolean variable that accounts for the period between October 2001 and March 2002 (*i.e*. for 6 months after the terrorist attack). Instead, $LAPR99$ represents a deeper structural change, which impacted CO market from April 1999 to March 2003 (Ye *et al.*, 2005). Both periods are not included in the time-horizon that is involved in this thesis (*i.e.* from 2010 to 2015) and need to be neglected. The second alternative model is the AR model:

$$P_t = a + b \cdot AR(1) + \sum_{i=0}^{5}\left(c_j \cdot D_j 911\right) + d \cdot LAPR99 + e \cdot AR(12) + \varepsilon_t \qquad (31)$$

where $AR(1)$ and $AR(12)$ are the 1st and 12th order autoregressive terms, respectively. The third comparison model is the Modified Alternative model, which is specified by:

$$P_t = a + b \cdot P_{t-1} + \sum_{i=0}^{5}\left(c_j \cdot D_j 911\right) + d \cdot LAPR99 + e \cdot STK_{t-1} + f \cdot ANN_{t-1} + g \cdot T + \varepsilon_t \qquad (32)$$

in which STK is the total OECD industrial inventory level, and ANN is the difference between STK_{t-1} and STK_{t-12}.

The forecast model that is simulated in this Chapter is the one proposed in Ye *et al.* (2009):

$$P_t = \alpha_0 + \alpha_1 \cdot P_{t-1} + \sum_{k=0}^{2}(\beta_k \cdot RIN_{t-k}) + \gamma_1 \cdot CAP_{t+3} + \gamma_2 \cdot CUMCAP_t + \varepsilon_t \qquad (33)$$

where P_t is the WTI CO spot price in [USD/bbl], α_0, α_1, β_k, γ_1, and γ_2 are adaptive parameters whose values are presented in Table 7, RIN_t is the relative inventory level in [Mbbl], CAP_t is the deviation from the critical level of excess production capacity in [Mbbl/d], $CUMCAP_t$ is the absolute value of the cumulative excess production capacity, and ε_t is the error term. The capacity variable is defined using lead months (*i.e.* in financial terms, the soonest month in which a contract expires) because anticipated OPEC production constraints in the near future (*i.e.*, $t+3$) affect current oil prices.

Figure 53 shows the results of the simulations that were carried out in *Matlab©*. In particular, the one-step-ahead simulation uses the previous real data of the involved variables at each step, while the partial predictive simulation exploits the true values of prices at each step, but uses as RIN_t, CAP_t, and $CUMCAP_t$ values the monthly forecast data over a one-year period. Indeed, the model that was developed by Ye and coauthor had a monthly time-granularity, *i.e.* it worked with monthly prices and input variables. Partial predictive simulation forecasts twelve monthly values for RIN_t, CAP_t, and $CUMCAP_t$ variables and corrects them after one year by means of the real data.

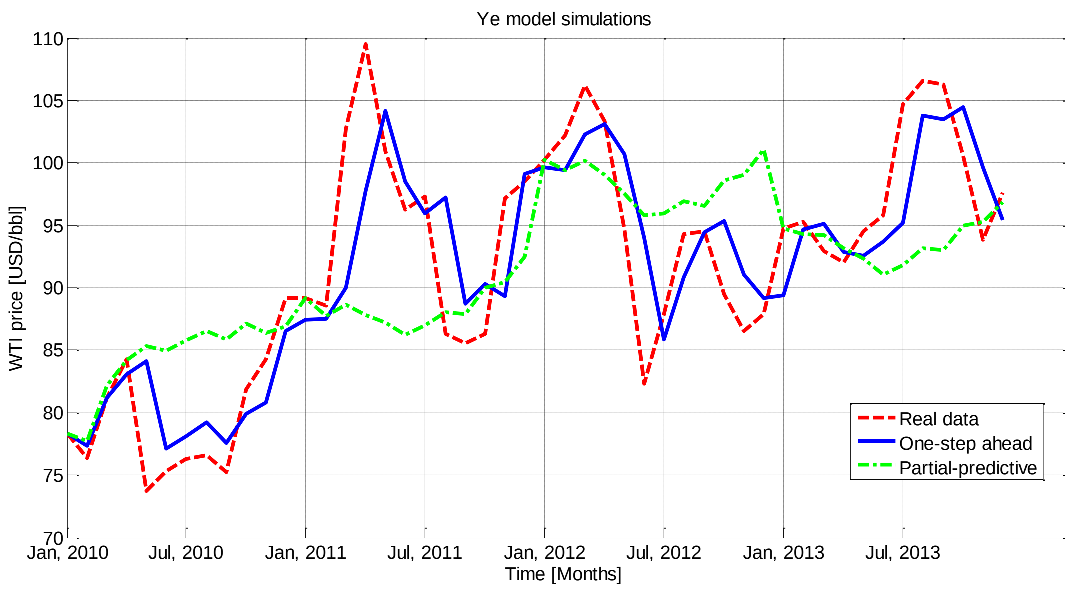


Figure 53 - WTI price simulations on monthly basis with one-step-ahead (red solid line) and partial-predictive (green dotted line).

The evaluation of the regression parameters goes through a linear regression procedure based on a multidimensional unconstrained optimization algorithm that minimizes the summation of square errors between the real data and the model values. Table 7 collects the models parameters that are evaluated by means of the *Matlab*© minimization function *fminsearch* and the correlation coefficient.

Table 7 - Adaptive parameters in Equation (33) for the model of WTI and Brent quotations and correlation coefficient.

Parameter	*Value*
α_0	55.2964
α_1	0.7177
β_0	0.0162
β_1	-0.0132
β_2	0.0430
γ_1	0.5897
γ_2	0.9969
R	0.86

The results of the forecasting model calculated in the sample period from January, 2010 to December, 2013 are only in part consistent with the expected real values, because of a number of weaknesses. First of all, the model requires as input values the future values of inventories and excess capacity, which are not available over long time intervals. The real volume of proved reserve and inventories is clouded by secrecy and the estimates of oil companies are not verified. It would be possible to model them roughly, but the weaknesses of the model are yet hardly surmountable. Secondly, that model is limited, as far as its robustness is concerned, by the time delay between true and calculated prices. However, some results are consistent with expectations. The cumulative excess production capacity has a positive impact on WTI prices, showing that when spare capacity is low, the price will be higher. Nevertheless, even if it is expected that relative inventory has a significant negative impact on price, this is not true.

The model structure (*i.e.* estimated parameters and Boolean variables that account for war, terroristic events, and changes due to market or financial issues) requires to be updated periodically because there should be a fundamental market change or a shift in the normal level of inventories that can underrate the characteristic time-period involved in PSE and DCD interests. Over time, forecast equation coefficients may need to be re-estimated to improve accuracy, and normal levels may need to be updated to reflect changes in trend and infrastructures. At my knowledge, I propose that this model should be updated yearly. However, the importance of this model lies in the definition and analysis of inventories and excess production capacity, which reflect both the supply-and-demand sides and are proposed in all the economic models though in different ways.

4.3 Chevallier's model

4.3.1 Determining fundamental factors: physical, economic, and financial variables

Chevallier (2014) considered the price relationships of CO futures by using a stepwise regression and pointed out that physical, macroeconomic, and financial factors are involved in the definition of CO quotations, suggesting the presence of excessive speculation during the 2008 oil price swing episode (see Section 1.1.2). Indeed, oil creates two distinct demands in the physical and financial markets: a demand for physical oil and a demand for paper oil. The different players in these markets (*i.e.* producers and users, swap dealers or banks, and money managers) may have different objects: some of them are the defenders of physical and macroeconomic fundamentals (*e.g.*, the demand from emerging countries, the fears of a shortage, the economic crisis, the role of exchange and interest rates), while the others represent the financial power (*e.g.*, the development of new products like commodity index funds, or the behavior of spot and futures markets). These features allow understanding the anomalous trends of CO quotations that occurred in recent years.

In analogy with Ye *et al.* (2009), Chevallier (2014) proposed a model with a high number of variables that are difficult to account for and implement whenever PSE/CAPE problems are concerned. Nevertheless, Chevallier's model provides some interesting interactions among variables involved in CO market.

Physical fundamentals define the dynamic balance between supply-and-demand. On the supply side, the OPEC CO spare capacity (*i.e.* "*the volume of production that can be brought on within 30 days and sustained for at least 90 days*" according to EIA, 2015) serves as an important indicator of the market tightness, because it shows how much the supply can theoretically increase within a short time horizon, faced with either growing demand or supply disruption (Chevallier, 2014). As for macroeconomic fundamentals, the recent global recession played an important role in the subsequent collapse of CO prices. A weaker US dollar means lower prices in Euro, which should translate into higher oil demand in Eurozone and tighten the global supply-and-demand balance. Financial fundamentals (*e.g.* S&P Goldman Sachs Commodity Index, Energy Spot Price Index, and Dow Jones Heating Oil Sub Index) go beyond the CO market alone and contribute to the operation of financial markets as a whole, where different types of assets are constantly competing with each other (Medlock and Jaffe, 2009).

4.3.2 Model description

The model proposed by Chevallier (2014) derives from a strategy that is referred to as *backward elimination*. In other words, the full model was estimated with all the independent variables, then the author withdrew one by one non-significant variables and re-estimated the model. At last, the restricted model is:

$$P_t = \alpha + AR(1) + MA(1) + \beta_1 \cdot OPECSPARECAP_t + \beta_2 \cdot OTHERASIANCONS_t + \beta_3 \cdot GSENSPT_t + \beta_4 \cdot PPIFEG_t + \beta_5 \cdot GDP_t + \beta_6 \cdot UMCSENT_t + \beta_7 \cdot MONEYM_t + \beta_8 \cdot PRODMERC_t + \beta_9 \cdot SWAP_t + \beta_{10} \cdot DUMMYREFIN + \beta_{11} \cdot WORKINGT_t + \varepsilon_t \qquad (34)$$

where $AR(1)$ is the 1st order autoregressive term, $MA(1)$ is the 1st order moving average term, $OPECSPARECAP$ is OPEC CO spare capacity [Mbbl/d], $OTHERASIANCONS$ is Asian countries' consumption except China [Mbbl/d], $GSENSPT$ is S&P Energy Spot Price Index [-], $PPIFEG$ is the producer price index [-], GDP is the Gross Domestic Product [MUSD], $UMCSENT$ is University of Michigan consumer sentiment index [-], $MONEYM$, $PRODMERC$,and $SWAP$ are the net positions of Money Manager, Producer, and Swap Dealer respectively [contracts of 1000 bbl], $DUMMYREFIN$ is the Boolean variable that takes into account the refining crisis during July and August 2008 [-], $WORKINGT$ is the working T index [-], ε_t is the error term [USD/bbl], and the coefficients are adaptive parameters. As in the case of Ye *et al.* (2006), Chevallier's model accounts for a past event that does not affect CO prices anymore, *i.e.* the refining and financial crisis that influenced quotations in July and August, 2008.

Some results of the paper deserve few words. Among the physical fundamentals of oil prices, the positive sign of $OPECSPARECAP$, which is the most significant variable, indicates that more supply of CO is translated into a price increase in a context of limited supply. Among the economic variables, GDP and $UMCSENT$ have a negative sign, which suggests that economic downturn fostered price decreases on the CO futures market. As for the economic fundamentals, the positive sign of $GSENSPT$ means that finance fosters CO price increases.

4.3.2.1 Disaggregated CFTC data

Chevallier (2014) used the data that the U.S. Commodity Futures Trading Commission (CFTC) has published since September 4, 2009, by releasing weekly time-series on the net positions in futures of different types of agents on U.S. commodity exchanges. The main interest in using disaggregated CFTC data consists in decomposing the "commercial" and "non-commercial" agents in order to detect the potential speculative behavior of market participants. As the CFTC categorizes agents as being commercial or non-commercial, many uncertainties arise concerning this classification. Indeed, the raw data before 2009 are not publicly available and the model proposed by Chevallier (2014) cannot be used in its original form.

4.3.3 Simulation of a new model inspired by Chevallier (2014)

Since the model cannot be simulated because of the lack of data, the present work of thesis shows a revised model that takes into account the confluence of physical, economic, and financial factors into CO price forecast. The adopted methodology is different than backward elimination. To analyze another possible model, some comparisons among the variables contained in the EIA databank (EIA, 2015) have been performed by calculating the correlation index, which measures in a non-dimensional way the pairwise interaction between variables. The higher the correlation index the better the functional dependency of these variables related to the economy and market of CO. As available statistics literature reports that values of the correlation index higher than 0.5 mean a rather good correlation between variables (Manca, 2013), the revised model is described by the following equation:

$$P_{co} = a \cdot OTCP + b \cdot MEPC + c \cdot TOP + d \cdot ONCLP + e \cdot NOFS + f \cdot TNOL + g \cdot NCP + h \cdot TOPS + l \cdot NOFC + m \cdot TWLFC + o \cdot GDP + p \cdot Q \quad (35a)$$

where P_{CO} is CO price [USD/bbl], $OTCP$ is OPEC total CO production capacity [Mbbl/d], $MEPC$ is OPEC middle east CO production capacity [Mbbl/d], TOP is OPEC total production capacity [Mbbl/d], $ONCLP$ is OPEC non-CO liquids production [Mbbl/d], $NOFS$ is non-OECD total CO and liquid fuels supply [Mbbl/d], $TNOL$ is non-OPEC total liquids production [Mbbl/d], NCP is global non-CO liquids production [Mbbl/d], $TOPS$ is OPEC total petroleum supply [Mbbl/d], $NOFC$ non-OECD total liquid fuels consumption [Mbbl/d], $TWLFC$ is global liquid fuels consumption [Mbbl/d], GDP is World real Gross Domestic Product Index [Index 2010 = 100], Q is a Boolean variable for July-August 2008 (*i.e.* it is taken into account only in July and August, 2008 to catch the influence of the speculative bubble just occurred), and all the coefficients (*a*-*p*) are adaptive parameters, whose values are calculated with the *Matlab*© minimization function *fminsearch*, and reported in Table 8. The different values of Brent parameters respect to WTI ones are mainly due to the dissimilar evolution of Brent and WTI quotations in recent years.

Table 8 - Adaptive parameters in Equation (35a) for the model of WTI and Brent quotations and correlation coefficient.

Parameter	*Brent*	*WTI*
$\boldsymbol{a}$	-15.1069	-11.5219
$\boldsymbol{b}$	17.8407	13.3263
$\boldsymbol{c}$	7.5056	0.1590
$\boldsymbol{d}$	-4.6115	-27.4094
$\boldsymbol{e}$	-4.8194	-2.9405
$\boldsymbol{f}$	-8.5461	-10.5231
$\boldsymbol{g}$	-0.6700	3.4104
$\boldsymbol{h}$	8.6260	12.8732
$\boldsymbol{l}$	1.0849	2.3534
$\boldsymbol{m}$	1.0348	-0.2983
$\boldsymbol{o}$	3.1102	2.9875
$\boldsymbol{p}$	24.2741	26.1497
R	0.92	0.87

The data used for the parameter estimation and simulations are taken from EIA (2015). The more interesting signs to determine fundamental factors are the ones related to OPEC CO production capacity, non-OECD total CO and liquid fuels supply, global liquid fuels consumption, and real GDP, which synthetize the main contributions of supply-and-demand variables, and macroeconomic influences.

The positive sign of OPEC CO production capacity is consistent with Ye's model and indicates in the same way that increased supply capacity of CO by OPEC brings to a price increase in a context of limited supply, because OPEC possesses the largest spare production capacity and has the power to influence the CO prices by changing the offer. Instead, non-OECD total CO and liquid fuels supply has a negative sign, because an increase in the offer leads to a decrease of prices. On the demand side, global liquid fuel consumption variable behaves in an uncommon way, because it has a different sign for Brent and WTI. This anomaly is detectable also in the global non-crude production, and can be interpreted by means of the dissimilar evolution of the quotations in European and American markets that brought WTI to a discount position respect to Brent. The positive sign of GDP signifies that an increase in the gross value determines a global growth of CO price.

Since aim of PSE forecasting activities (*i.e.* planning, scheduling, and supply chain management) is to get a trend of future prices according also to different geopolitical, economic, and financial incidents, the economic model should not provide a unique price trend, but show different trends (*i.e.* scenarios) that depend on the stochastic variations of CO quotations. Equation (35a) can be modified by introducing a stochastic (*i.e.* random) term related to CO price volatility, as follows:

$$P_{co} = (a \cdot OTCP + b \cdot MEPC + c \cdot TOP + d \cdot ONCLP + e \cdot NOFS + f \cdot TNOL + g \cdot NCP + h \cdot TOPS + l \cdot NOFC + m \cdot TWLFC + o \cdot GDP + p \cdot Q) \cdot (1 + randn \cdot \sigma_v) \qquad (35b)$$

where $randn$ is a function that returns a random number belonging to a normal distribution with average value 0 and standard deviation 1, and σ_v is the standard deviation of CO price shocks that in the analyzed time period (*i.e.* from January, 2005 to August, 2014) is equal to 0.0803 for Brent and 0.0823 for WTI. Figure 54, Figure 55, Figure 56, and Figure 57 report the results of one-step-ahead simulations and partial-predictive scenarios of Brent and WTI, respectively.

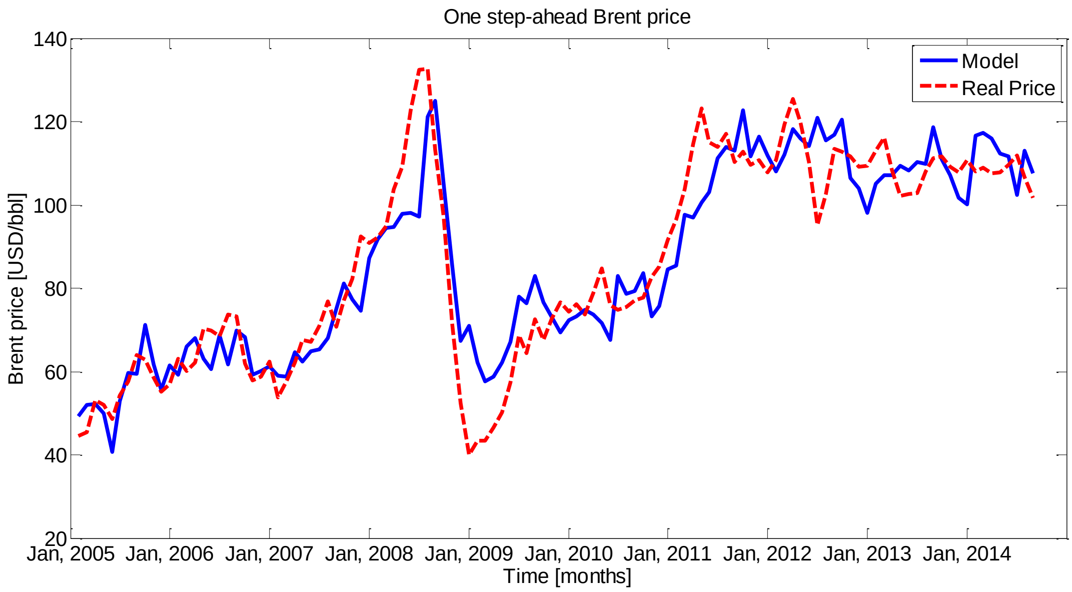


Figure 54 - One-step-ahead simulation of Brent monthly prices from January, 2005 to August, 2014 (real data from EIA).

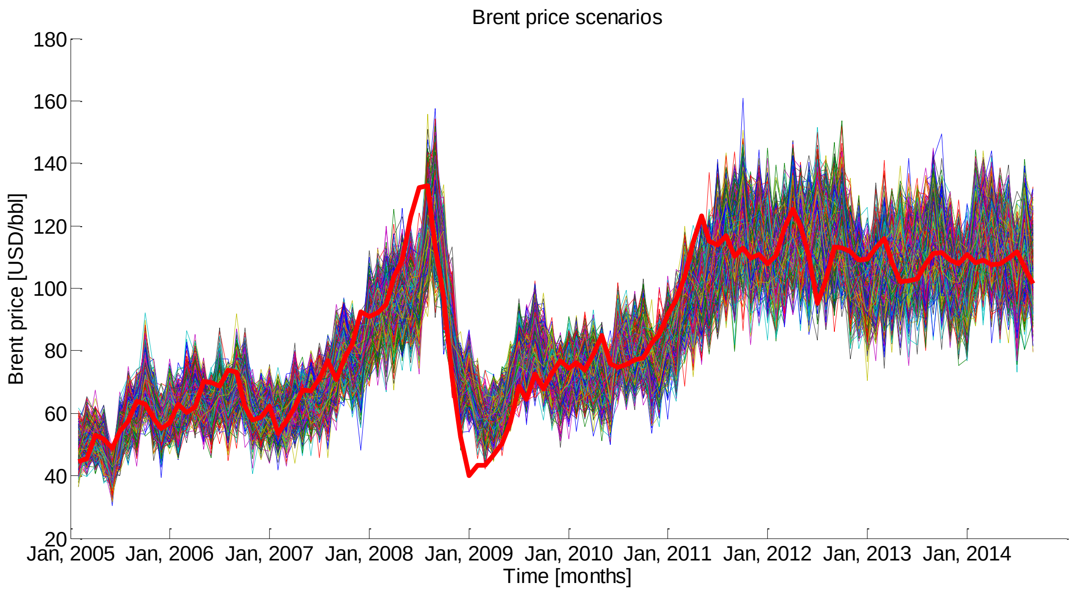


Figure 55 - Partial-predictive forecast scenarios and real Brent monthly prices (red solid line) from January, 2005 to August, 2014 (real data from EIA).

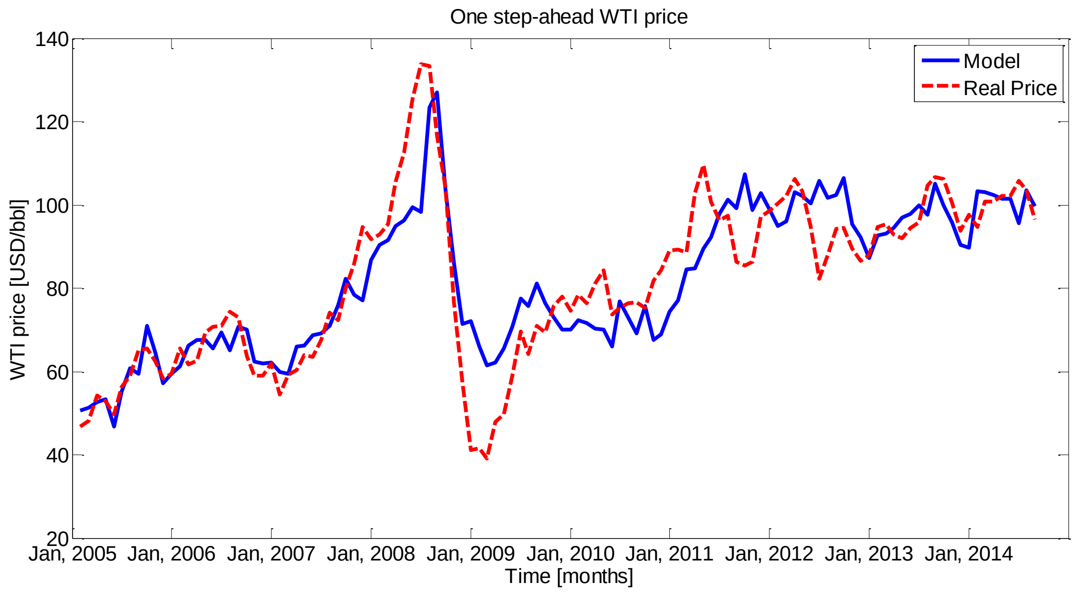


Figure 56 - One-step-ahead simulation of WTI monthly prices from January, 2005 to August, 2014 (real data from EIA).

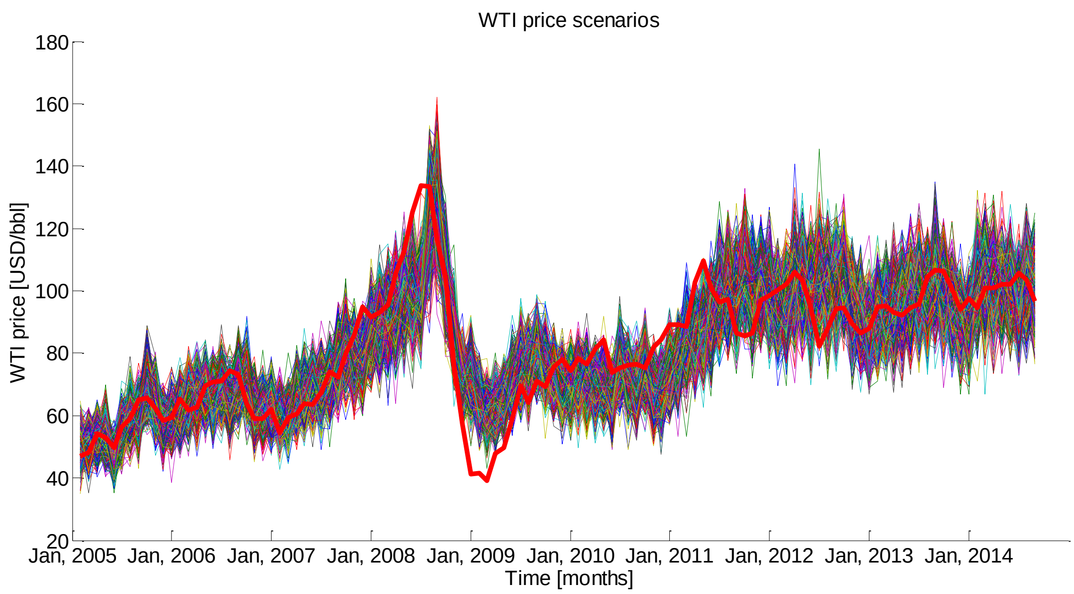


Figure 57 - Partial-predictive forecast scenarios and real WTI monthly prices (red solid line) from January, 2005 to August, 2014 (real data from EIA).

The accuracy of the results is acceptable despite the lag of few months. The created scenarios are defined as partial-predictive because they take into account stochastic variations of CO scenarios in Equation (35b), but employ the true past values of the independent variables. As already discussed for Ye *et al.* (2009) paper, the main disadvantage of this model is that it requires the future values of all the involved variables, which means further eleven predictive models. Conversely, the main power lies on the theoretical understanding of the physical and economic fundamentals related to CO price forecast, which are present also in the model described in the following Chapter (*i.e.*, the OPEC-based model).

Chapter 5 A new OPEC-based model for PSE applications

This Chapter is the heart of this thesis and describes the new economic model based on OPEC features. After a presentation of OPEC organization, the main historical events, and political decisions of oil countries, a brief review of OPEC-based models is offered. The innovation of the proposed model is that it takes into account the recent oil overproduction due to US shale oil, but also the stochasticity that is intrinsic to political incidents and decisions. This Chapter shows the model, its main characteristic, and proposes an identifiability analysis by means of the Fisher Information Matrix method and DAISY software, which is based on differential algebra. The model bets on CO price stabilization by means of production quotas, and considers both supply-and-demand variables. Thanks to a sensitivity analysis these input variables can be manipulated in order to create future scenarios, with an overall bullish or bearish trend, and simulate possible demand crisis, situations of oversupply, or economic and technological developments.

5.1 Historical review of OPEC behavior

The role of OPEC in the worldwide oil market has been examined by both the press (*e.g.*, Laherrère, 2011; Davis and Fleming, 2014; Suratman, 2015) and the academic community (Dahmani and Al-Osaimy, 2001; Molchanov, 2003; Sandrea, 2003; Horn, 2004; Kaufmann *et al.*, 2004; Dees *et al.*, 2007; Dvir and Rogoff, 2014) over the last decade. As already discussed in Section 1.2, OPEC is a permanent intergovernmental organization that is headquartered in Vienna and was created at the Baghdad Conference on 10-14th September 1960, by Iraq, Kuwait, Iran, Saudi Arabia, and Venezuela. Later, nine more governments joined OPEC and some of them left in different periods: Qatar (1961), Libya (1962), Indonesia (1962-2009, then included again in 2015), United Arab Emirates (1967), Algeria (1969), Nigeria (1971), Ecuador (1973-1992, then included again in 2007), Gabon (1975-1994), and Angola (2007).

OPEC was formed when the international oil market was largely dominated by a group of multinational companies known as the *Seven Sisters*, which tried to eliminate competitors and control the world's oil resources. Indeed, prior to 1970s these oil industries controlled around 85% of the global petroleum reserves, and comprised Anglo-Persian Oil Company (now BP), Gulf Oil, Standard Oil of California (SoCal), Texaco (now Chevron), Royal Dutch Shell, Standard Oil of New Jersey (now Exxon), and Standard Oil Company of New York (now part of ExxonMobil). In recent decades the dominance of these companies and their successors declined as a result of the increasing influence of the OPEC cartel (*i.e.* an organization created from a formal agreement among a group of producers in order to regulate supply and manipulate prices) and of the emerging market economies (*i.e.* BRIC countries). For instance, the expression *New Seven Sisters* indicates the group of most influential national oil and gas companies that are based in countries outside OECD and comprises OPEC and BRIC countries, such as Saudi Arabia, Iran, Venezuela, China, Russia, and Brazil (Hoyos, 2007). Almost 72% of the world's proven oil reserves are located in OPEC countries, with the bulk of CO reserves in the Middle East, which counts for 66% of the OPEC totals (OPEC, 2014; BP, 2015). The Saudi Arabia reserves are about 20% of the World's total conventional oil reserves and its production amounts to 40% of the stated reserves (EIA, 2014). The reader can find more information about OPEC reserves, production, consumption, and refinery capacity in Section 1.2.

The OPEC Conference is the supreme authority of the organization, and consists of delegations normally headed by the Ministers of Oil, Mines, and Energy of member countries. The Conference usually meets twice a year (*i.e.* in March and September) and in extraordinary sessions whenever required. As for OPEC's aims and policy, its mandate is to coordinate and unify the petroleum policies of its members and to ensure the stabilization of oil markets in order to ensure an efficient, economic, and regular supply of petroleum to consumers, a steady income to producers, and a fair return on capital for those investing in the petroleum industry (OPEC Statute, 2012). Actually, its formation represented a collective act of sovereignty by oil exporting nations, and marked a turning point in state control over natural resources (Molchanov, 2003). For instance, in the 1960s OPEC ensured that oil companies could not unilaterally cut prices. In the 1970s, OPEC began to gain influence and steeply raised oil prices. In particular, in October 1973, OPEC declared an oil embargo in response to the United States' and Western Europe's support to Israel in the Yom Kippur War of 1973. Although the embargo only lasted a year, during that time oil prices quadrupled (from 3 USD/bbl to 12 USD/bbl starting on 17th October 1973 and ending on 18th March 1974) and OPEC members discovered that their oil could be used as a political and economic weapon against other countries. Other factors of the rise in gasoline prices included the market and consumer panic reaction, the peak of oil production in the USA (Hubbert, 1956; Hubbert, 1962), and the devaluation of the USD.

In the 1980s, the price of oil was allowed to rise before the adverse effects of higher prices caused demand and price to fall. The OPEC countries experienced severe economic hardship from lower demand for oil and consequently cut production in order to boost again the price of oil. In response to the high oil prices of the 1970s, industrial nations tried to reduce their dependence on CO. Utilities switched to using coal, natural gas, and nuclear power, while national governments initiated multi-billion dollar research programs to develop alternatives to oil. Demand for CO dropped by 5 Mbbl/d, while oil production outside of OPEC rose by 14 Mbbl/d since 1986. During that time, the percentage of oil produced by OPEC fell from 50% to 29% (EIA, 2015). The result was a six-year price decline that culminated with a 46% price drop in 1986 (Figure 58).

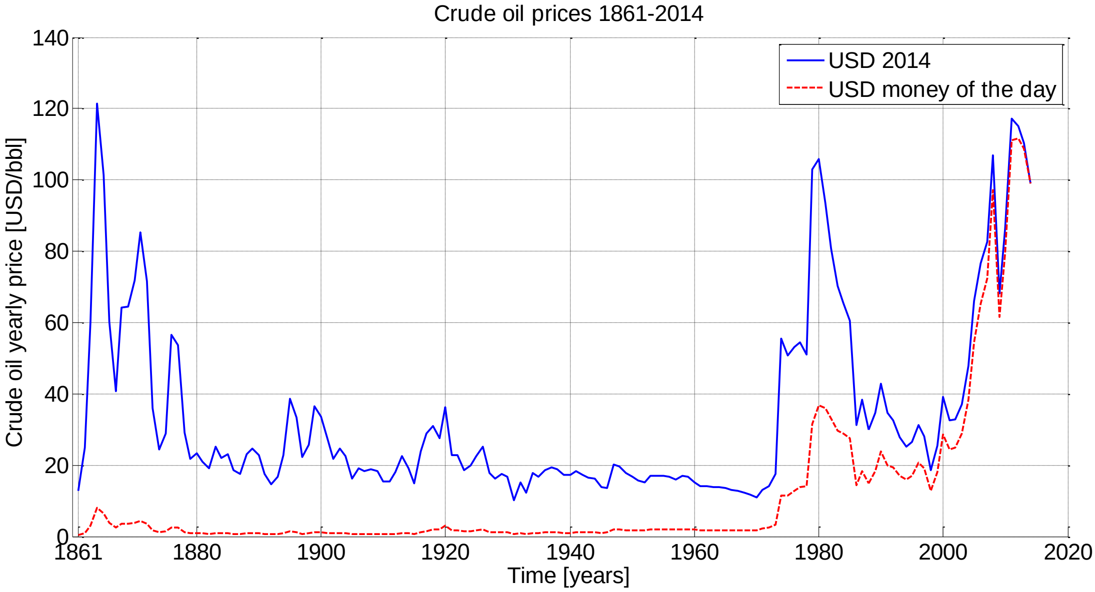


Figure 58 - Historical crude oil yearly price data from 1861 to 2014 with the current USD value and the money value of that year (data taken from BP, 2014).

The invasion of Kuwait and the Gulf War in 1990-1991 marked a low point in the cohesion of OPEC. Once supply disruption fears, which accompanied these conflicts, dissipated, oil prices began to fall dramatically. After CO prices slumped at around 15 USD/bbl in the late 1990s (EIA, 2015), joint diplomacy achieved a slowing down of CO production beginning in 1998.

Following the price crash of 1998, when oil prices hit 10 USD/bbl in February 1999, in March 2000 OPEC members decided to stabilize oil prices within a range of 22-28 USD/bbl by tuning output at their discretion. That mechanism was set up by OPEC to adjust production levels for both OPEC as whole and individual member countries, if prices moved out of the band for a previously defined period (Dahmani and Al-Osaimy, 2001; Sandrea, 2003; Horn, 2004). In the case of high prices (above the range), a production increase is needed, while low prices and cut in production have the opposite effect on the market. As an example of the disposition to keep prices in line, when CO prices began climbing to the upper limit of the band at the end of September 2002, OPEC increased production by 760 kbbl/d. By the end of October of the same year, crude prices had fallen more than 10%, returning to their target range (Sandrea, 2003).

Some of the last decade market trends and historical events about OPEC have been already discussed in Chapter 1. In front of the emerging scenarios (*i.e.* the increasing concern about shale oil in the USA, which could become independent of OPEC decisions about quotas and production in the near future), Middle East producers decided not to cut their quotas at the end of 2014 to maintain their production competitive with the global CO spot market. The highest authorities of Saudi Arabia declared acceptable that prices remained low for long periods if that would reduce investments in shale oil and rebalance global markets (Bellomo and Negri, 2014; Sole 24 Ore, 2014). OPEC argued that this drop in the price of oil was not exclusively attributed to oil market fundamentals, such as ample supply, moderate demand, a stronger US dollar, and uncertainties about global economic growth, but also the speculative activity in the oil market has contributed to the drop in price (Davis and Fleming, 2014).

It is noteworthy to spend few words about the country members that changed their position in OPEC in a more or less close past. Ecuador and Gabon were early members of OPEC, but Ecuador withdrew on 31st December 1992, because it was unwilling or unable to pay a 2 million USD membership fee and needed to produce more oil than it was allowed to do under the quota system, although it rejoined in October 2007. Similar concerns prompted Gabon to suspend membership in January 1995. Iraq remains a member of OPEC, but Iraqi production has not been part of any OPEC quota agreements since March 1998. In May 2008, Indonesia announced that it would leave OPEC because it became a net importer of oil and was unable to meet its production quota. The departure from OPEC did not likely affect the amount of oil produced or imported by Indonesia. In June 2015, the Southeast Asian country re-joined OPEC and increased its business relations with Saudi Arabia oil companies (Suratman, 2015).

The economic needs of OPEC members have often affected the internal politics behind OPEC production quotas. Various members pushed for reductions in production quotas to increase the price of oil and thus their own revenues. Part of the basis for this policy is the Saudi Arabia concern that expensive oil or supply uncertainty would drive developed nations to conserve and develop alternative fuels. The following Sections discuss in more detail the quota system and its implications on CO historical price trend.

5.1.1 OPEC quota policy

Problems in the international oil market usually result from an imbalance between supply and demand (Dahmani and Al-Osaimy, 2001). In order to avoid such an imbalance, OPEC adopted various strategies at different times. Since 1982, the Organization has self-imposed to fix an overall production ceiling with individual quotas for each member in accordance with each country’s output capacity. The purpose has been to achieve and maintain equilibrium between supply and demand, in order to ensure CO prices that are stable and at reasonable levels. The production ceiling has been frequently adjusted in a systematic and timely manner (see Section 5.2.3.6 for more details), according to prevailing market conditions. The present work of thesis does not discuss the original quota system that OPEC set up in 1982, as it was revised several times and finally suspended in 1986. In 1993, OPEC countries adjusted the quotas distributions to the upgraded production potentials of the member countries. OPEC countries have analyzed market demand and CO spot prices constantly to control their output and market in general. In particular, OPEC conducted an analysis of the allocating quota system and defined eight criteria that have influenced members' production ceilings, *i.e.* reserves, production capacity that provides the practical upper limit on a country's quota, historical production share, domestic oil consumption, production costs, population, dependence on oil exports, and external debt (Sandrea, 2003; Laherrère, 2011). Indeed, also today members with economies that are most heavily dependent on oil revenues for export earnings (primarily Gulf States) tend to overproduce respect to countries with a more diversified international trade profile, *e.g.* Venezuela and Indonesia (Molchanov, 2013). It is reasonable to surmise that an OPEC member with high population relative to other members might have greater need for revenues to pay for social services, which could be raised by producing above quota. A rise in the public debt burden may lead to greater overproduction, particularly for members that lack a diversified export sector and are unable to easily raise revenues for financing the debt through other means. On the other hand, it is entirely possible that the pressure of a rising population may lead a government to violate quotas, as seems to be the case of Algeria and Nigeria (Molchanov, 2013).

The quota system is needed to distribute the output ceiling among the countries by assigning production aliquots to each member. The country aliquots are determined by the following formula (Sandrea, 2003):

$$Q_i = Q_{cap,i} \cdot (Q_c - Q_{obt}) + Q_{ob,i} \tag{36}$$

where Q_i is the country aliquot expressed in Mbbl/d, $Q_{cap,i}$ is the basic country quota expressed as a fraction of the total OPEC output capacity, $Q_{ob,i}$ is the country's bonus quota for new field discoveries in Mbbl/d, Q_{obt} is the total OPEC bonus quotas in Mbbl/d, Q_c is the established production ceiling in Mbbl/d, and i is the country index. It should be pointed out that in Equation (36) the total bonus quotas are subtracted from the production ceiling in order to guarantee that production cutbacks will be distributed equitably among all members, and the bonus quotas are subsequently added to each member allocation. Table 9 shows the quota percentage of OPEC countries in different time periods (*i.e.* 1982-1986 and 1993-2002). It is evident that they made major adjustments to bring every nation in line with their current output capacity. In particular, Qatar, Venezuela, Saudi Arabia, and Iran received substantial quota increases, while all other ones were adjusted downwards.

Table 9 - Quota percentages of OPEC countries (data taken from Sandrea, 2003).

Country	Quotas [%]	
	1982 - 1986	1993 - 2002
Algeria	5.03	2.89
Indonesia	5.67	4.98
Iran	11.72	12.99
Iraq	12.79	8.42
Kuwait	8.70	7.36
Libya	5.30	4.94
Nigeria	7.51	7.32
Qatar	1.69	2.36
Saudi Arabia	24.03	29.68
UAE	8.91	8.09
Venezuela	8.54	10.98

Production generally exceeded the ceiling (Figure 59) and, in many cases, the excess was significant. Some papers (Dahmani and Al-Osaimy, 2001; Molchanov, 2003; Sandrea, 2003; Horn, 2004; Laherrère, 2011) described the tendency of OPEC members to exceed their production quotas.

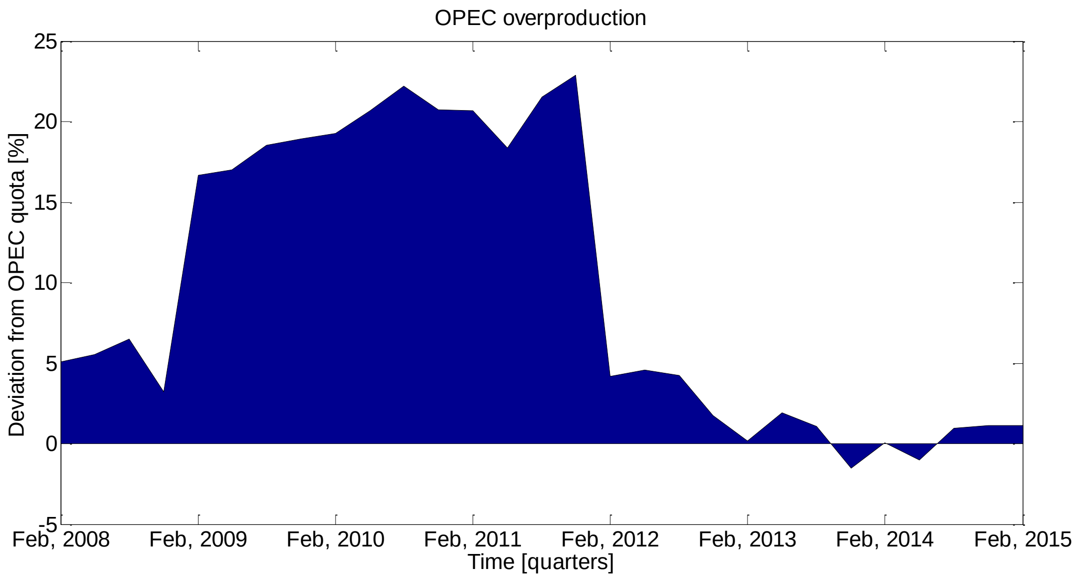


Figure 59 - OPEC deviation from quota (data from EIA, 2014).

The analysis of OPEC overproduction starts from Molchanov (2003), who led an interesting analysis about OPEC quota and overproduction trends, and modeled it as a rational economic strategy. That paper used the definition of deviation from the official quota as the "*ratio of the difference between the production and the quota to the quota itself*" over a period of time:

$$D_t = \frac{P_t - Q_t}{Q_t} \cdot 100 \tag{37}$$

A positive value given by this formula represents overproduction above the allotted quota, while a negative value signifies underproduction.

A moderate level of overproduction (5% to 7% above quota) has been the norm for OPEC over the last twenty years, but it reached also more than 20% (as reported in Figure 59). It may in fact be useful to think of the quotas not as immutable output ceilings, but rather as floors, beneath which actual production almost never falls. There are two sets of factors that were analyzed in literature to discuss the nature and extent of quota-breaking over the last years, *i.e.* common features of the international oil market and country-specific characteristics. The OPEC Statute asserts that "*Member Countries shall fulfill, in good faith, the obligations assumed by them*," but it also refers to members' "*sovereign equality*" (Molchanov, 2013). In practice, collective decisions, though binding, are not backed by any disciplinary action, such as suspension of membership. Indeed, OPEC has no way to enforce compliance by its members with the agreed-upon quotas. Indeed, there have always been members dissatisfied with their assigned quotas (Sandrea, 2003).

Dahmani and Al-Osaimy (2001) stated several oil market fundamentals as explanatory variables that affect OPEC's overall level of overproduction. First of all, they showed that most statistically significant influence on deviation from quotas comes from OECD oil stocks and that the correlation is strongly negative. OECD demand is also significant with positive correlation, as greater demand should logically lead producers to expand their CO outputs. Furthermore, member overproduction tends to fall when OPEC's total production ceiling is increased, even when such an increase benefits members proportionally. Small producers (*e.g.*, Qatar or United Arab Emirates) have historically engaged in far greater quota-breaking than the cartel's largest members. Indeed, a member with a quota that is 1% larger than another member's production allocation tends to produce less in proportion, *i.e.* it has a level of overproduction that is more than 1% lower.

5.1.2 OPEC behavior and price trend

As it is represented in Figure 58, when OPEC became the dominant producer in the early 1970s, price volatility increased tremendously (Dees *et al.*, 2007). In order to explain that trend, Dvir and Rogoff (2009) argued that the real price of oil went through three distinct periods. First, from 1861 to 1878 the price of CO was generally high in real terms and was moreover highly persistent (*i.e.* it moved in the same direction, upward or downward) and volatile. Then a much less volatile period came between 1878 and 1973, and prices were generally lower and not at all persistent. This long period can be further divided into two sub-periods, before and after 1933, where price volatility was significantly lower after 1933 compared with previous years. Finally, from 1933 onwards there is a recurrence of high persistence and volatility accompanied again by high prices. In both periods that were characterized by high price persistence (*i.e.* 1861-1878 and 1973-nowadays) two forces have coincided. First, demand was high and very persistent; second, access to supply was restricted by agents who had the capability and incentive to do so. In particular, after 1973 all the excess capacity was in the Middle East, where producers were more interested in maintaining high prices by manipulating the offer than in accommodating demand increases (Dvir and Rogoff, 2014). Ye *et al.* (2009) further subdivided the CO prices from early 1990s to present into three distinct regimes. Indeed, during the early 1990s OPEC had excess production capacity that could be used to meet unexpected demand increases. However, as world CO demand grew, that excess capacity diminished and reduced the perceived ability of producers to meet demand increases in short-time horizons. The first sub-period, which represented a stable market with WTI prices averaging about 20 USD/bbl, was between January 1992 and June 1999. The second one reflected OPEC's attempt to reestablish control of the CO market and extended from July 1999 to May 2004 with WTI prices averaging 30 USD/bbl. The last period ran from June 2004 to December 2007 with CO prices increasing with virtually no excess production capacity for CO. The last eight years (2008-2015) were discussed extensively in Chapter 1.

The next Section presents the economic models that were inspired by these political and historical considerations about OPEC behavior, and a revised model that catches the recent market trend and the new variables involved in the global background.

5.2 Models description

5.2.1 Economic models inspired by the behavior of OPEC

It is a standard modeling practice to represent the world oil market in terms of a supply and demand equilibrium (Weiqi *et al.*, 2011). Existing simple models (*e.g.* econometric models presented in Chapter 3) lack a theoretical economic basis, while more complex models (*e.g.* Chevallier, 2014) provide insight into economic fundamentals behind the market behavior, but they are not practically useful as they require too much expertise and specific data to be easily implemented (*e.g.* Section 4.3.2.1). The specific characteristics of the CO market make its modeling a particularly complex endeavor, as oil prices react in a complex fashion to changes in market conditions. Indeed, within CO markets, OPEC and non-OPEC behaviors need to be distinguished (Dees *et al.*, 2007). The former is modeled according to a cooperative behavior, in which OPEC matches production to demand, while non-OPEC production is modeled by means of geological and economic factors. Indeed, non-OPEC production has had a significant effect on OPEC's share of the world CO supply and, as a consequence, on OPEC's ability to influence its prices. Such a model would produce a price at which OPEC is ready to act as a swing producer (*i.e.* a supplier that controls the global deposits of this raw material and possesses large spare production capacity), given new demand conditions, and would include market indicators that reflect the effect of the dominant producer behavior. Given a certain price, demand determines the optimal quantity of CO sold. Non-OPEC countries adapt their production to this new price and OPEC acts as a balancing producer to equilibrate supply and demand to their optimal levels.

The scientific literature identified a cartel model, where OPEC is the price maker that has the power to influence the CO prices it charges, and a competitive model, where OPEC is the price taker that can alter its rate of production without significantly affecting CO price and market (Kaufmann *et al.*, 2004; Dees *et al.*, 2007; Al-Qahtani *et al.*, 2008). Regression results indicate that OPEC is able to influence CO prices by means of capacity utilization, setting production quotas, and the degree to which OPEC production exceeds these quotas. According to Dees *et al.* (2007), oil prices are defined by a rule based on changes in market conditions and OPEC behavior. OPEC decisions about quota and capacity utilization have a significant and immediate impact on oil price and some models were developed to comply with these hypotheses. Indeed, oil markets are characterized by the existence of a cartel together with the presence of independent producers. That model is credited by its own power to include both the CO producers and consumers that for the sake of simplicity can be assumed to be clustered in either OPEC or OECD (see Section 1.2). Some authors studied the behavior of OPEC and its decisions (Cooper, 2003; Kaufmann *et al.*, 2004; Hamilton, 2005; Dees *et al.*, 2007). All these studies used the variables and equations summarized in the work of Kaufmann *et al.* (2004), who estimated a model for real oil prices that includes variables that represent market conditions, such as OECD stocks for CO, capacity utilization by OPEC, OPEC announced quotas, and the degree to which its production adheres to these quotas. It contains also a dummy variable for the Persian Gulf War (1990-1991), which is not included in the quarters analyzed in the sample period of this thesis. The involved variables are not easy to model and forecast because they are function of the economic activities of the countries (either producers or consumers). The abovementioned papers are not clear about how to forecast the inventories and production capacity. The work of Kaufmann and coauthors proposed the following CO price model:

$$PRICE_t = \alpha_0 + \alpha_1 \cdot DAYS_t + \alpha_2 \cdot Quotas_t + \alpha_3 \cdot Cheat_t + \alpha_4 \cdot Caputil_t + \alpha_5 \cdot Q_1 + \alpha_6 \cdot Q_2 + \alpha_7 \cdot Q_3 \qquad (38)$$

where $PRICE_t$ is the CO quarterly price at time (quarter) t [USD/bbl]; $DAYS_t$ is the number of days of forward consumption of OECD CO stocks, which is calculated by dividing OECD CO stocks by OECD CO demand; $Quotas_t$ is the OPEC production quota [Mbbl/d]; $Cheat_t$ is the difference between OPEC CO production and OPEC quotas [Mbbl/d]; $Caputil_t$ is the capacity utilization by OPEC, which is calculated by dividing OPEC production [Mbbl/d] by OPEC capacity of production [Mbbl/d], Q_1, Q_2, and Q_3 are dummy (*i.e.* Boolean) variables for quarters 1, 2 and 3, respectively, which improve the modeling of the price over the time interval analyzed. Equation (38) is linear in the adaptive parameters α_i and the prices of CO can be calculated directly with the following matrix formula:

$$\begin{pmatrix} \widehat{P_1} \\ \hat{P}_2 \\ \cdots \\ \hat{P}_n \end{pmatrix} = \begin{pmatrix} 1 & Days_1 & Quotas_1 & Cheat_1 & Caputil_1 & 1 & 0 & 0 \\ 1 & Days_2 & Quotas_2 & Cheat_2 & Caputil_2 & 0 & 1 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & Days_n & Quotas_n & Cheat_2 & Caputil_n & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \widehat{\alpha_0} \\ \widehat{\alpha_1} \\ \cdots \\ \widehat{\alpha_7} \end{pmatrix} \tag{39}$$

The main disadvantage of this model is that it cannot forecast the unpredictable conflicts, tenses, and events that may occur worldwide and their influence on the CO market, such as what happened in the first months of 2011, when the conflicts in Libya and the tsunami in Japan, combined with the following Fukushima nuclear disaster, played a significant role in increasing the petroleum prices. Similar comments can also be made for the political situation of Iraq which impacted significantly the quotations of CO in the recent past.

The forecast of future prices of CO calls for developing suitable models capable of describing the evolution of different input variables (*i.e.* OECD demand, OECD inventories, OPEC production, and OPEC production capacity). The input variable models that are included in the existing literature are described in next Sections to show their failures and the consequent breakthrough of the new economic model devised in this thesis. We anticipate that, according to Dees *et al.* (2007) and Kaufmann (1995), the oil supply for non-OPEC producers was derived from a competitive behavior taking into account the effect of geological and economic variables (Hubbert, 1962; Fisher, 1981), while econometric estimations showed that the demand for oil is explained by the economic activity, the real price of CO, and a time trend representing technological developments related to energy efficiency. The new price and input variable models should take into account the main pros and cons of the OPEC-based model that are included in the literature, especially in view of the application to planning, scheduling, and feasibility studies of chemical plants.

5.2.2 New economic crude oil price model

The new CO model takes into account those market variables that reflect the effect of behavior among the dominant producers (*i.e.* OPEC), the competitive producer (*i.e.* USA), and the price takers (*i.e.* other OECD countries, where most of the demand is concentrated). The economic estimation by Kaufmann *et al.* (2004) of the price rule indicates that CO prices are affected by OPEC capacity utilization, OPEC production quotas, the degree to OPEC members cheat on those quotas, and CO stocks in OECD countries. The new economic model takes into account also the production surplus originated by US shale oil. According to the New York Times (Clifford, 2014) the oil-drilling boom in the United States has increased oil production by over 70% since 2008 and has reduced the United States oil imports from OPEC by 50%. Indeed, the oversupply due to US shale oil has been counted as the main cause of the 50% drop in CO prices from July 2014 to December 2014.

The new economic model is:

$$PRICE_t = \alpha_0 + \alpha_1 \cdot Days_t + \alpha_2 \cdot Quotas_t + \alpha_3 \cdot Cheat_t + \alpha_4 \cdot Caputil_t + \alpha_5 \cdot Delta_t \tag{40}$$

where $PRICE_t$ is the CO quarterly price at time (quarter) t [USD/bbl]; $DAYS_t$ is the number of days of forward consumption of OECD CO stocks, $Quotas_t$ is the OPEC production quota [Mbbl/d], $Cheat_t$ is the difference between OPEC CO production and OPEC quotas [Mbbl/d]; $Caputil_t$ is the capacity utilization by OPEC, which is calculated by dividing OPEC production [Mbbl/d] by OPEC capacity of production [Mbbl/d], $Delta_t$ is the oversupply of OPEC respect to the USA production in [Mbbl/d], which seems to be relevant since four quarters ago (*i.e.* Summer 2014), and α_i are the adaptive parameters that are calculated by means of a multi-linear regression function in *Matlab*©. In mathematical terms, the involved variables can be rewritten by means of the input variables:

$$Days_t = \frac{Inventory_t}{Demand_t} \tag{41}$$

$$Cheat_t = Production_t^{OPEC} - Quotas_t \tag{42}$$

$$Caputil_t = \frac{Production_t^{OPEC}}{Capacity_t^{OPEC}} \tag{43}$$

$$Delta_t = Production_t^{OPEC} - Production_t^{USA} \tag{44}$$

The first innovation of this model is that it takes into account the offer surplus originated by the shale-oil revolution. The last term of Equation (40) accounts for the recent decrease between OPEC and US productions that has fallen under 22 Mbbl/d since the fourth quarter of 2013. The simulation of the economic model takes into account also the recent disturbances that influenced and changed significantly the CO markets. Sections 5.2.6 and 5.3 provide a more detailed description of the possible price scenarios.

Figure 60 shows the historical data of the involved variables in a 3-D domain. As it can be observed, the greater the oversupply due to increasing $Delta$ the bigger the CO price. The same consideration is valid for the variables $Caputil$ and $Cheat$. Historical data show that the relationship is reverse for $Days$, as either a stock increase or a demand reduction push the CO prices to fall. Adaptive parameters calculation and sensitivity analysis provide consistent results (see Section 5.2.5.1 for more details).

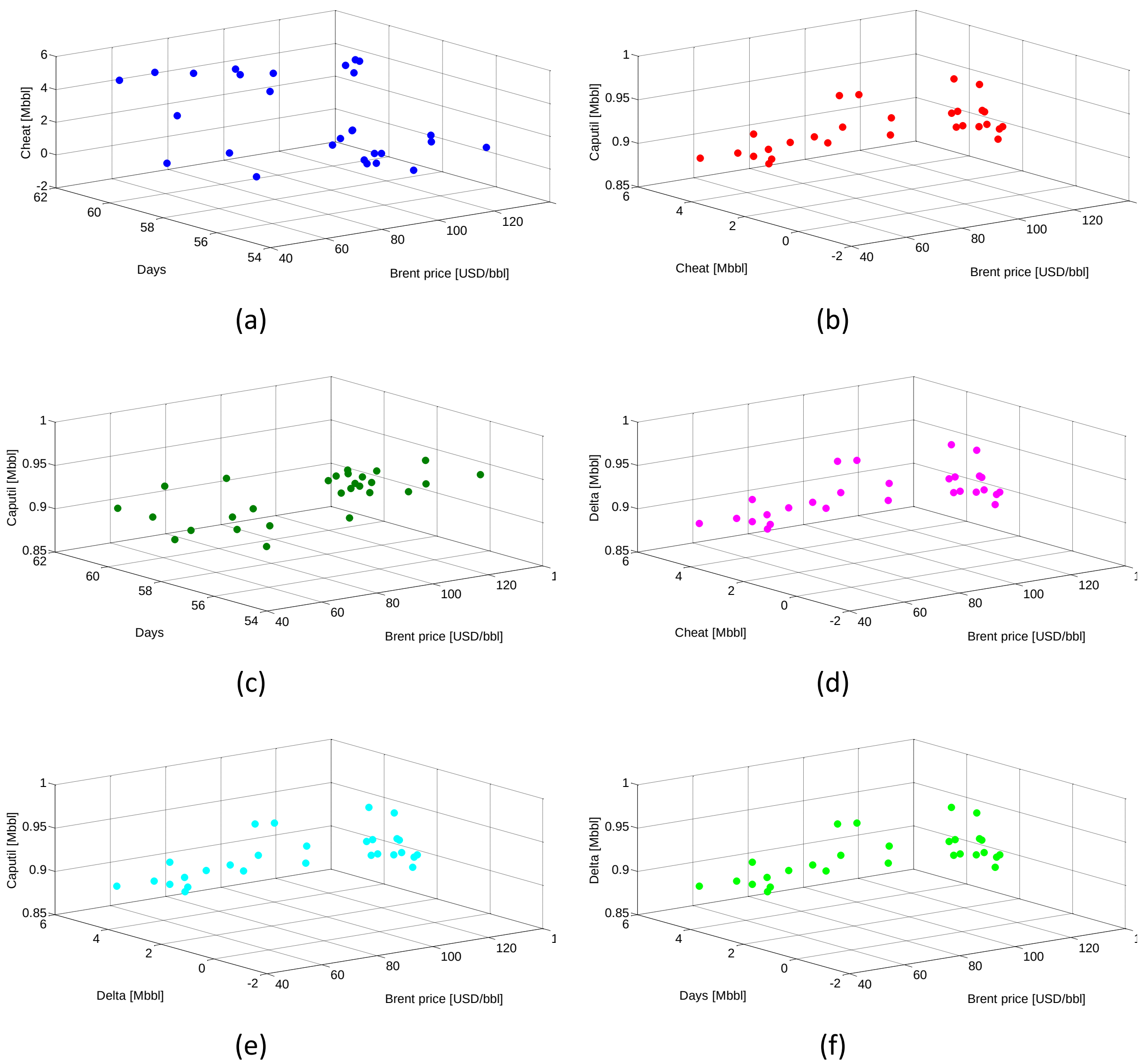


Figure 60 - 3D plots of (a) Brent price, Days, and Cheat; (b) Brent price, Cheat, and Caputil; (c) Brent price, Days, and Caputil; (d) Brent price, Cheat, and Delta; (e) Brent price, Delta, and Caputil; (f) Brent price, Days, and Delta.

Price and input data have a quarterly (*i.e.* seasonal) granularity, because this is the frequency of political decision about OPEC quotas and availability of supply and demand variables. The quarterly data of Brent and WTI prices come from the US Energy Administration databank (www.eia.gov/forecast/steo/query). OPEC production capacity, OPEC production, OECD commercial inventories are taken from EIA (2015), while demand data come from IEA databank (http://www.iea.org/statistics/relatedsurveys/monthlyoildatasurvey/). OPEC quotas are more difficult to be found since January 2009 (Laherrère, 2011) and are taken by cross-checking data from both OPEC Annual (OPEC, 2015) and EIA (2015) reports. The identification of the adaptive parameters α_i is performed by using as input data the values of the real variables from the first quarter of 2013 (*i.e.* Feb, 2013) to the first quarter of 2015 (*i.e.* Feb, 2015). The choice of the last eight quarters is due to the most suitable length of forecast horizon as it is discussed in Section 5.2.6. As time changes, the adaptive parameters need to be re-estimated to improve accuracy and better follow the last trend of CO prices in that particular historical period. Several one-step-ahead simulations were performed in *Matlab*©, from the first quarter of 2010. Every simulation covers eight quarters, starts one-quarter forward respect to the previous one, and takes the real value of the current input variables. Figure 61 and Figure 62 show the actual and fitted series of CO price over the simulation period (*i.e.* Feb, 2013-Feb, 2015) and indicate that the model reproduces past developments in CO markets satisfactorily.

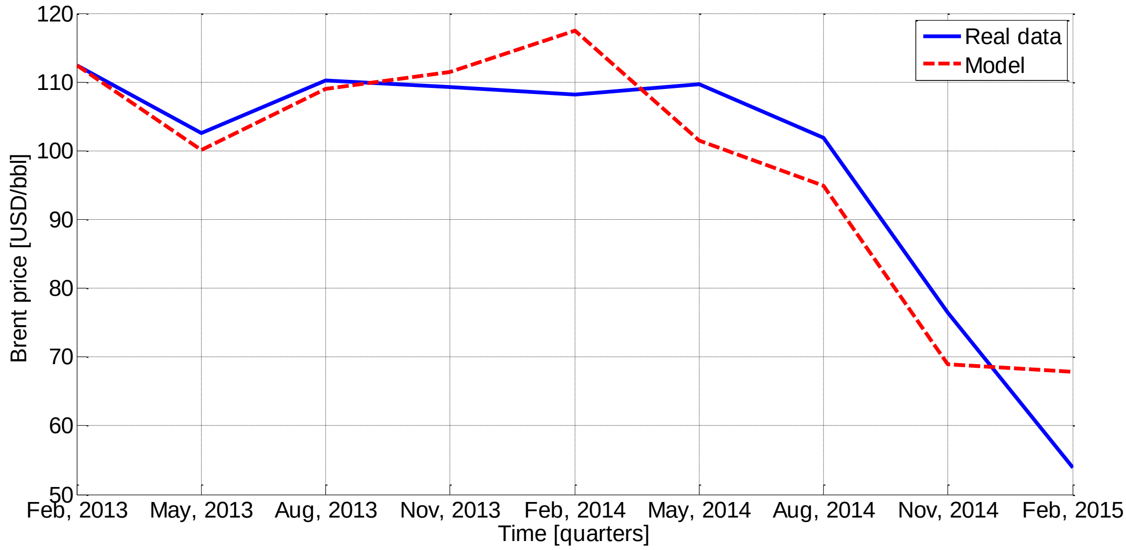


Figure 61 - One-step-ahead simulation of Brent quarterly prices from Feb, 2013 to Feb, 2015 (real data from EIA).

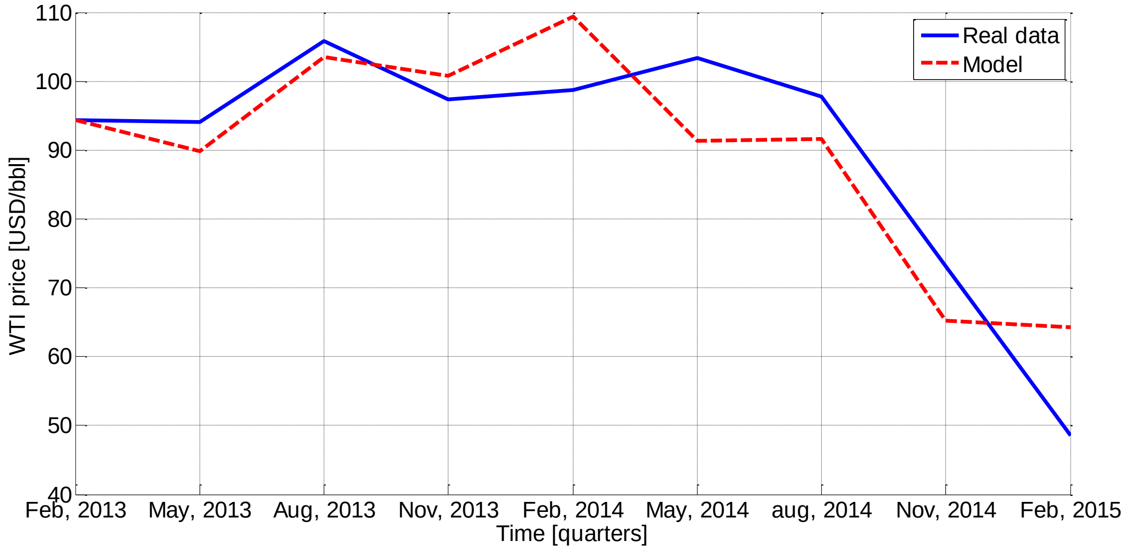


Figure 62 - One-step-ahead simulation of WTI quarterly prices from Feb, 2013 to Feb, 2015 (real data from EIA).

Figure 63, Figure 64, Figure 65, Figure 66, Figure 67, and Figure 68 show the trend of the estimated parameters of each simulation for Brent model. For the sake of conciseness, Appendix A reports the model trends for WTI. Table 10 shows the correlation between real and model values for each simulation. It is worth observing that the values are higher enough to conclude that the model is appropriate for PSE purposes.

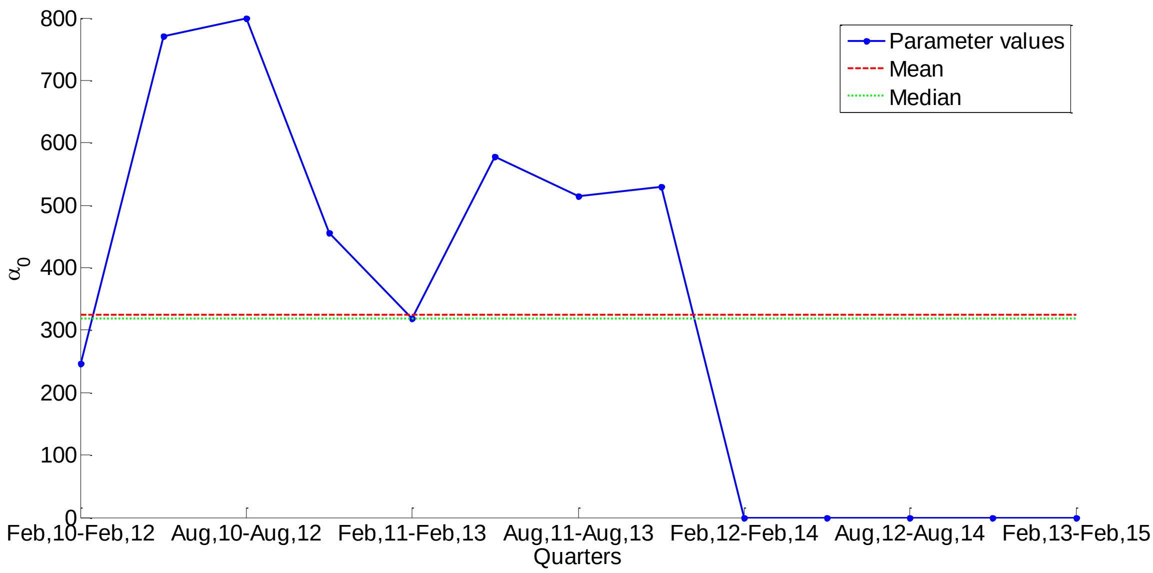


Figure 63 - α_0 values for each eight quarter-long simulation.

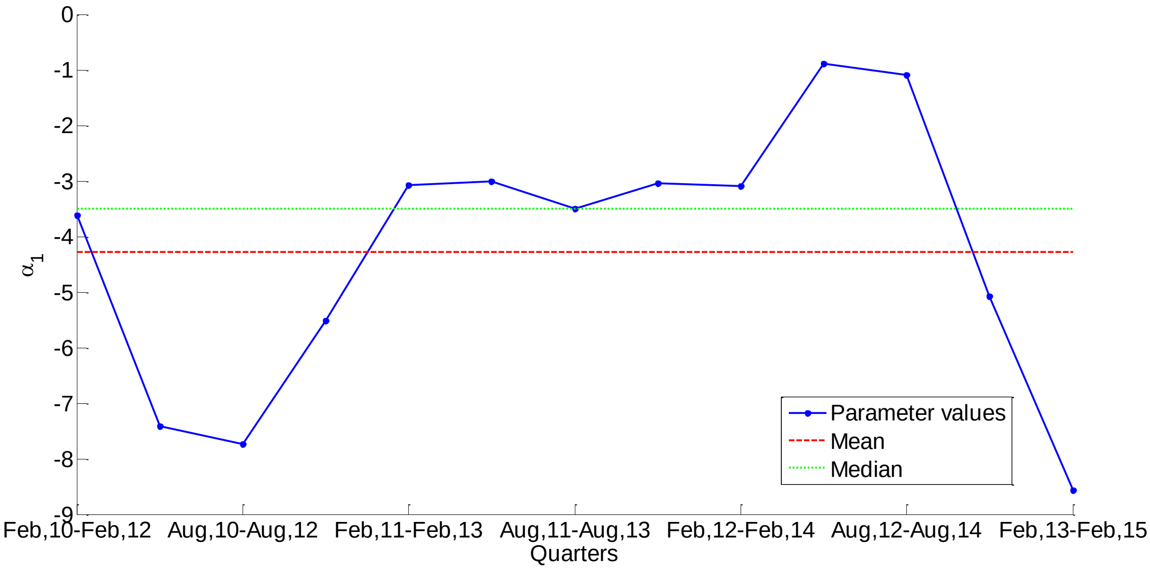


Figure 64 - α_1 values for each eight quarter-long simulation.

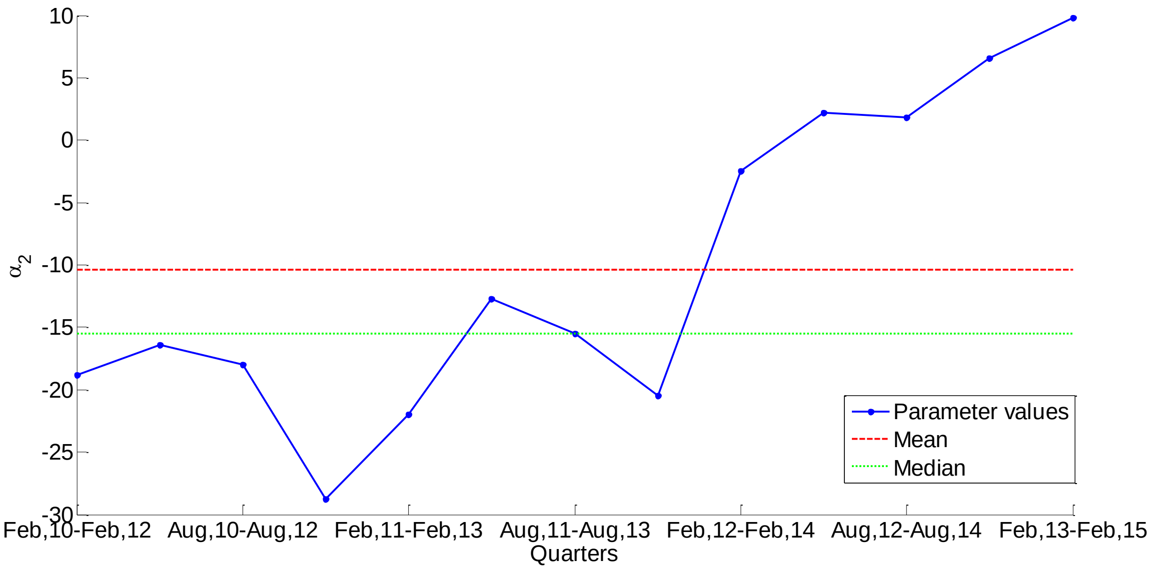


Figure 65 - α_2 values for each eight quarter-long simulation.

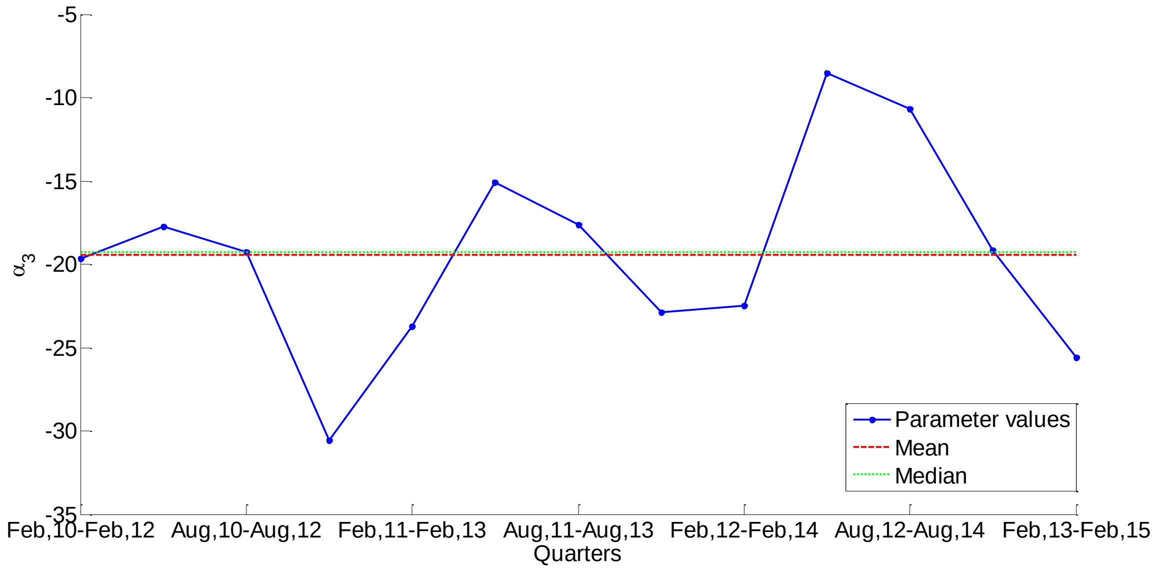


Figure 66 - α_3 values for each eight quarter-long simulation.

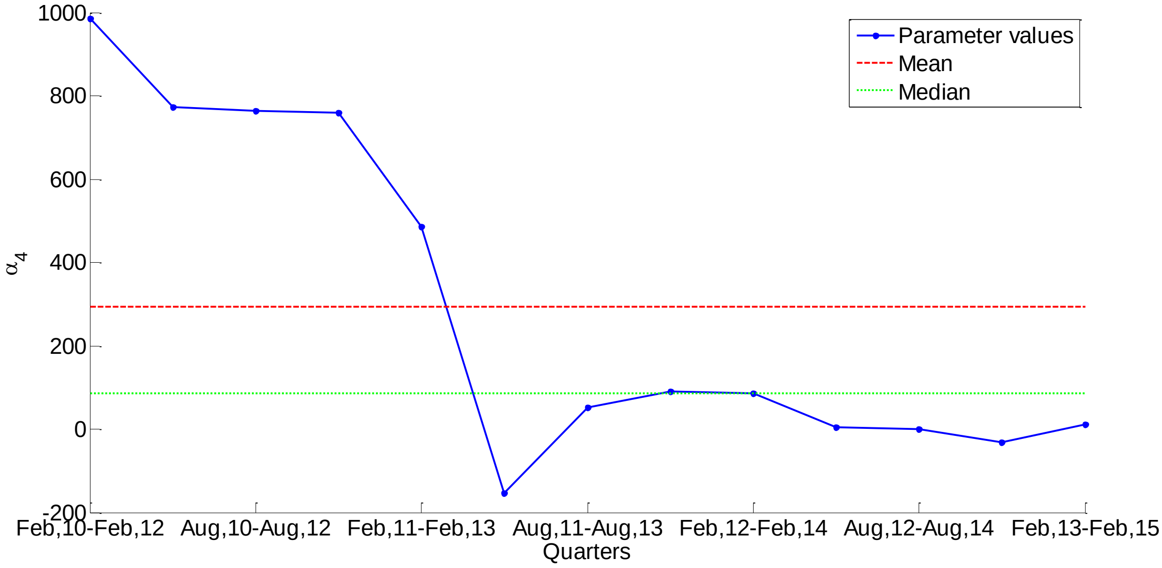


Figure 67 - α_4 values for each eight quarter-long simulation.

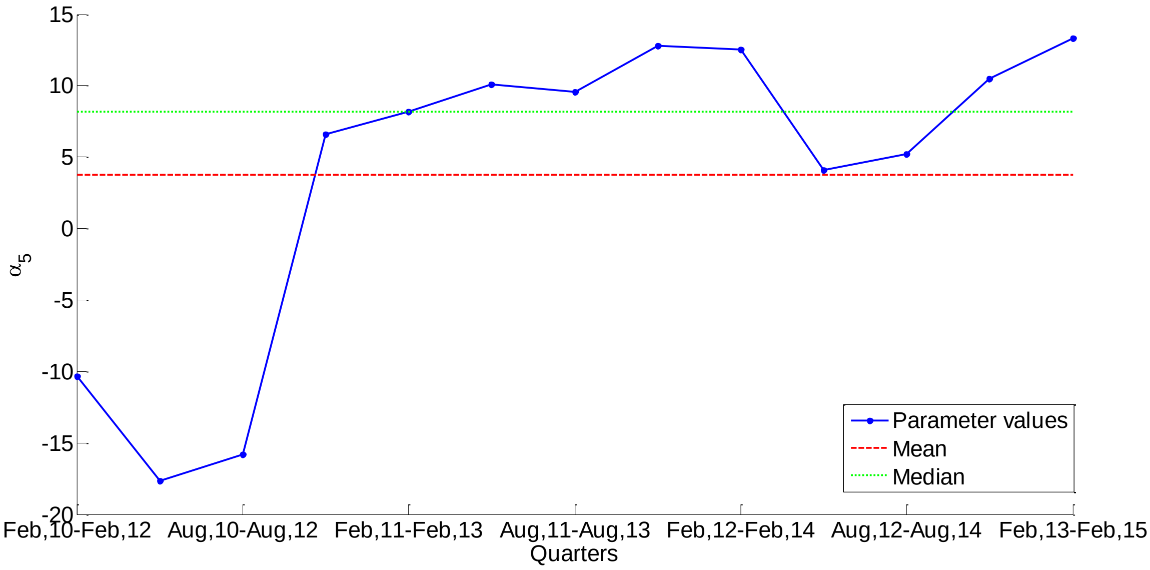


Figure 68 - $\alpha_5$values for each eight quarter-long simulation.

Table 10 - Correlation coefficients between real and model prices of Brent and WTI.

Period	Brent	WTI
Feb,10-Feb,12	0.970236	0.952081
May,10-May,12	0.989026	0.99435
Aug,10-Aug,12	0.986477	0.977481
Nov,10-Nov,12	0.968168	0.760532
Feb,11-Feb,13	0.900819	0.713259
May,11-May,13	0.939644	0.845981
Aug,11-Aug,13	0.916921	0.991109
Nov,11-Nov,13	0.91111	0.981163
Feb,12-Feb,14	0.91437	0.872074
May,12-May,14	0.496166	0.846983
Aug,12-Aug,14	0.760845	0.781139
Nov,12-Nov,14	0.936489	0.859042
Feb,13-Feb,15	0.92363	0.866048

Table 11 reports the identified α_i values of both Brent and WTI quotations as of data from the first quarter of 2013 to the first quarter of 2015. The signs of the estimated parameters are in part consistent with previous results described by Dees *et al.* (2007) with regard to $Days$, $Caputil$, and $Cheat$ variables in WTI equation.

Table 11 - Adaptive parameters in Equation (40) for the model of WTI and Brent quotations.

Parameter	Value	
	Brent	WTI
α_0	0	0
α_1	-8.57475	-9.07955
α_2	9.807808	2.278991
α_3	-25.5884	-21.2083
α_4	10.51636	305.3278
α_5	13.32928	11.93709

The regression coefficient associated with $Days$ is negative because an increase in stocks reduces real oil price by diminishing reliance on current production. Similarly, an increase in $Production^{OPEC}$ relative to their quotas increases the supply whilst the price decreases. The sign of the regression coefficient associated with $Caputil$ is positive, because increases in capacity utilization tend to increase prices. The sign of the regression coefficient associated with $Delta$ is positive, because a decrease in the difference between OPEC and the USA production causes the decrease of the CO quotations, as we have seen in recent quarters. The positive sign of the coefficient associated with the variable $Quota$ seems anomalous, as a decrease in allocations should increase CO prices, and vice versa. The positive sign is due to the recent market trend and OPEC members' decision of not cutting their quotas at the end of 2014 to maintain their production competitive with the global CO market, instead of facilitating price recovery. At the same time, it is worth observing the different signs and values of Brent parameters respect to WTI ones. This is mainly due to the dissimilar evolution of Brent and WTI quotations subject to separate and rather different dynamics produced by the events occurred in recent years (see Section 1.1). As it is reported in Appendix A, the accuracy of one-step-ahead simulations is good and the forecast errors produced by the model are small.

5.2.3 Input variable models

According to Equation (40) the input variables that need to be modeled and forecast for the PSE/CAPE purposes of this thesis are OECD demand, OECD inventories, OPEC production, OPEC production capacity, USA production, and OPEC quota. Table 12 summarizes the *physical* variables that are either directly or indirectly involved in the new OPEC-based model, their definitions, and their historical ranges.

Table 12 - Input variables involved in Equations (40-44).

Variable	Definition	Range	Source
OECD inventory [Mbbl] $Inventory^{OECD}$	Actual OECD reserves of unrefined petroleum.	2550 – 2799	Ye *et al.* (2005) EIA databank IEA databank
OECD demand [Mbbl/d] $Demand^{OECD}$	Actual OECD demand of CO.	44.5 – 48	EIA databank IEA databank
Production capacity [Mbbl/d] $Capacity^{OPEC}$	Total CO production capacity.	31.49 – 34.06	EIA databank IEA databank
OPEC production [Mbbl/d] $Production^{OPEC}$	Actual OPEC CO production.	28.98 – 31.74	EIA databank IEA databank
USA production [Mbbl/d] $Production^{USA}$	Actual USA CO production (including light tight shale oil).	4.73 – 9.26	EIA databank
OPEC quota [Mbbl/d] $Quotas$	Oil production allocations, also called *ceilings*.	24.84 – 30	EIA databank IEA databank

Although oil demand conditions are correctly modeled in Cooper (2003), the supply modeling is extremely difficult as oil markets reflect and translate the complex background of production conditions and OPEC behavior (Dees *et al.*, 2007). According to PSE/CAPE applications, it is recommended to customize the different models by avoiding undesired dependencies and over-parameterizations. However, the scientific literature does not report any simple forecasting models for inventories, production, and capacity utilization by OPEC.

Input data are taken from EIA (2015) and IEA (2015). The time interval chosen to identify both models is from the first quarter of 2013 (*i.e.* Feb, 2013) to first quarter of 2015 (*i.e.* Feb, 2015) for the same reasons that were discussed for price model. That interval is chosen rather short to be as close as possible to the current situation, but long enough to improve the accuracy.

Figure 69, Figure 71, and Figure 72 show respectively the trend of demand, OPEC production, OECD inventories, OPEC production capacity, and OPEC quota. The trend of the USA production and its comparison with Saudi Arabia one have already been discussed in Chapter 1 (see Figure 8). As for demand, the greatest difference between two following quarters equals 4.3%, while the overall variations is 7.3%. The inventory percentages are similar. In particular, the overall variation is 8.1%, while the maximum variation between adjacent time periods is 5%. OPEC production has shown the higher variations, with an overall percentage of 12% and a local variation of 7%.

Figure 70 represents the comparison between normalized OECD demand and OPEC production, which are calculated by dividing the actual value by 48.5 Mbbl/d and 33 Mbbl/d respectively (*i.e.* values that are slightly larger than the upper limits), in order to obtain values that are between 0 and 1. It is worth observing that OPEC CO production follows mostly demand trend in order to serve consumers' needs, except for the period between February 2012 and February 2013, when the production capacity and OPEC production increased and determined CO price instability. As for the OPEC production and production capacity, it is worth observing that OPEC production is always lower than its capacity. Hence, OPEC members produce less CO than the quantity they can manufacture, but more than their production quotas. Instead, inventories do not show a significant trend, except for reaching their highest value in the last quarter (*i.e.* Feb, 2015).

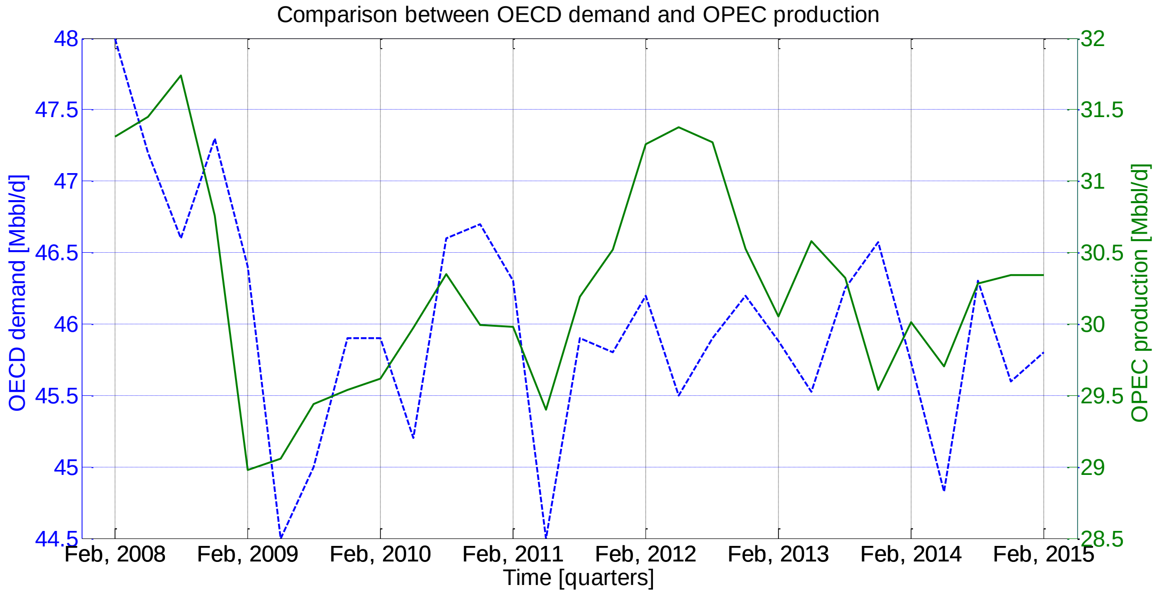


Figure 69 - OECD demand and OPEC production from the first quarter of 2008 to the first quarter of 2015 (data from EIA).

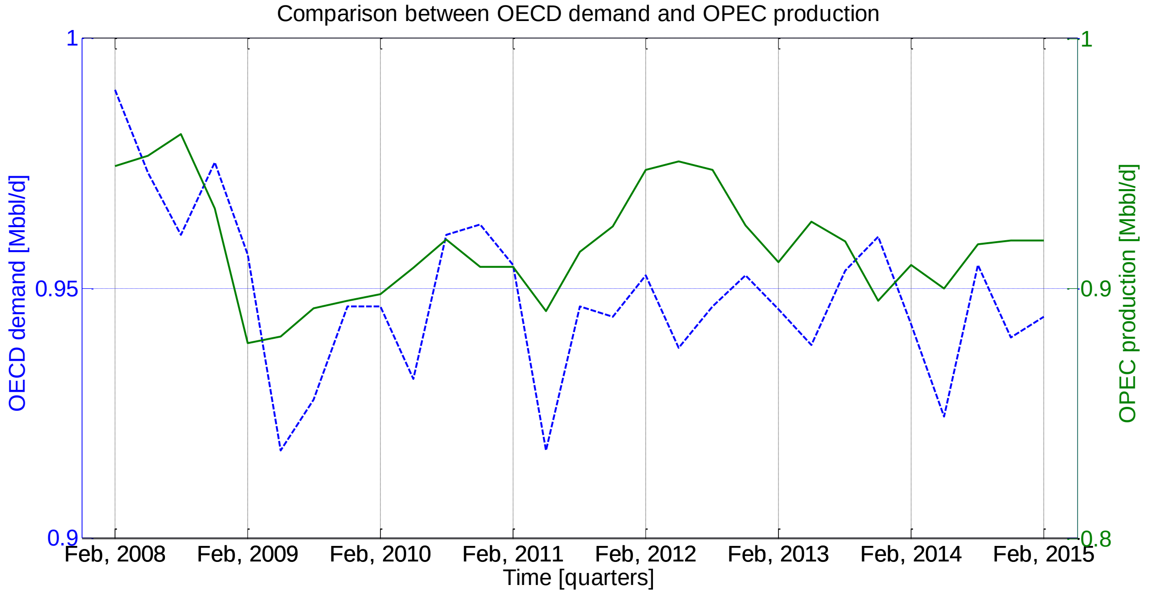


Figure 70 - Normalized OECD demand and normalized OPEC production from the first quarter of 2008 to the first quarter of 2015.

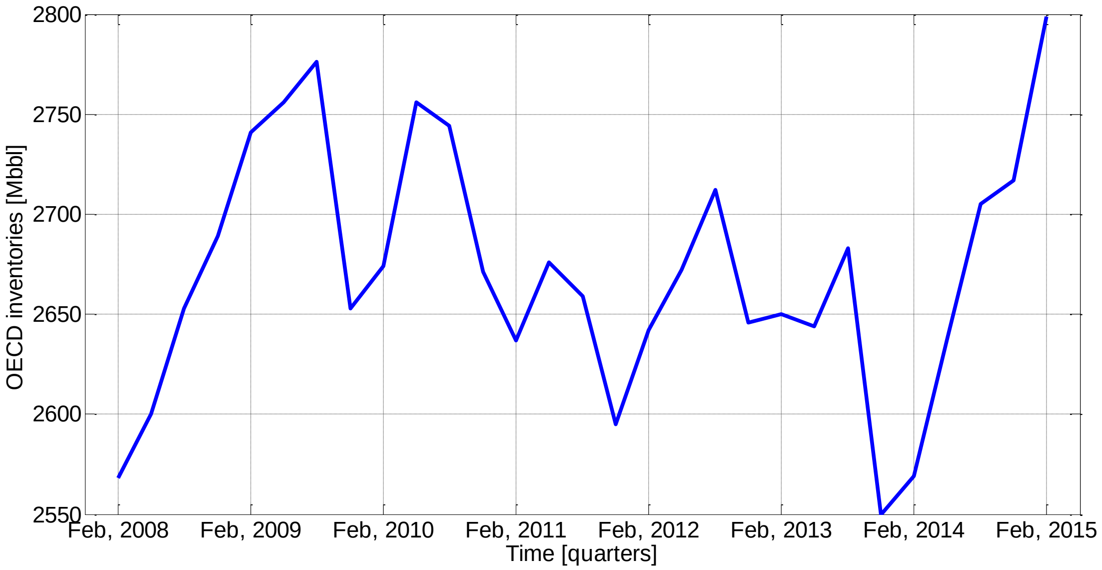


Figure 71 - OECD inventories from the first quarter of 2008 to the first quarter of 2015 (data from EIA).

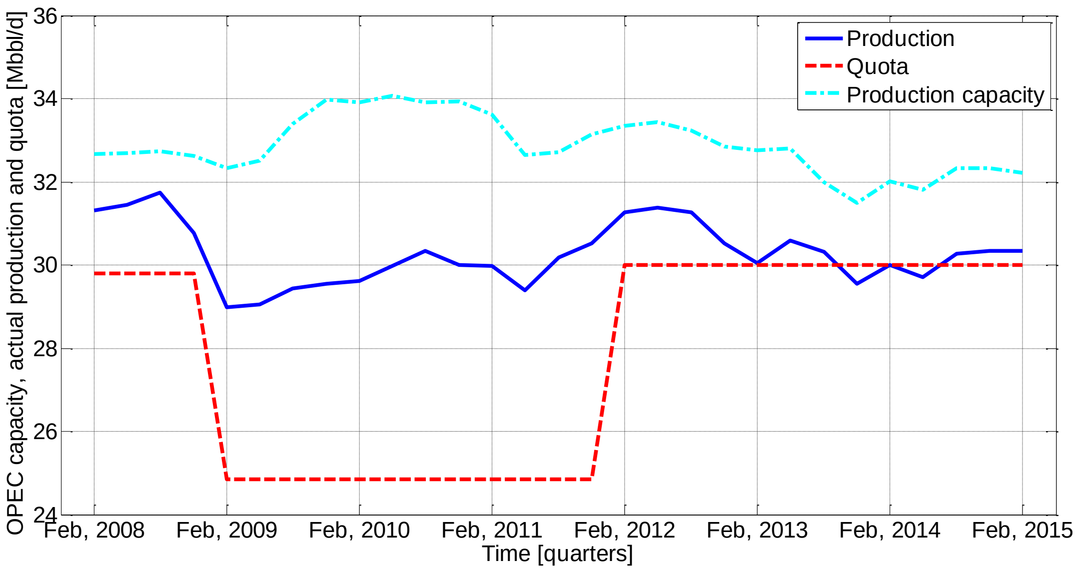


Figure 72 - OPEC production, OPEC production capacity, and OPEC quota from the first quarter of 2008 to the first quarter of 2015 (data from EIA).

Table 13 summarizes the correlation coefficients between real and model values for each eight quarter-long simulation starting from the first quarter of 2010. The models are described in next Sections, but it is worth observing that the models fit well the input variables and price trends. Appendix A reports further simulations of the input variables, some of which are more significant than the last ones.

Table 13 - Correlation coefficients between real input and model values for Equations (45-46-50-51-54).

Period	Capacity	Demand	Inventory	OPEC Production	USA Production
Feb,10-Feb,12	0.588297	0.369269	0.679217	0.433089	0.975956
May,10-May,12	0.548701	0.640437	0.763727	0.438179	0.977628
Aug,10-Aug,12	0.404001	0.620156	0.708188	0.431371	0.987027
Nov,10-Nov,12	0.63057	0.528194	0.4763	0.428827	0.979473
Feb,11-Feb,13	0.553888	0.437743	0.560825	0.403239	0.987392
May,11-May,13	0.458227	0.823319	0.665777	0.729914	0.984283
Aug,11-Aug,13	0.214066	0.380912	0.599055	0.673775	0.98596
Nov,11-Nov,13	0.644365	0.296693	0.867109	0.745269	0.991938
Feb,12-Feb,14	0.776942	0.34588	0.895032	0.894638	0.99545
May,12-May,14	0.780448	0.73494	0.911702	0.827044	0.99798
Aug,12-Aug,14	0.680294	0.655205	0.580236	0.858015	0.994492
Nov,12-Nov,14	0.636794	0.68148	0.41185	0.770356	0.993342
Feb,13-Feb,15	0.642385	0.309231	0.413727	0.644923	0.992501

As the following Section report and comment, the model curves do not follow the historical data exactly. However, such curves can be considered rather good as the variables involved are not easy to be modeled and forecast, because they are functions of the economic activities of the countries (either producers or consumers).

5.2.3.1 OECD demand

Dees *et al.* (2007) proposed an economic model also for the demand, which is intended to assess CO market developments. In particular, oil demand is explained by a *behavioral equation* that relates demand to domestic activity, real price of CO, and a time trend that represents the technological developments linked to energy efficiency. As a result of that paper, an increase in demand can be considered to be followed by increases in both OPEC production and world CO price. According to Cooper (2003), oil demand is explained by the same dependence on domestic activity and real CO price.

In the proposed price model, OECD demand is symbolized in the $Days$ variable. The model to forecast OECD demand is adapted from Cooper (2003) and takes the following form:

$$Demand_{t+1}^{OECD} = \beta_0 GDP_{t+1} + \beta_1 Price_t + \beta_2 \tag{45}$$

with GDP being the global Gross Domestic Product of the corresponding quarter, which seems to be stable at around 2% (in term of year-on-year prices), and $Price$ the CO price. For the sake of simplicity, compared to the model of Cooper, Equation (45) uses the CO price of the previous quarter and does consider neither a random error term nor the contribution of the Boolean variables. These variables were introduced in Manca *et al.* (2015) to describe the impact of the different quarters starting from the same time period of demand and GDP and improve the modeling of demand periodicity. Indeed, as shown in Figure 69, demand seems periodic, even if some incidents may impact the periodicity. Therefore, the values of the adaptive coefficients that are associated with the quarterly Boolean variables are almost equal (Figure 73). Hence, the quarterly component can be excluded from the demand model, which simply substitutes the Boolean variables with β_2. Moreover, the demand model without Boolean elements has a greater accuracy than the model with the Boolean variables. For instance, the demand model in Equation (45) gives a CO price of 99.57 USD/bbl in Aug, 2014, while the model with the Boolean variables would provide a result of 86.9 USD/bbl (against the real Brent quotation that was 101.9 USD/bbl).

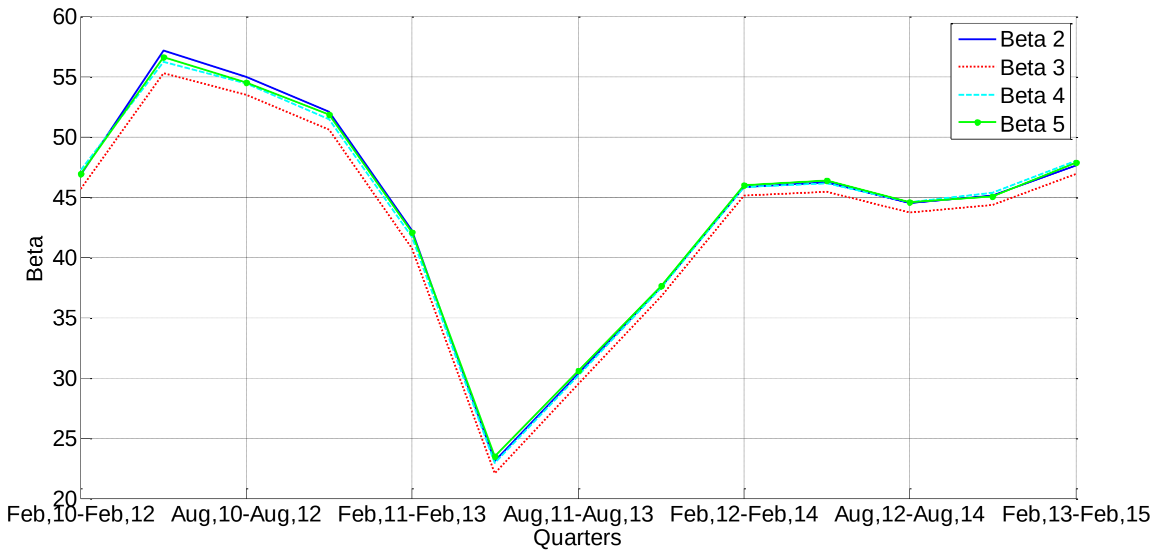


Figure 73 - Comparison of the Boolean variables adaptive coefficients for demand model.

Table 14 contains the values of the adaptive parameters in Equation (45). Figure 76 represents the one-step-ahead simulation results for OECD Brent demand from the first quarter of 2013 (*i.e.* Feb, 2013) to the first quarter of 2015 (*i.e.* Feb, 2015). The model seems to reach the average trend of the involved variable, as the greatest variation in the last time interval was 1.8%. Figure 74 and Figure 75 collect the results of the two previous simulations.

Table 14 - Adaptive parameters in Equation (45) for the demand model.

Parameter	Value	
	Brent	WTI
$\boldsymbol{\beta_0}$	-3.04e-14	-1.25e-14
$\boldsymbol{\beta_1}$	-0.02242	0.002732
$\boldsymbol{\beta_2}$	50.76692	46.64156

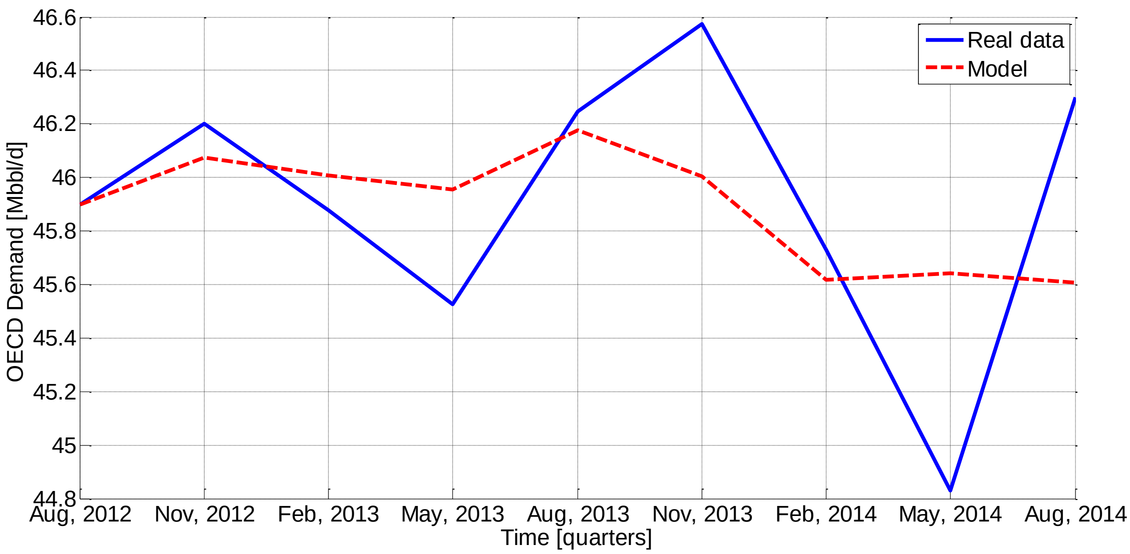


Figure 74 - One-step-ahead simulation of quarterly OECD crude oil demand from Aug, 2012 to Aug, 2014 (real data from EIA).

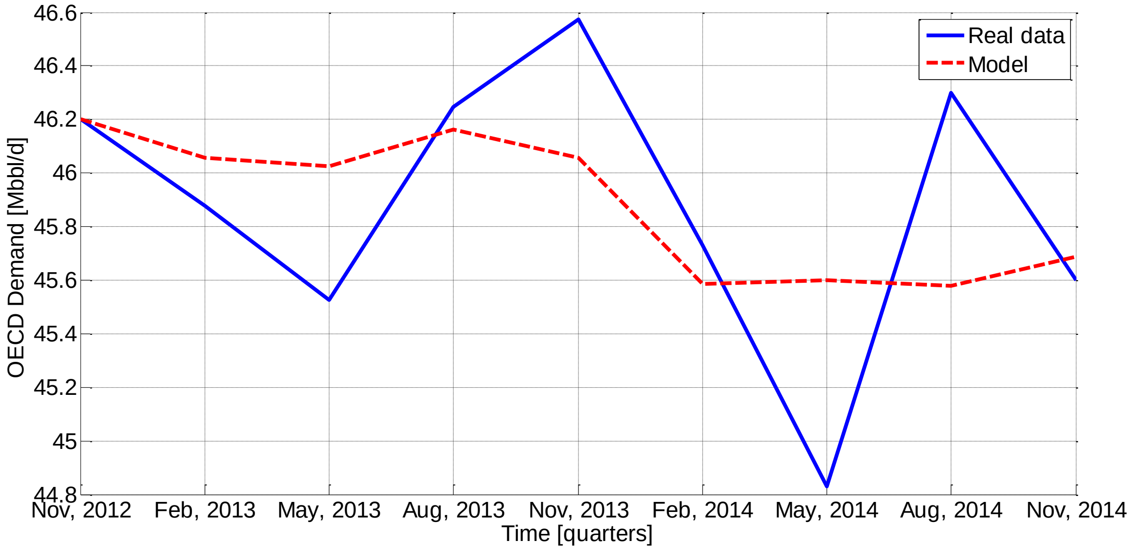


Figure 75 - One-step-ahead simulation of quarterly OECD crude oil demand from Nov, 2012 to Nov, 2014 (real data from EIA).

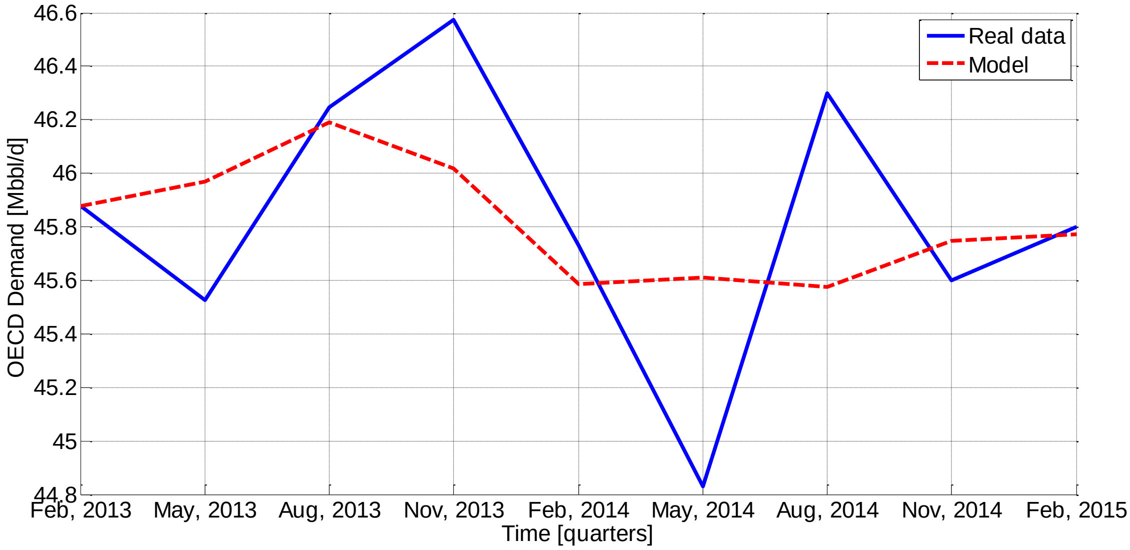


Figure 76 - One-step-ahead simulation of quarterly OECD crude oil demand from Feb, 2013 to Feb, 2015 (real data from EIA).

The effect of yearly global GDP is interesting since it can create different scenarios according to the degree of economic activity in a country, region, and in the world. Different simulations based on the last collected value of GDP (*i.e.* 2013), a constant annual increase of 2%, and a constant annual decrease of 2% are reported in Section 5.4.

5.2.3.2 OECD inventories

Inventories are reserves of unrefined petroleum that are measured in number of barrels. Oil producers use CO stockpiles to smooth out the impact of changes in supply-and-demand, and prevent unforeseen circumstances that could cut down oil supply. Figure 77 explains the meaning of this variable.

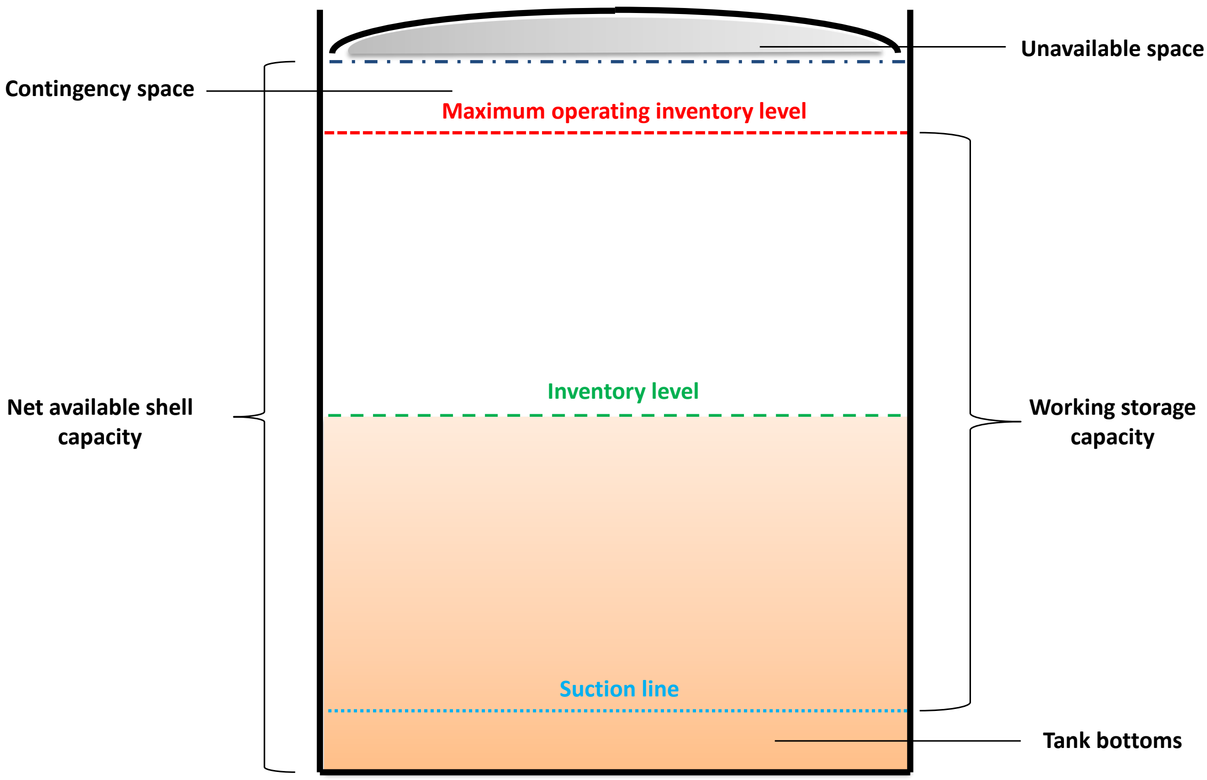


Figure 77 - Schematic representation of the inventory concept.

Inventory levels are affected by OPEC's production decisions, political events, tax policy changes, and other factors (*e.g.* tenses on both supply-and-demand sides). At the same time, inventory levels affect the price of oil with higher inventories leading to lower prices (Kaufmann *et al.*, 2004; Dees *et al.*, 2007). As a matter of fact, the relationship between inventories and prices is no longer simple, being a function of production behavior. In particular, causality between these variables runs in both directions, *i.e.* inventories may rise or fall as a function of countries' income for providing CO or through investing capitals on refining infrastructure. Dvir and Rogoff (2014) predicted how inventories and price behave according to income and production conditions. Specifically, in periods when production is flexible, *i.e.* when a rise in income, which raises demand, is predicted to result in a commensurate rise in production, inventories should fall, as the effect of high demand should dissipate quickly. This should help mitigate any rise in price associated with the surge in demand so that inventories and price should exhibit a negative relationship. Conversely, in periods when production is inflexible, *i.e.* when a rise in income is not predicted to significantly increase production, inventories should levitate. This would actually enhance the price increase, so that inventories and price should exhibit a positive relationship.

As already discussed in Chapter 4, Ye *et al.* (2009) investigated four potential inventory variables, *i.e.* total of industrial and governmental petroleum inventories, total of industrial and governmental CO inventories, industrial petroleum inventories, and industrial CO inventories. The present work adopts the commercial stocks available in the EIA databank as a measure of supply-and-demand balance, rather than total stocks (EIA, 2015). The reason is that total stocks, which include strategic petroleum reserves, could generate a misspecification of the equilibrium between supply and demand because of their strategic nature. EIA (2015) estimated that OECD commercial oil inventories totaled 2.75 Gbbl at the end of 2014, which is the highest end-of-year level on record and equivalent to roughly 60 days of consumption. Figure 71 shows the historical trend of commercial inventories from February 2008 to February 2015.

In the proposed price model, OECD inventories are accounted for by the $Days$ variable. OECD inventories are the result of the difference between the OECD demand and supply (Ye *et al.*, 2006). The efforts accomplished in this work on the study of inventories show that they are correlated with both OECD demand and OPEC production capacity. Consequently, it is advisable to adopt the following model to forecast the OECD inventories:

$$Inventory_{t+1}^{OECD} = \gamma_0 + \gamma_1 Capacity_t^{OPEC} + \gamma_2 Demand_{t+1}^{OECD} \quad (46)$$

Table 15 contains the values of the adaptive parameters of Equation (46). The global trend shown in Figure 78, Figure 79, and Figure 80 is respected and the overall model for prediction of CO prices exploits a positive contribution from the inventory term. The maximum error occurred in the last eight quarters is equal to 4.6%. Appendix A reports further one-step-ahead simulation results for this input variable and shows that the new inventory model is functional.

Table 15 - Adaptive parameters in Equation (46) for the inventory model.

Parameter	Value	
	Brent	WTI
γ_0	1641.698	1641.698
γ_1	52.47368	52.47368
γ_2	-14.6816	-14.6816

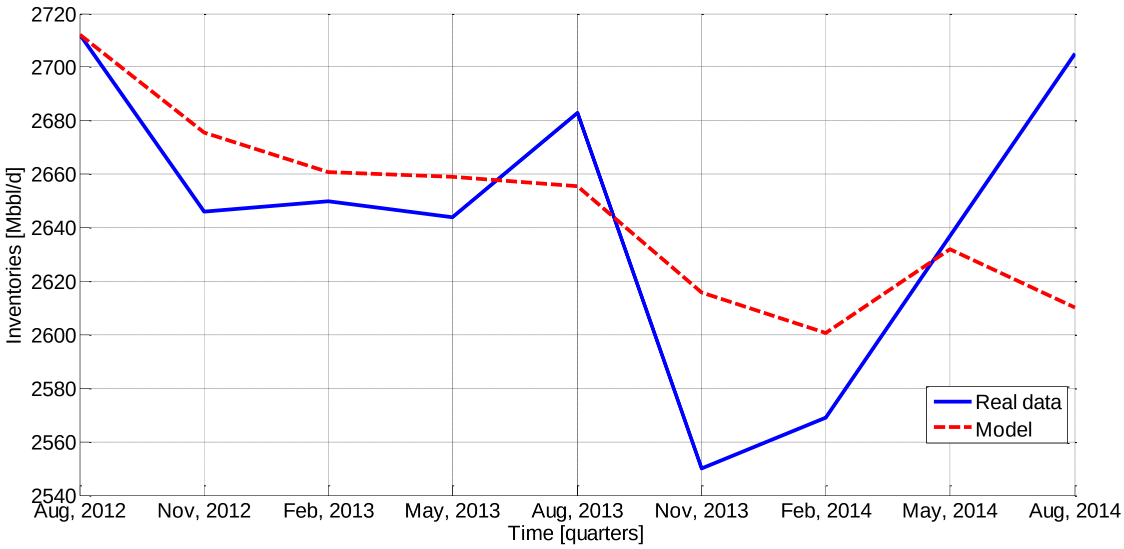


Figure 78 - One-step-ahead simulation of quarterly OECD crude oil inventories from Aug, 2012 to Aug, 2014 (real data from EIA).

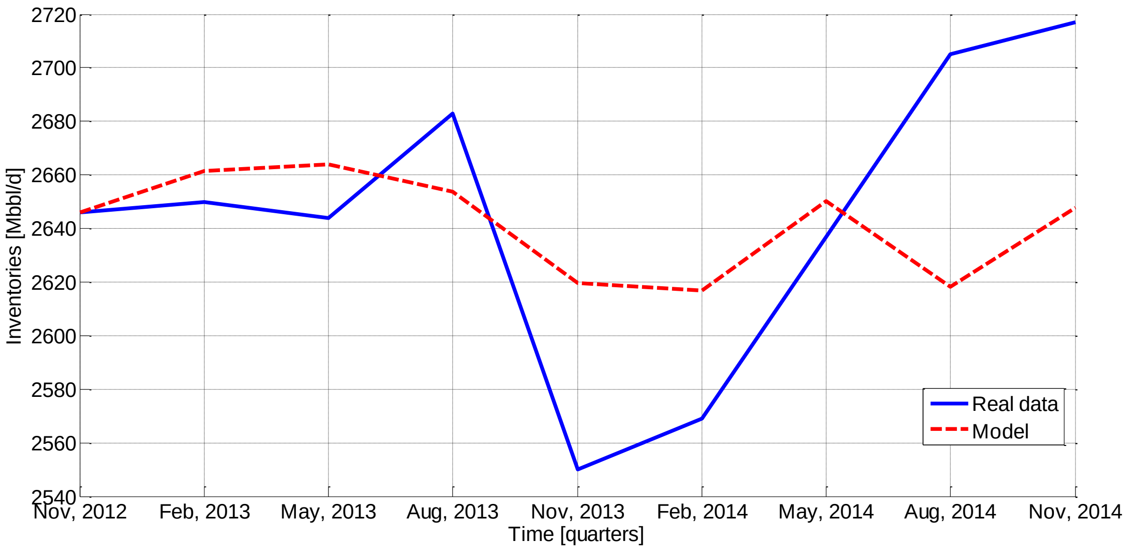


Figure 79 - One-step-ahead simulation of quarterly OECD crude oil inventories from Nov, 2012 to Nov, 2014 (real data from EIA).

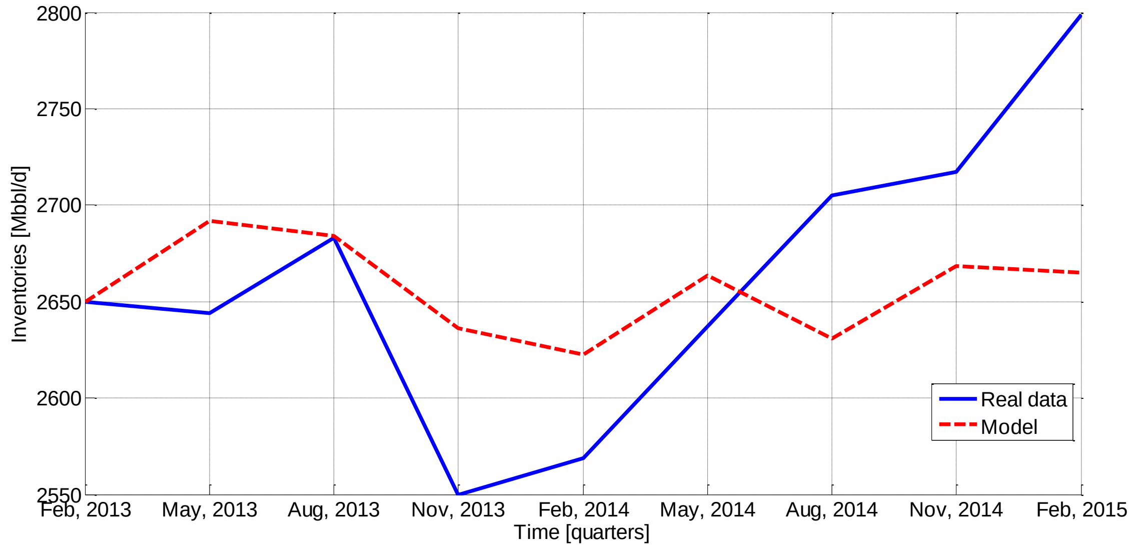


Figure 80 - One-step-ahead simulation of quarterly OECD crude oil inventories from Feb, 2013 to Feb, 2015 (real data from EIA).

5.2.3.3 OPEC production

According to EIA (2015), OPEC member countries produce about 40% of the global CO and represent about 60% of the total petroleum traded worldwide. Hence, OPEC CO production is an important factor that affects global oil prices. OPEC sets production targets for its members and generally, when OPEC production targets are reduced, oil prices increase. OPEC itself tries to regulate the production by means of compensations in case of an unintended production decrease or increase from a country member. Nevertheless, OPEC production depends on several ambiguous and unpredictable parameters, such as the local political situation in the different countries of the cartel.

Within the proposed price model, OPEC production appears in the $Caputil$, $Cheat$, and $Delta$ variables. The scientific literature reports several models for this input variable. For instance, Griffin (1985) used data from individual OPEC nations to estimate the following equation:

$$ln\ Q_{i,t} = \alpha_i + \gamma_i \cdot lnP_t + \beta_i \cdot lnQ_{i,t}^{00} + \varepsilon_{i,t} \tag{47}$$

where $Q_{i,t}$ is CO production by OPEC country i at time t, P is CO price, Q^{00} is CO production by OPEC nations other than country i, and ε is the error term.

According to the cooperative model (Kaufmann, 1995; Dees *et al.*, 2007), OPEC restrains production from existing capacity to match the demand for CO, which is equal to the difference between demand DEM and non-OPEC supply $PROD^{non-OPEC}$, providing the following relation:

$$PROD^{OPEC} = \sum_i DEM_i + \Delta Stocks^{OECD} - NGLS - \sum_j PROD_j^{non-OPEC} - PG \quad (48)$$

where $\Delta Stocks^{OECD}$ is the level of stocks reported by OECD countries, $NGLS$ is non-gas liquid supply, PG is net processing gains, i and j are the indexes for OPEC and OECD countries, respectively.

The existing models also adopt a competitive behavior for OPEC supply, which implies that OPEC countries compete not only among themselves but also with non-OPEC producers, leading OPEC to increase production to levels that are consistent with operable capacity. The cartel tries to produce less than its capacities allow so to be able to intervene in case of sudden variations in CO market, and to impact on CO prices. A very basic model representing the OPEC competitive behavior considers:

$$Production^{OPEC} = 0.95 \cdot Capacity^{OPEC} \quad (49)$$

For the sake of clarity, the model used in this paper depends on current OPEC capacity and on the previous price of CO:

$$Production_{t+1}^{OPEC} = \xi_0 + \xi_1 Capacity_{t+1}^{OPEC} + \xi_2 Price_t \quad (50)$$

Table 16 reports the values of the adaptive parameters in Equation (50).

Table 16 - Adaptive parameters in Equation (50) for the OPEC production model.

Parameter	Value	
	Brent	WTI
ξ_0	11.8352	13.37011
ξ_1	0.599866	0.537235
ξ_2	-0.00971	-0.00563

Figure 81, Figure 82, and Figure 83 show the results of one-step-ahead simulations over the analyzed quarters. The model of OPEC production capacity follows the trend of real data with a maximum error of 1% for the last eight quarters. Appendix A shows the graphs of other one-step-ahead simulations between 2010 and 2015.

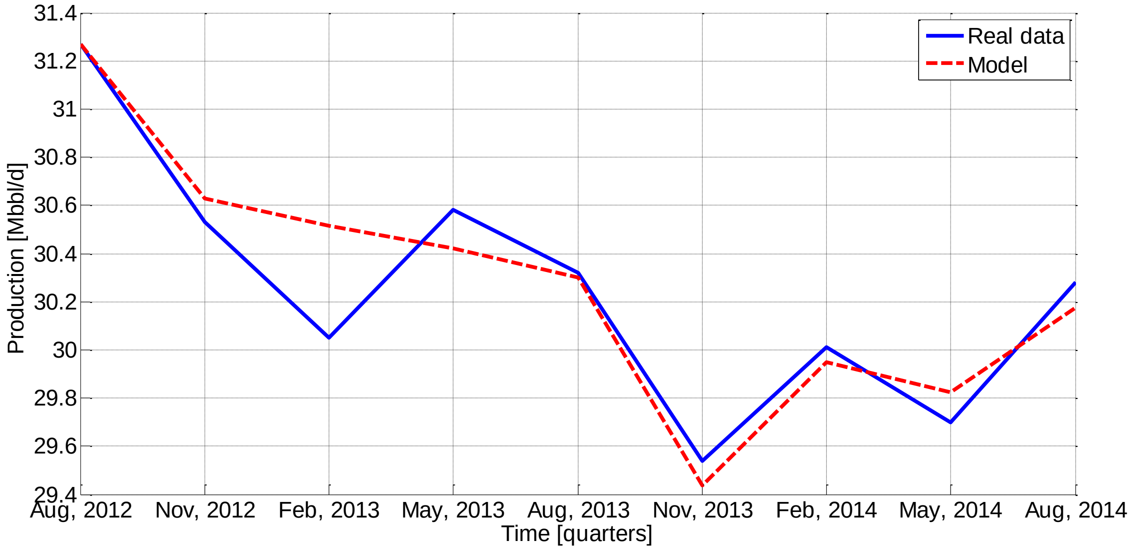


Figure 81 - One-step-ahead simulation of quarterly OPEC production from Aug, 2012 to Aug, 2014 (real data from EIA).

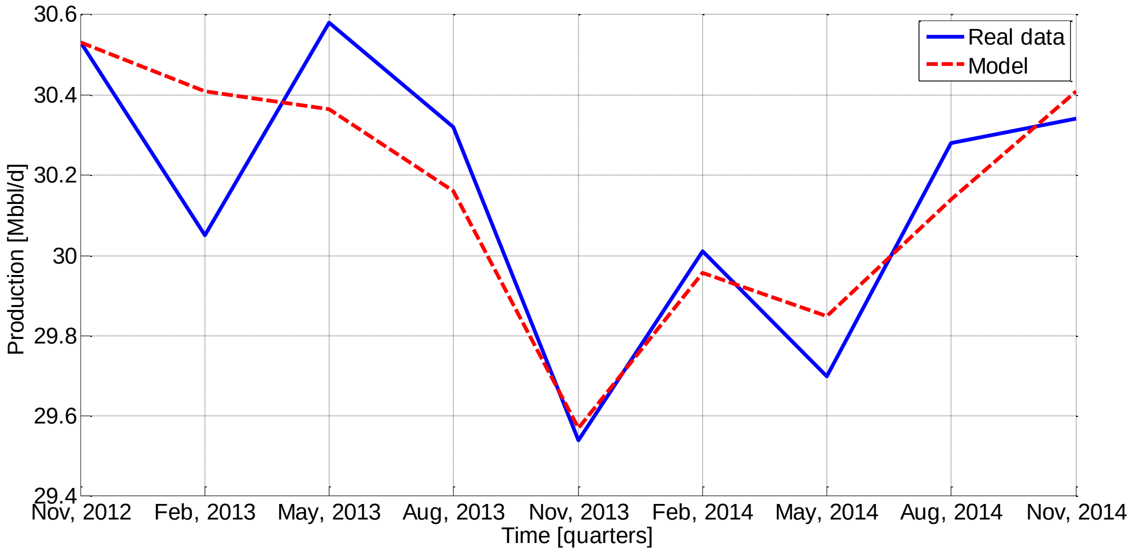


Figure 82 - One-step-ahead simulation of quarterly OPEC production from Nov, 2012 to Nov, 2014 (real data from EIA).

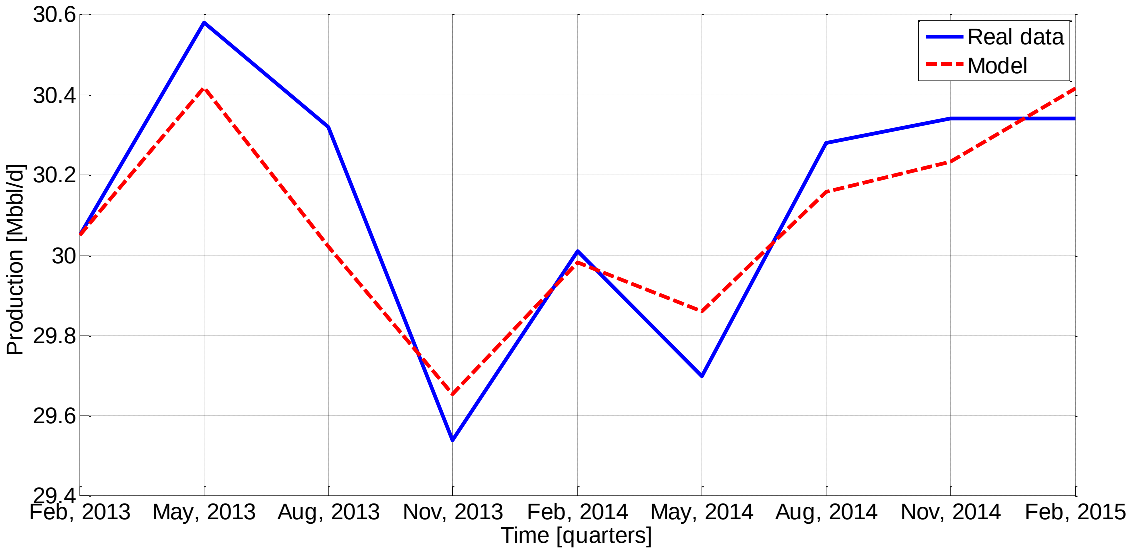


Figure 83 - One-step-ahead simulation of quarterly OPEC production from Feb, 2013 to Feb, 2015 (real data from EIA).

5.2.3.4 OPEC production capacity

The scientific literature and most important databanks (*e.g.*, EIA, IEA, ICIS) report different definitions of capacity as a function of its attributes. For example, EIA (2015) defines *spare capacity* as "*the volume of production that can be brought on within 30 days and sustained for at least 90 days*" and provides an indicator of the world oil market's ability to respond to potential crises that could reduce oil supplies. Saudi Arabia, which is the largest oil producer within OPEC and the world's largest oil exporter, historically has had the greatest spare capacity and has usually kept more than 1.5-2 Mbbl/d as a margin of flexibility on hand for market management. Instead, the *refining capacity* is defined as the maximum amount of CO that can be processed in a calendar year divided by the number of days in the corresponding year and characterizes how well the downstream sector is developed (Moebert, 2007).

In Equation (40), OPEC production capacity appears in the variable $Caputil$. The proposed capacity model appears simple but also rather accurate:

$$Capacity_{t+1}^{OPEC} = \varepsilon_0 + \varepsilon_1 Capacity_t^{OPEC} + \varepsilon_2 Production_t^{OPEC} \qquad (51)$$

Table 17 contains the values of the adaptive parameters in Equation (51). As shown in Figure 84, Figure 85, and Figure 86 the trend of the capacity model is close to the real one. The model of OPEC production capacity follows the trend of real data with a maximum error of 1%. Appendix A shows further graphs of one-step-ahead simulations between 2010 and 2015.

Table 17 - Adaptive parameters in Equation (51) for the production capacity model.

Parameter	Value	
	Brent	WTI
ε_0	30.42316	30.42316
ε_1	1.241432	1.241432
ε_2	-1.26974	-1.26974

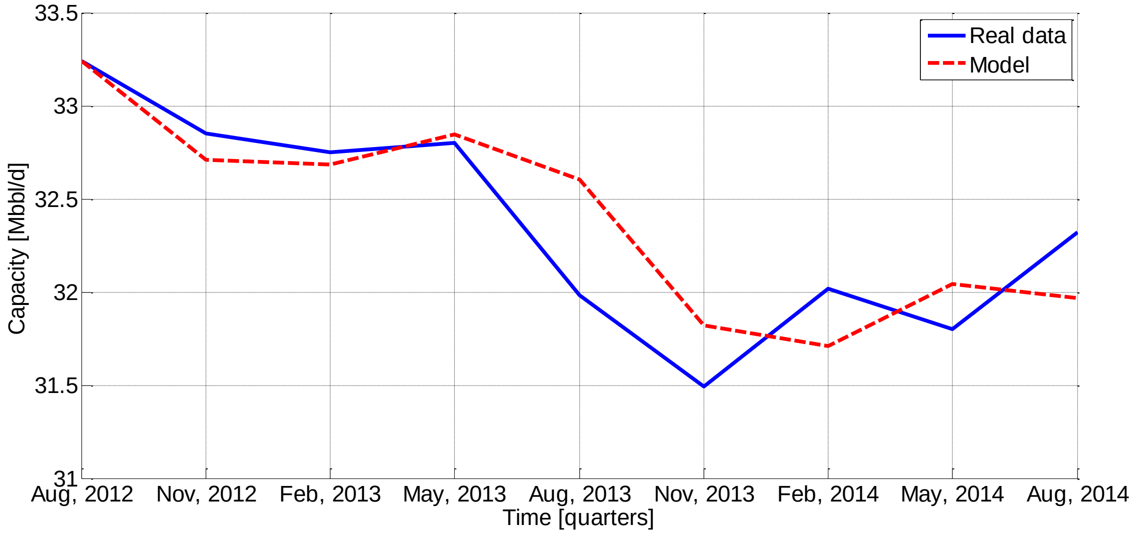


Figure 84 - One-step-ahead simulation of quarterly OPEC production capacity from Aug, 2012 to Aug, 2014 (real data from EIA).

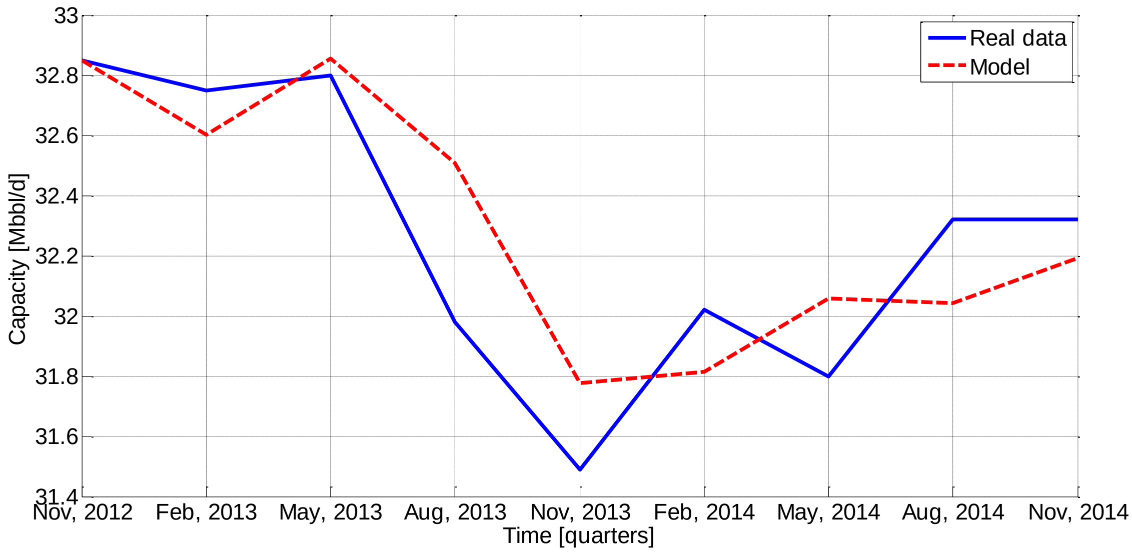


Figure 85 - One-step-ahead simulation of quarterly OPEC production capacity from Nov, 2012 to Feb, 2014 (real data from EIA).

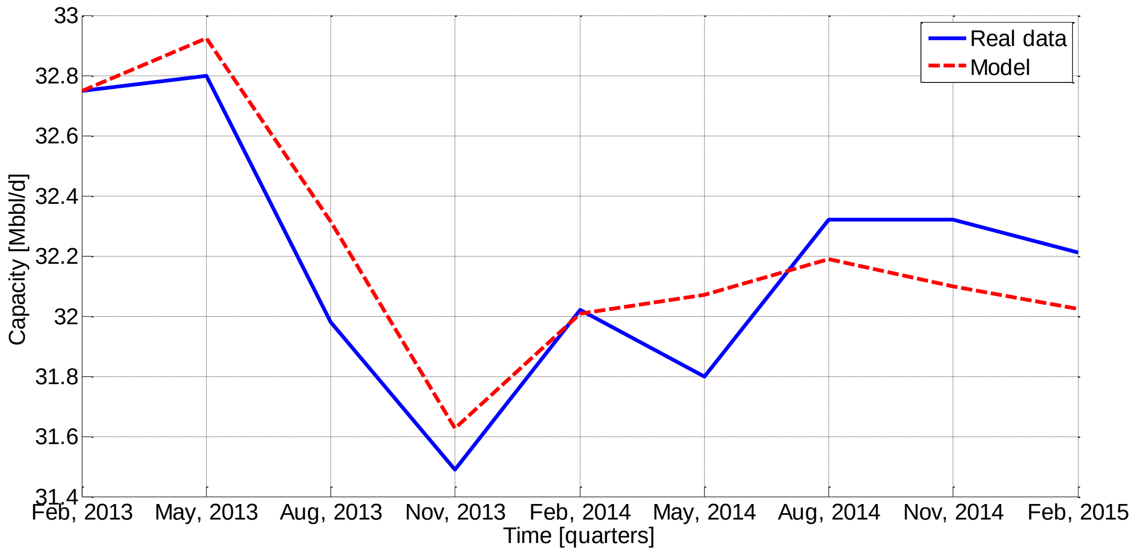


Figure 86 - One-step-ahead simulation of quarterly OPEC production capacity from Feb, 2013 to Feb, 2015 (real data from EIA).

5.2.3.5 USA production

Economic models of non-OPEC production generally have proved unreliable for PSE/CAPE applications, because of the lack of a simple relation between CO prices and production. Indeed, to model non-OPEC production, a complex interrelation among resource depletion, technical changes, economic incentives, and political considerations has to be taken into account (Dees *et al.*, 2007). The relation among these forces was modeled in Kaufmann (1995) by combining the curve fitting technique developed by Hubbert (1962) with the econometric model of supply that was proposed by Fisher (1981). The Hubbert theory, also known as Peak Oil Theory, is extensively discussed in Bardi and Yaxley (2005), Bardi (2009), Al-Bisharah *et al.* (2009), Bardi and Lavacchi (2009), Nashawi *et al.* (2010), Murphy and Hall (2011), and Miller and Sorrell (2013), and also presented in Section 2.1. According to Hubbert (1962), the production curve is bell-shaped and approximately symmetric. His theory was verified with good approximation for the case of oil production in the United States that peaked in 1971. There are several other regions of the world where oil production followed a single shaped curve (*e.g.* Brazil, Kazakhstan) and other where a few bell curves can be identified (*e.g.* Algeria, India, Kuwait, Venezuela) (Bardi and Yaxley, 2005; Nashawi *et al.*, 2010). According to Murphy and Hall (2011), Hubbert predicted that peak oil for the world would occur in 2000 based on an Estimated Ultimate Recoverable (EUR) quantity for global oil of 1,250 Gbbl. Deffeys (2001) predicted that global CO production would peak in 2003 based on a EUR of 2,120 Gbbl. Also Campbell (1998) estimated that the peak in global oil would occur in 2003, but his estimate was based on a EUR of 1,800 Gbbl.

Dees *et al.* (2007) estimated the logistic curve of Hubbert (1961) for cumulative oil production:

$$Q_t = \frac{Q^{\infty}}{1+a\cdot e^{-b(t-t_0)}} \tag{52}$$

where Q_t is cumulative oil production at time t, Q^{∞} is the ultimate recoverable supply, and t_0 is the start date of the analysis. Al-Bisharah *et al.* (2009) provided the demonstration of the formula. The first difference of the logistic curve gives an estimate for the annual rate of production. As the physical characteristics of the oil fields do not entirely determine production, the model that is described in Dees *et al.* (2007) incorporated the effects of economic and political variables. The annual rate of production evaluated with the production curve (ΔQ_t) was used as an explanatory variable in the following relation:

$$PROD_t = \alpha + \beta_1 \cdot \Delta Q_t + \beta_2 \cdot ROIL_t + \beta_3 \cdot Dummy + \beta_4 \cdot Asym + \varepsilon_t \tag{53}$$

where $PROD$ is CO production, $ROIL$ is the real price of CO, $Dummy$ is a dummy (*i.e.* Boolean) variable that may affect local production of non-OPEC countries, $Asym$ is a variable designed to test the symmetry of the production curve and can be used only for regions where production has continued beyond the peak of the production curve, and ε_t is the error term.

The analysis that is conducted in the present work takes into account the period from the first quarter of 2010 (*i.e.* Feb, 2010) to the first quarter of 2015 (*i.e.* Feb, 2015), when the overall US production rose from 5.5 Mbbl/d to 9.18 Mbbl/d (*i.e.* a 67% increase). US production approached Saudi Arabia offer, and has heightened the global oversupply. Indeed, in the new CO price model the US production appears in the $Delta$ variable.

Hence, the US production results monotonically increasing in the last quarters and can be modeled as:

$$Production_{t+1}^{USA} = \omega_0 + \omega_1 Production_t^{USA} + \omega_2 Price_t \tag{54}$$

Table 18 contains the values of the adaptive parameters in Equation (54) and Figure 87, Figure 88, and Figure 89 show the results of one-step-ahead simulations over the analyzed quarters. The trend of the model is close to the real one with a maximum overestimation error of 4.4%. Appendix A reports further one-step-ahead simulations between 2010 and 2015.

Table 18 - Adaptive parameters in Equation (54) for the USA production model.

Parameter	Value	
	Brent	WTI
$\boldsymbol{\omega_0}$	-1.43231	-0.8788
$\boldsymbol{\omega_1}$	1.097971	1.041671
$\boldsymbol{\omega_2}$	0.008628	0.008405

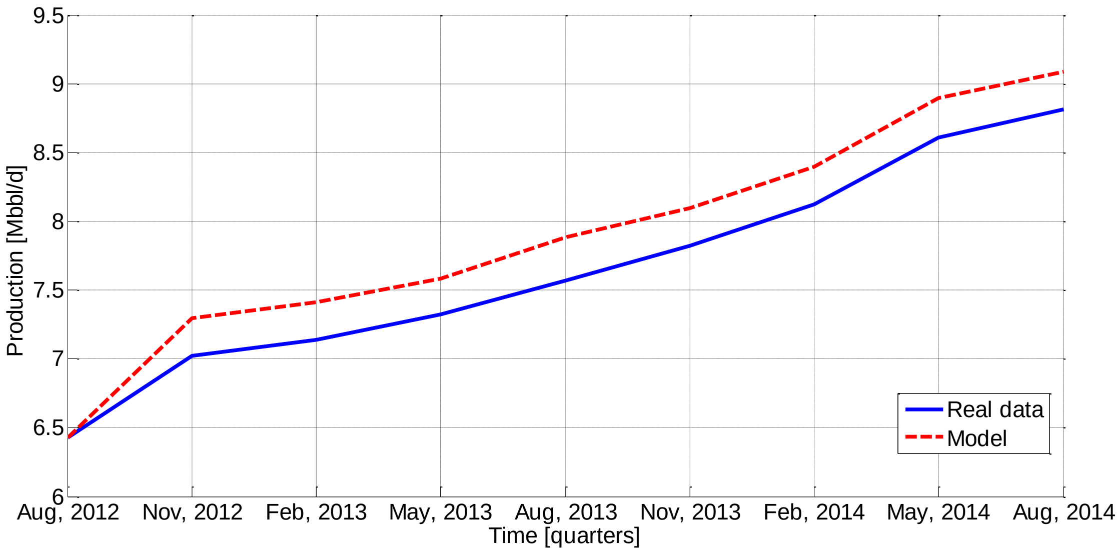


Figure 87 - One-step-ahead simulation of quarterly USA production from Aug, 2012 to Aug, 2014 (real data from EIA).

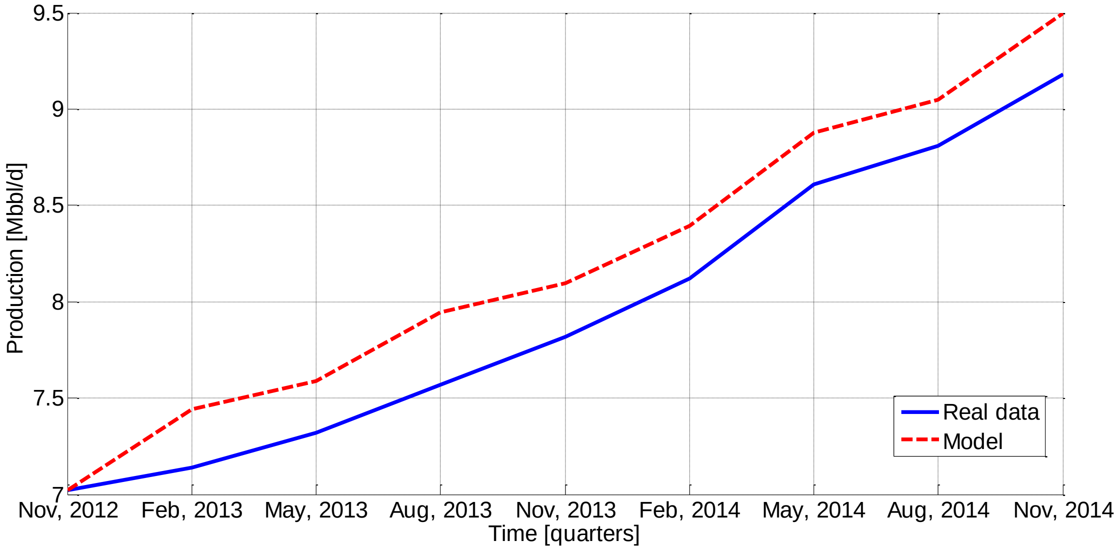


Figure 88 - One-step-ahead simulation of quarterly USA production from Nov, 2012 to Nov, 2014 (real data from EIA).

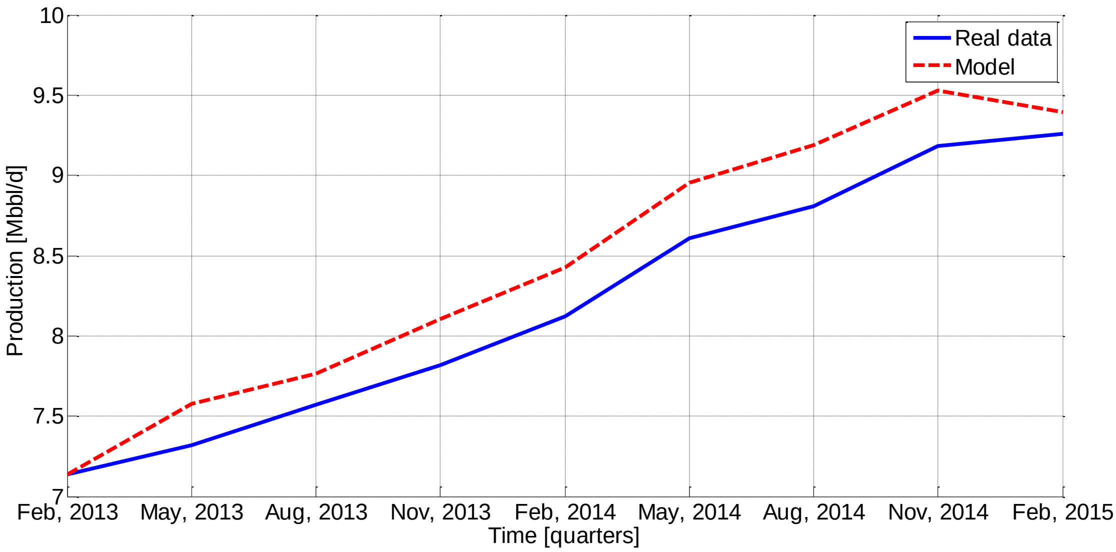


Figure 89 - One-step-ahead simulation of quarterly USA production from Feb, 2013 to Feb, 2015 (real data from EIA).

5.2.3.6 OPEC quotas

The input variable quota appears in the $Cheat$ variable, which explains how production levels in the investigated countries differ from their respective quotas. Dahmani and Al-Osaimy (2001) examined the correlations and the causal relations between compliance level and fundamentals of oil market. The compliance level is measured by the deviation of the production level from the respective quota for OPEC member countries, while oil market fundamentals are represented by OECD CO demand and stock levels (*i.e.* inventories), OPEC supply and CO price. In the new economic model a forecast equation for OPEC quotas in not proposed. Instead, according to the OPEC behavior that has influenced the CO market literature and historical events, the model acts as the swing producer itself. OPEC bets on stabilization for several reasons. A too low price of CO (less than 40 USD/bbl) would produce serious consequences on the producer countries (OPEC) that need to make profits from their investments. On the other side, an excessive cost of the barrel would penalize the consumers (OECD), decrease the demand, and favor the exploitation of other sources of energy. In other words, if the prices decreased during a prolonged period (*e.g.*, two quarters) the quotas would automatically change from 30 Mbbl/d to 27 Mbbl/d. On its turn, the decrease in CO production would make its quotations increase in the following few quarters. Eventually, this bullish trend would produce also a return of the quotas to the initial value of 30 Mbbl/d. On the contrary, in case the prices increased for two consecutive quarters the quotas would automatically change from 30 Mbbl/d to 33 Mbbl/d, in order to re-equilibrate the prices.

It is worth observing that the sign of the adaptive parameter α_2 that is associated to OPEC quotas does not comply with the abovementioned comments. This is due to the recent events that influenced CO market. In particular, the predicted sign by Dees *et al.* (2007) for the variable $Quota$ was negative, as an increase in OPEC production quotas should induce a supply increase and a price decrease, and *vice versa*. However, the recent political decision of not cutting production quotas in December 2015 instead of inducing a CO price collapse and going back to the breakeven point of oil countries has modified the abovementioned considerations. Indeed, the sign of the variable $Quota$ that derives from the regression between the first quarter of 2013 (*i.e.* Feb, 2013) and the first quarter of 2015 (*i.e.* Feb, 2015) is positive. In particular the sign has become positive in the period Nov, 2011 - Nov, 2013 (see Figure 65).

Before proceeding with the model simulation and validation of the forecast horizon, Section 5.2.4 shows some mathematical considerations about model parameters, in order to understand their identifiability and deeper meaning. Indeed, without a guarantee of *a priori* identifiability, the estimates of the parameters which could be obtained by some numerical optimization or regression algorithms are totally unreliable and inconsistent.

5.2.4 Model identifiability

The following step of the model analysis is the identifiability assessment. Identifiability assessment is a critical step in the process of parameter estimation. It addresses whether it is possible to determine univocally the model parameters from a given data set (*e.g.* historical data of CO prices, inventories, demand, production). In other words, a model is identifiable if it is theoretically possible to determine the true value of its parameters by means of the available variables observations and data. Mathematically, this is equivalent to saying that different values of the parameters must generate different probability distributions of the observable variables (Eisenberg and Hayashi, 2014). Identifiability is a necessary condition for successful model parameter estimation. Identifiability is largely used in pharmacokinetics, pharmacodynamics, cellular biology, and physiology to determine which parameters should be measured experimentally to recover/guarantee model identifiability (*i.e.* experimental design). Analogously, this thesis analyzes the identifiability of OPEC-based model to assess its robustness and consistency (*i.e.* model design).

Several numerical approaches to determine identifiability of different models were developed in the literature (Bellu *et al.*, 2007; Dietrich *et al.*, 2010; Chis *et al.*, 2011; Fortunati *et al.*, 2012; Eisenberg and Hayashi, 2014; Hines *et al.*, 2014; Wittman, 2015). We opted to use the method based on the Fisher Information Matrix (FIM) and the Differential Algebra implemented in the DAISY computer program (Saccomani *et al.*, 2003; Bellu *et al.*, 2007; Saccomani *et al.*, 2010) as numerical approaches are more computationally tractable than analytical ones (*e.g. Matlab*©, Octave, Fortran, Reduce). Even if the OPEC behavior model is not a differential system model, it is possible to use the same mathematical tools proposed in Bellu *et al.* (2007), Eisenberg and Hayashi (2014), and Wittman (2015), since that model is dynamic-discrete. Before starting with the description of the aforementioned techniques, few words should be spent on the definition of identifiability, by avoiding a useless mathematical formalism that would be excessive for the application of interest.

5.2.4.1 The concept of identifiability

Identifiability concerns uniqueness of the model parameters determined from input-output data, under ideal conditions of noise-free observations and error-free model structure. In other words, this analysis explores and tries to answer the following question: given an input, model, and experimental output, is it possible to uniquely identify the model parameters? If the analyzed parameters take on individually unique values which allow evaluating a given output, the model is then considered *globally* (or uniquely) structurally identifiable. If there are unique values of the parameters within a local neighborhood of parameter space, the model is considered *locally* structurally identifiable. Eventually, if there is a range of values even for one single parameter which yields the function output, then the model is considered *unidentifiable*. A model is said to be globally structurally identifiable if all the parameters are globally structural identifiable. If any parameters are locally structurally identifiable or unidentifiable, the model is respectively considered locally structurally identifiable or unidentifiable. In the case of model unidentifiability, the model parameters rather often form identifiable combinations, *i.e.* combinations of parameters that are identifiable even though the individual parameters are unidentifiable (Eisenberg and Hayashi, 2014). While structural identifiability is often treated as a formal/mathematical property and therefore evaluated with suitable analytical approaches (*e.g.* Laplace transform method), a range of numerical approaches was developed in the scientific literature to investigate the structural identifiability of a model (*e.g.* FIM). Numerical approaches often address local structural identifiability at a particular point in the parameter space, as they typically require numerical values for the parameters to be used. This can often be partially mitigated by testing a wide range of parameter values (Eisenberg and Hayashi, 2014).

The present thesis uses the FIM and Differential Algebra to test the identifiability of the OPEC-based model and the input-variable models that were presented in Section 5.2.2 and in Section 5.2.3, respectively. Since the models are discrete in time, it was decided to choose a particular point in the parameter space (*i.e.* the parameters that were discussed and reported in the previous Sections), and as input and output observations a specific point of the historical data series of CO prices and input variables.

5.2.4.2 Fisher Information Matrix

The FIM approach is a simple method for model design and validation. It requires to know only the model and the measurement uncertainties, which are not considered for the data taken from historical time series of CO quotations. In literature there are a few definitions and applications of the FIM (Dietrich *et al.*, 2010; Fortunati *et al.*, 2012; Eisenberg and Hayashi, 2014; Wittman, 2015), and the most relevant for our purposes is discussed in the following. Analytically, the Fisher matrix F is the inverse of the covariance matrix C, which represents the uncertainties in model parameters. For N model parameters $A = \{p_1, \dots, p_N\}$, the FIM is a $N \times N$ symmetric matrix that represents the amount of information contained in the data y^*, about parameters A. If F is singular, A is unidentifiable. Since F is difficult to calculate explicitly, it is evaluated by using numerical approximations of the parameter sensitivities. In practice, A is typically considered unidentifiable when the determinant of the FIM is non-zero but small or better (numerically) the condition number is rather high. In addition, the rank of F corresponds to the number of identifiable parameters or parameter combinations in A. By inverting the FIM, one obtains the Cramér-Rao bound Covariance Matrix C, whose diagonal entries correspond to the individual variances of parameters in A. If A is a singleton parameter set $\{p\}$, the variance C is simply the reciprocal of the squared parameter sensitivity $\frac{1}{\left(\frac{dy}{dp}\right)^2}$. Matrix C does not properly exist when the model is unidentifiable. In the present thesis the FIM is calculated by means of *Matlab*© and the condition number of the matrix is calculated before inverting it to C (it is numerically necessary to invert it as a simple and more robust/efficient factorization would not provide the required matrix information). The condition number respect to the inversion procedure measures the sensitivity of the solution of a system of linear equations to errors in the data. If it is not much higher than 1, the matrix is well-conditioned (*i.e.* the output (*i.e.* regressed) value of the parameters does not change significantly for a small change in the input data), otherwise it is ill-conditioned.

Given a model and corresponding output function, the FIM for the parameter set $\{p\}$ can be computed as follows:

- Generate the sensitivity matrix: $X = \left(s(t, p_1), \dots, s(t, p_N)\right)$ where $s(t, p_i) = \left[\frac{\partial y}{\partial p_i}(t_0), \dots, \frac{\partial y}{\partial p_i}(t_t)\right]$;
- Compute the FIM: $F = X^T X$;
- Provided F is not singular (for instance through a condition number assessment), compute the covariance matrix C: $C = F^{-1}$.

If there are B observables $f_1, f_2, \dots, f_B$ (*i.e.* the input variables of CO price model), each one related to the model parameters by some equation $f_b = f_b(p_1, p_2, \dots, p_N)$, the FIM computation involves a summation over the observables:

$$F_{ij} = \sum_b \frac{\partial f_b}{\partial p_i} \frac{\partial f_b}{\partial p_j} \tag{55}$$

In order to calculate the FIM it is reasonable to assume a best value of the input data/variables and call that fiducial model, as the model structure does not change with the fiducial model (Wittman, 2015). To test the global identifiability of the model parameters, the explicit model is calculated by substituting the input variables in Equation (40) with the Equations (45-46-50-51-54):

$$PRICE_t = \alpha_0 + \alpha_1 \frac{\gamma_0+\gamma_1(\varepsilon_0+\varepsilon_1 Capacity_{t-1}+\varepsilon_2 Production_{t-1}^{OPEC})+\gamma_2(\beta_0 GDP_t+\beta_1 PRICE_{t-1}+\beta_2)}{\beta_0 GDP_t+\beta_1 PRICE_{t-1}+\beta_2} + \alpha_2 Quotas_t + \alpha_3[\xi_0 + \xi_1(\varepsilon_0 + \varepsilon_1 Capacity_{t-1} + \varepsilon_2 Production_{t-1}^{OPEC}) + \xi_2 PRICE_{t-1} - Quotas_t] + \alpha_4 \frac{\xi_0+\xi_1(\varepsilon_0+\varepsilon_1 Capacity_{t-1}+\varepsilon_2 Production_{t-1}^{OPEC})+\xi_2 PRICE_{t-1}}{\varepsilon_0+\varepsilon_1 Capacity_{t-1}+\varepsilon_2 Production_{t-1}^{OPEC}} + \alpha_5[\xi_0 + \xi_1(\varepsilon_0 + \varepsilon_1 Capacity_{t-1} + \varepsilon_2 Production_{t-1}^{OPEC}) + \xi_2 PRICE_{t-1} - \omega_0 - \omega_1 Production_{t-1}^{USA} - \omega_2 PRICE_{t-1}] \quad (56)$$

The actual input variables of the complete model result to be GDP_t, $Capacity_{t-1}$, $Quotas_t$, $PRICE_{t-1}$, $Production_{t-1}^{OPEC}$, and $Production_{t-1}^{USA}$, but this model definition was not used in the analysis of the present work because the parameters are often reciprocally multiplied and are not identifiable in a univocal combination.

Appendix B contains an algorithmic implementation in *Matlab©*, with the lines of code and procedure description, which can be used by the reader to compute the FIM quickly also for a variety of other models. It is worth observing that the *Matlab©* source code features the declaration of the independent variable of each involved model. Then, the analytical derivative of the model can be evaluated and the parameter values are calculated with a regression procedure (see Sections 5.2.2 and 5.2.3). Equation (55) provides the procedure to determine the FIM of CO price model and input variable models. The evaluation of the FIM condition number allows assessing the invertibility of such a matrix and eventually computes the covariance matrix. The implicit OPEC-based model described in Equation (40) with the support of Equations (45-46-50-51-54) results identifiable. On the contrary, the explicit model that is shown in Equation (56) is not identifiable, as there are combinations of identifiable parameters that bring to the given data set.

5.2.4.3 A differential algebra approach: DAISY algorithm

Differential algebra tools have been applied to study the identifiability of dynamic systems described by polynomial equations and rational function ODE models (Saccomani *et al.*, 2003). Differential algebra is used here to assess through a different method respect to FIM the identifiability of CO price model although it is not differential but is discretized in time.

Differential algebra approach requires knowing the concepts of *ring* and *ideal*. The former indicates a set having two binary operations (typically addition and multiplication), while the latter is a subring of a mathematical ring with certain properties, *i.e.* zero is an element of the ideal, the ideal is closed under addition, and the product of an element of the ideal and an element of the initial ring is an element of the ideal (*e.g.* the set of even integers as a subset of the integer ring). The idea that the characteristic set of differential ideal (*i.e.* an ideal that is close also with respect to differentiation) generated by the system dynamic polynomials provides the tool for testing global identifiability is due to Olliver (1990), and Ljung and Glad (1994). Nevertheless, this work takes as a reference the new algorithms that were developed in Bellu *et al.* (2007) because they are more suitable and intuitive for the current application. The same paper provides also a background on differential algebra, which is here summarized.

A characteristic set is a special basis (*i.e.* a minimal set of differential polynomials) that generates the same differential ideal generated by an arbitrary given set of differential polynomials. If one or more polynomials are rational, they are reduced to the same denominator. A binary matrix is assembled, so that all the information on the structure of the dynamic system can be summarized. This matrix has as many rows as the model equations, and as many columns as the model variables (*i.e.* inputs $\bar{u}$, outputs $\bar{y}$, and states $\bar{x}$) and their corresponding derivatives. In particular, the matrix is constructed by considering one equation, *i.e.* row, at a time, and by writing 1 if the corresponding variable/derivative is present in the equation and 0 if it is absent. The computation of the characteristic set is performed *via* the Ritt's pseudo-division algorithm, which requires the introduction of a ranking among the model variables (*i.e.* inputs, outputs, states, and their derivatives). The standard ranking used in system identification declares the input and the output components, which are known variables in this context, as the lowest ranked variables, and the highest rank is given to the state variable components (Bellu *et al.*, 2007):

$$u_1 < u_2 < \cdots < y_1 < y_2 < \cdots < \cdots < x_1 < x_2 < \cdots \qquad (57)$$

where $\bar{u}$ is the input vector (*i.e.* the independent variables of price and input variable models), $\bar{y}$ is the output vector (*i.e.* the dependent variable of each model described by Equations (40-45-46-50-51-54)), and $\bar{x}$ is the state variable vector (*i.e.* the variables used to describe the state of a dynamical system, which are not present in our discrete models). Given this rank among the variables, their derivatives can be ranked in different ways. According to Bellu *et al.* (2007), it turns out convenient to choose the following ranking:

$$u_1 < u_2 < \cdots < y_1 < y_2 < \cdots < \dot{u}_1 < \dot{u}_2 < \cdots < \ddot{u}_1 < \ddot{u}_2 < \cdots < \dot{y}_1 < \dot{y}_2 < \cdots < x_1 < x_2 < \cdots < \dot{x}_1 < \dot{x}_2 < \cdots \quad (58)$$

The pseudo-division algorithm to calculate the characteristic set is then applied with respect to the declared ranking (58), thus obtaining a family of differential polynomials, which are identically determined if all their coefficients are known. Each polynomial is compared with the previous ones. If it is of equal or higher rank, it is reduced with respect to the preceding ones by applying the pseudo-division algorithm. The characteristic set is formed when no further reductions can be performed. Then, the observability test states that if in the characteristic set a state component appears without derivatives then it is algebraically observable. The coefficients, which are functions of the unknown parameters, of the output, and of the model vector, and their derivatives, are extracted and provide the exhaustive summary of the original dynamic model. If parameter equality constraints are present, they are included in the exhaustive summary. Each function of the unknown parameters is evaluated at a pseudo-randomly chosen numerical value for the parameter vector and set equal to the obtained numerical value. Thus, a system of algebraic nonlinear equations in the unknown parameters is constructed. These algebraic nonlinear equations are submitted to the Groebner algorithm, in order to solve the system, return the solutions for each unknown parameter, and provide the model identifiability results, *i.e.* global or local identifiability or non-identifiability.

The software used in this thesis is *DAISY (Differential Algebra for Identifiability of Systems)*, which is provided by University of Cagliari and University of Padova (Italy), and can be downloaded at http://daisy.dei.unipd.it/. Appendix C contains the structure of the input and output files that are involved in the analysis of Equations (40-45-46-50-51-54) and Equation (56) by means of *DAISY*. It is worth observing that each input file comprises the declaration of the output variable (*i.e.* CO price, demand, capacity, inventories OPEC production, and USA production) and of the input variables of the involved model. In order to test the model identifiability it is reasonable to assume a best value of the input variable (called fiducial model) and choose an integer value (called *seed*) that is bigger than the number of unknown parameters. The *seed* represents the upper bound of the interval where the subroutine "random" will choose the numerical values corresponding to each parameter. In line with what was determined in Section 5.2.4.2, the model described by Equations (40-45-46-50-51-54) results globally identifiable, whilst the explicit price model (56) is not identifiable.

Thanks to these results, the model can be validated in the past years (see Section 5.2.6) and simulated over future time-horizons (see Section 5.3).

5.2.5 Sensitivity analysis

The sensitivity analysis accomplishes the already discussed signs and values of the model parameters and the relative importance of the input variables that are involved in the new CO price model. The sensitivity analysis allows also to carry out either bullish- or bearish-trend scenarios.

The term "sensitivity analysis" consists in the study of how the uncertainty in the output of a mathematical model or system can be ascribed to different sources of uncertainty in its inputs. In other words, the sensitivity analysis evaluates the effects that are induced on the results of the price model by changes/uncertainties/fluctuations in the values of the input variables. The parametric sensitivity matrix corresponds to the first step of the FIM procedure and can be computed in the following way:

$$P = \big(s(t, p_1), \dots, s(t, p_N)\big) \tag{59}$$

where:

$$s(t, p_i) = \left[\frac{\partial y}{\partial p_i}(t_0), \dots, \frac{\partial y}{\partial p_i}(t_t)\right] \tag{60}$$

The input variables sensitivity matrix can be expressed as:

$$S = \big(s(t, x_1), \dots, s(t, x_M)\big) \tag{61}$$

where:

$$s(t, x_i) = \left[\frac{\partial y}{\partial x_i}(t_0), \dots, \frac{\partial y}{\partial x_i}(t_t)\right] \tag{62}$$

and x, y, and p represent the input variables, CO price model, and model parameters, respectively, while N and M are the number of parameters and input variables of each model equation.

5.2.5.1 Parametric sensitivity

As shown in Figure 90, Figure 91, Figure 92, Figure 93, Figure 94, and Figure 95, the derivatives of the CO price respect to each model parameter are constant and equal to the parameter that is taken into account, because the model is linear in the adaptive parameters.

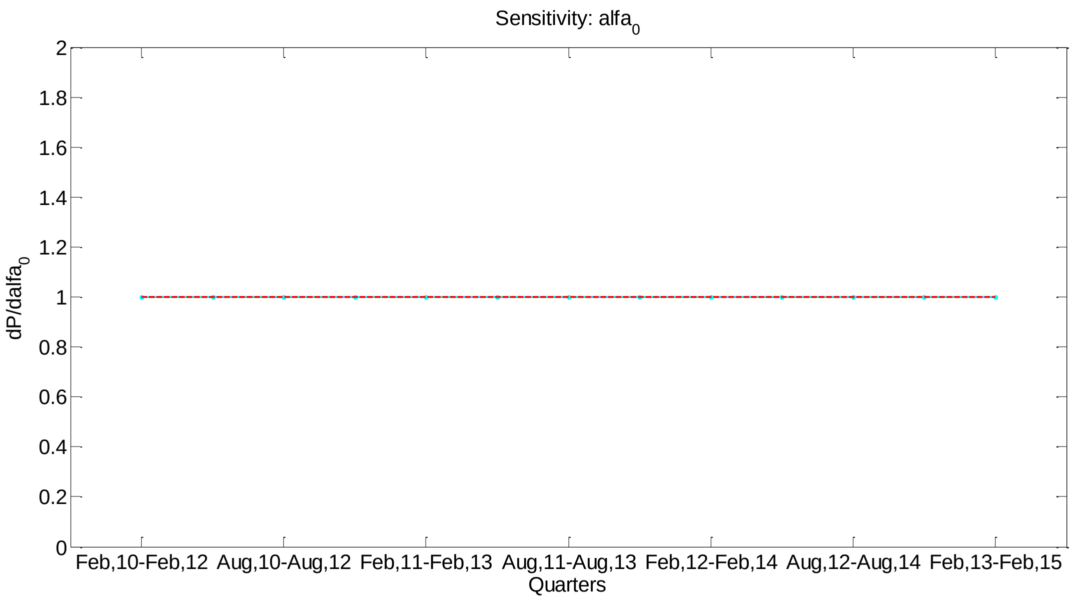


Figure 90 - Sensitivity analysis of the price model respect to the adaptive parameter α_0. The adopted values of the input variables are in ascending order for cyan, red, and blue lines.

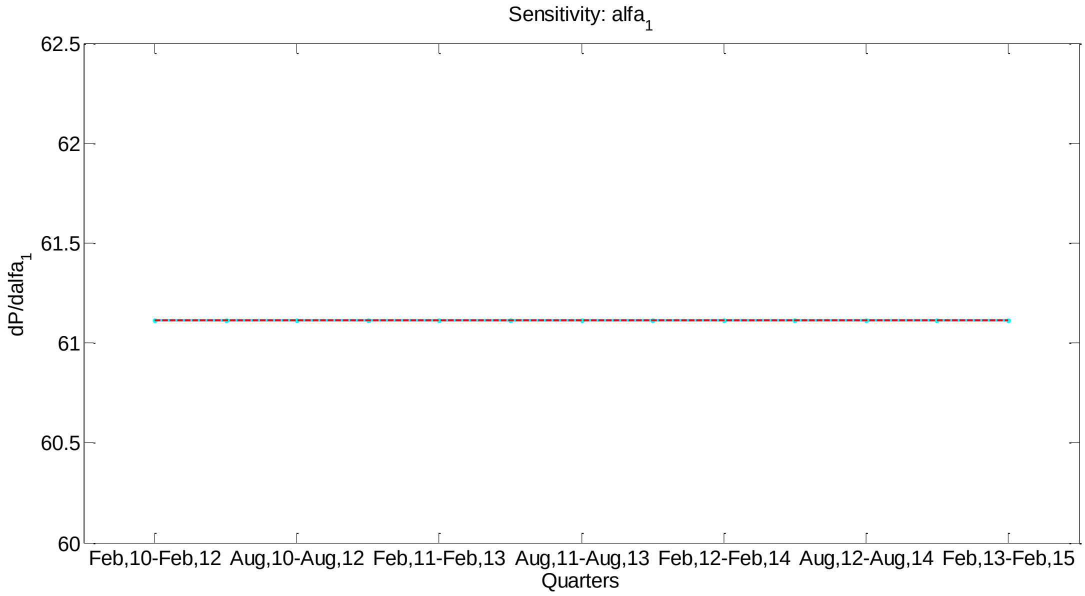


Figure 91 - Sensitivity analysis of the price model respect to the adaptive parameter α_1. The adopted values of the input variables are in ascending order for cyan, red, and blue lines.

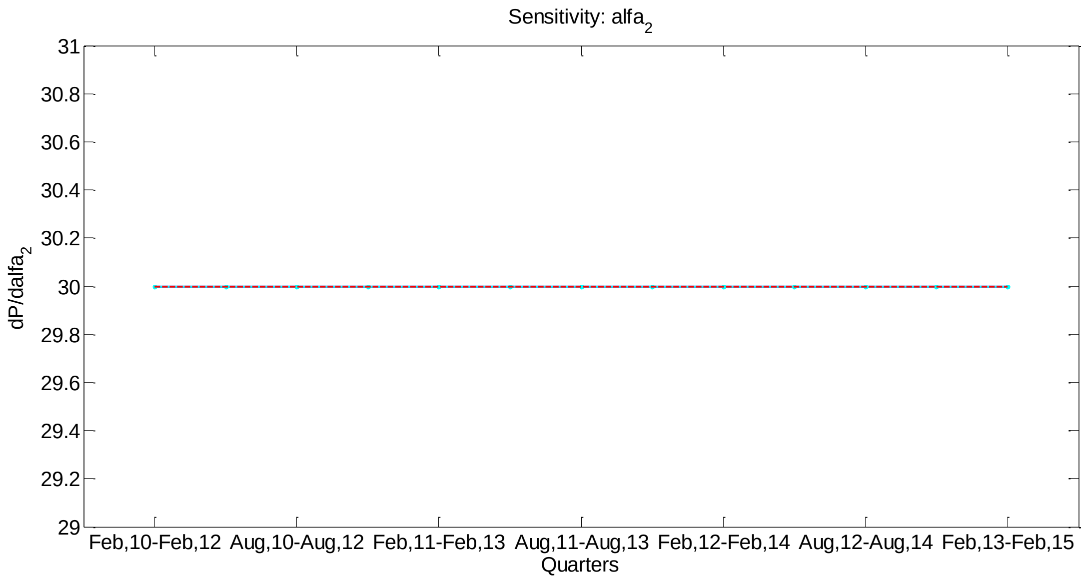


Figure 92 - Sensitivity analysis of the price model respect to the adaptive parameter α_2. The adopted values of the input variables are in ascending order for cyan, red, and blue lines.

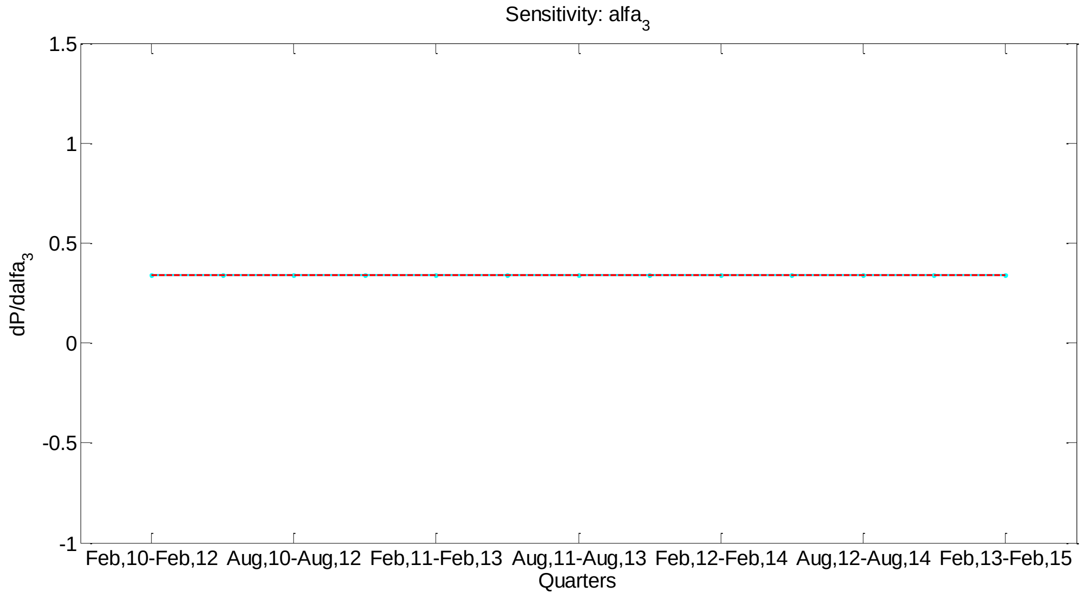


Figure 93 - Sensitivity analysis of the price model respect to the adaptive parameter α_3. The adopted values of the input variables are in ascending order for cyan, red, and blue lines.

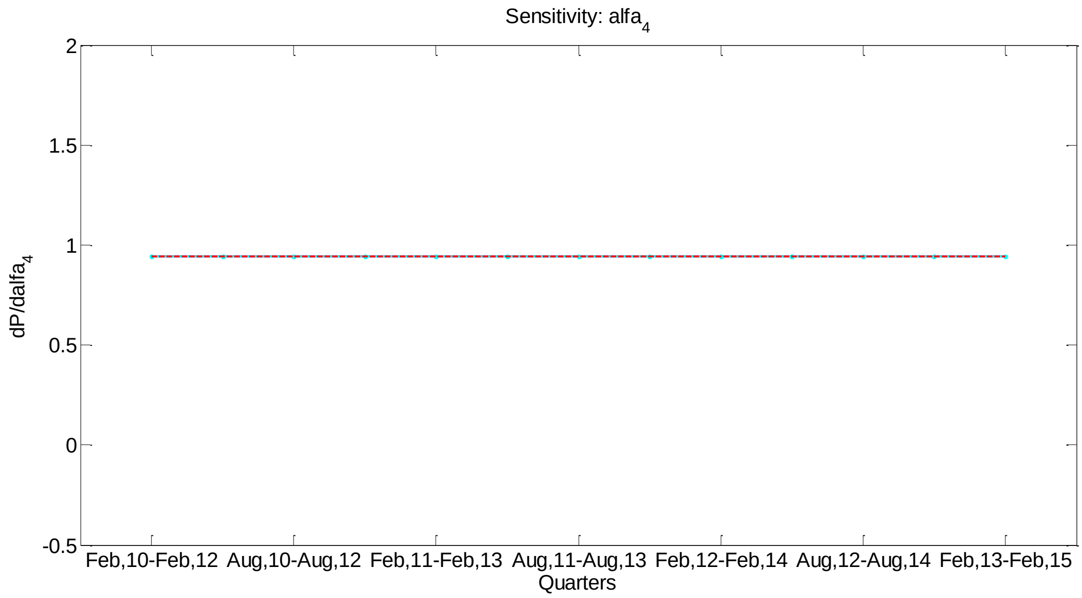


Figure 94 - Sensitivity analysis of the price model respect to the adaptive parameter α_4. The adopted values of the input variables are in ascending order for cyan, red, and blue lines.

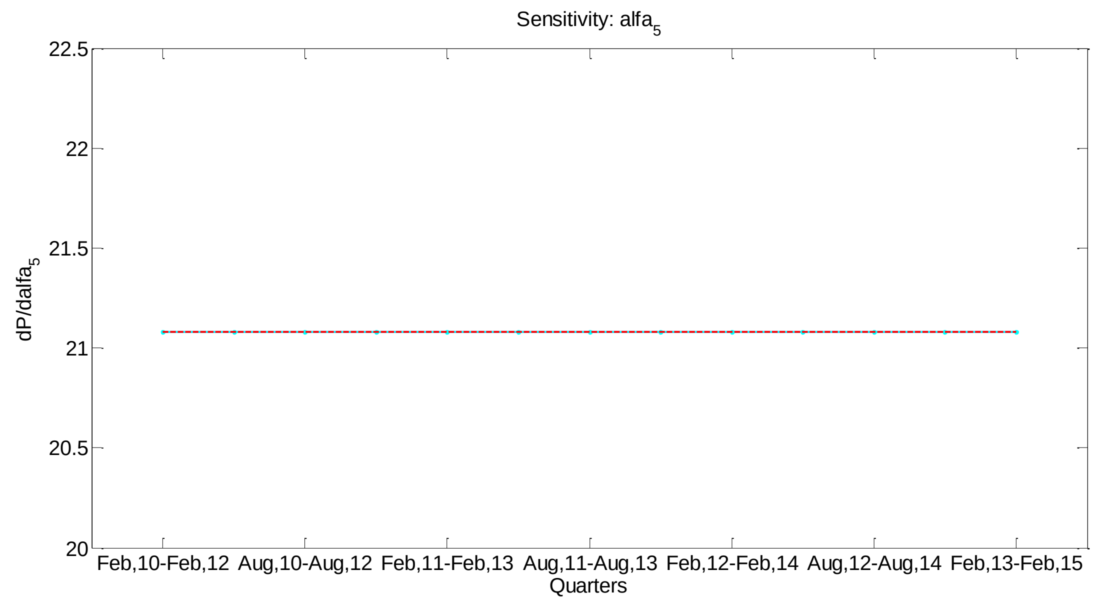


Figure 95 - Sensitivity analysis of the price model respect to the adaptive parameter α_5. The adopted values of the input variables are in ascending order for cyan, red, and blue lines.

5.2.5.2 Input variables sensitivity

The results of the sensitivity analysis respect to the input variables play a major role in understanding the relation between CO price and supply-and-demand variables so to better generate future price scenarios. In particular, given the adaptive parameter values in Table 11, Figure 96 shows that the derivative of price respect to OECD CO demand is not constant:

$$\frac{dPRICE_t}{dDemand_t} = -\alpha_1 \cdot \frac{Inventory_t}{Demand_t^2} \tag{63}$$

Furthermore, the greater the demand the lower the CO price variation, under the condition of considering constant all the other model variables. The same remark is valid for the OPEC production:

$$\frac{dPRICE_t}{dProduction_t^{OPEC}} = \alpha_3 + \frac{\alpha_4}{Capacity_t^{OPEC}} \tag{64}$$

Indeed, an increase in $Production^{OPEC}$ due to an increase in $Capacity^{OPEC}$ (α_3 is negative and α_4 is positive) increases the supply and decreases the prices (see Figure 97).

As for inventories and production capacity (see Figure 98 and Figure 99), an increase of these variables causes a raise of CO price differential, also due to a decreasing demand. In formulas:

$$\frac{dPRICE_t}{dInventory_t} = \frac{\alpha_1}{Demand_t} \tag{65}$$

$$\frac{dPRICE_t}{dCapacity_t^{OPEC}} = -\alpha_4 \cdot \frac{Production_t^{OPEC}}{{Capacity_t^{OPEC}}^2} \tag{66}$$

Since the model is linear respect to the USA production and OPEC quota (see Figure 100 and Figure 101), a unitary variation in these input variables causes the same variation in CO prices.

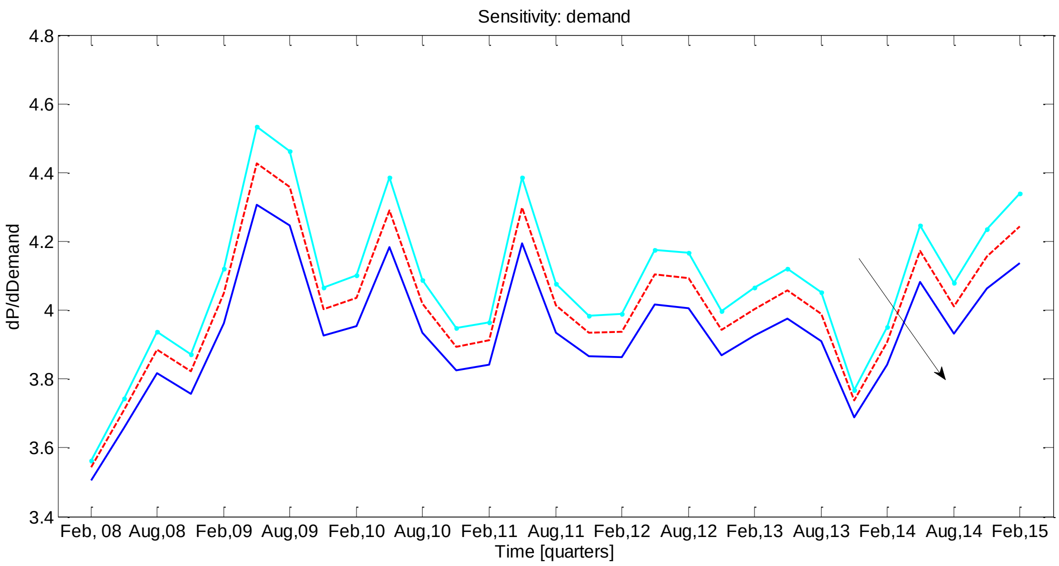


Figure 96 - Sensitivity analysis of the price model respect to OECD demand. The adopted values of the input variables are in ascending order for cyan, red, and blue lines.

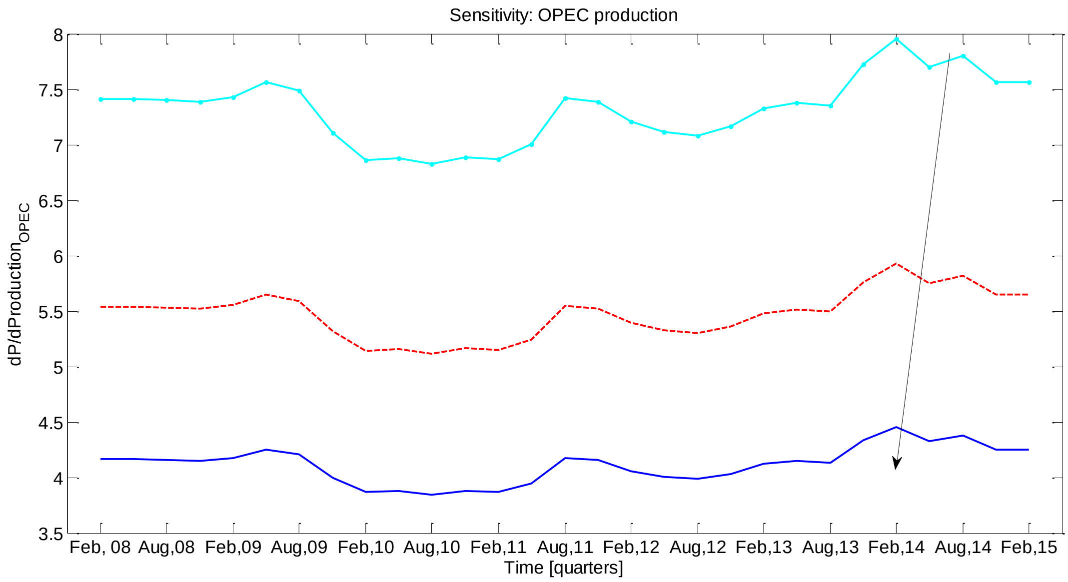


Figure 97 - Sensitivity analysis of the price model respect to OPEC production. The adopted values of the input variables are in ascending order for cyan, red, and blue lines.

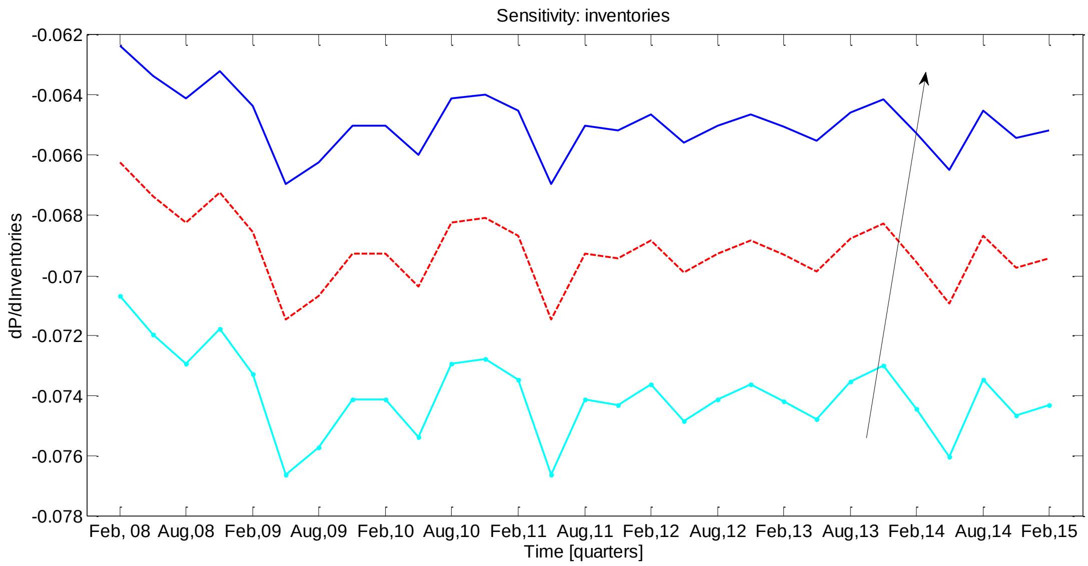


Figure 98 - Sensitivity analysis of the price model respect to OECD inventories. The adopted values of the input variables are in ascending order for cyan, red, and blue lines.

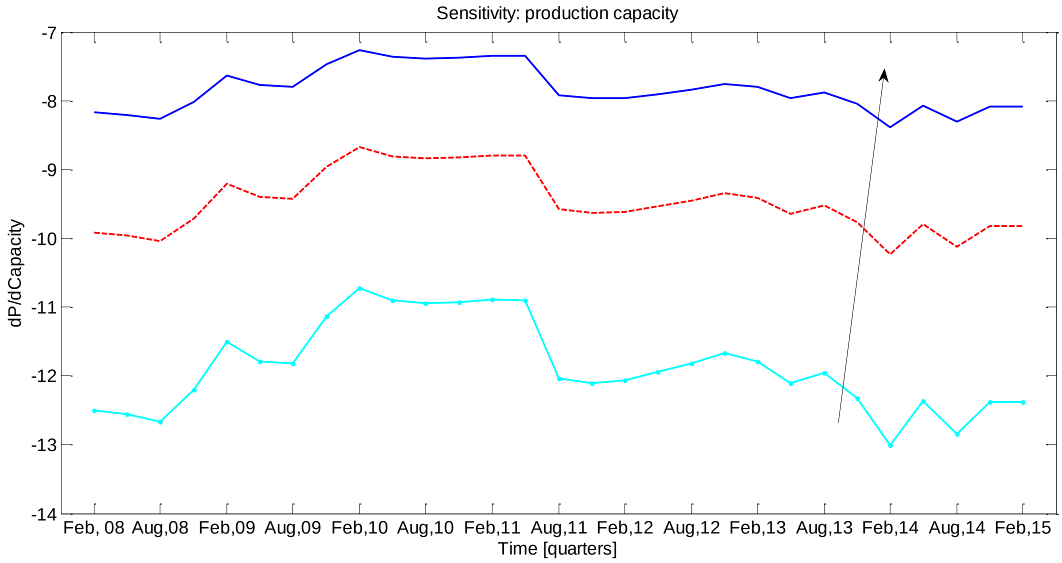


Figure 99 - Sensitivity analysis of the price model respect to OPEC production capacity. The adopted values of the input variables are in ascending order for cyan, red, and blue lines.

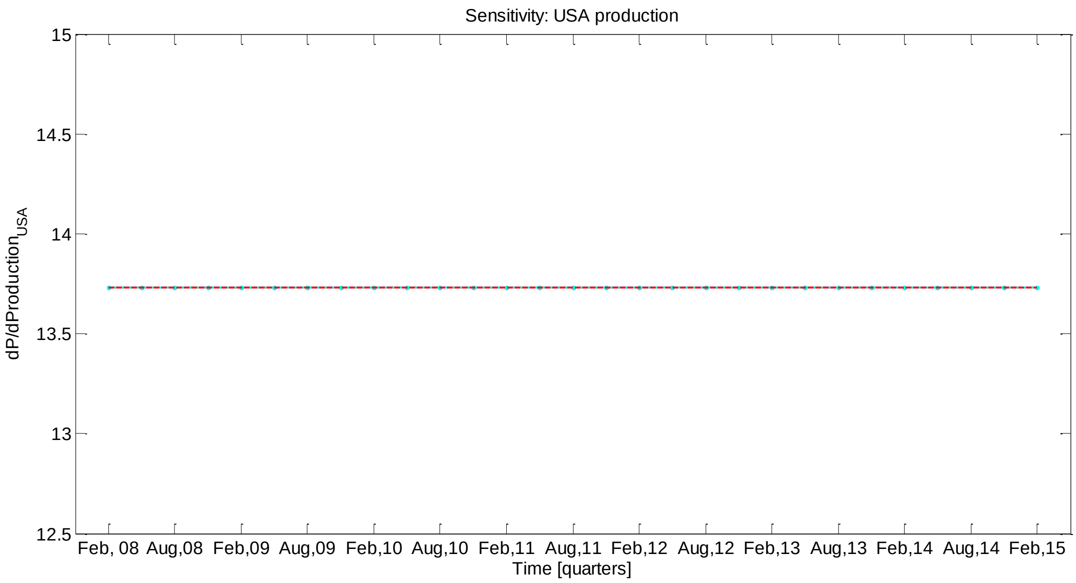


Figure 100 - Sensitivity analysis of the price model respect to USA production. The adopted values of the input variables are in ascending order for cyan, red, and blue lines.

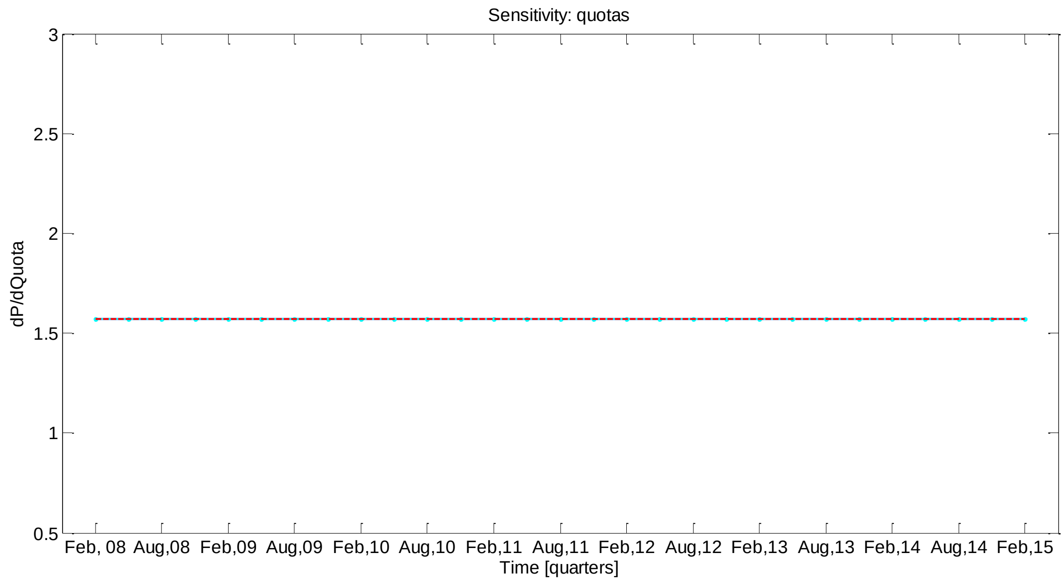


Figure 101 - Sensitivity analysis of the price model respect to OPEC quota. The adopted values of the input variables are in ascending order for cyan, red, and blue lines.

5.2.6 Model validation and selection of the forecast horizon

In order to investigate the forecast performance of the new model, Equations (40-45-46-50-51-54) are first applied to past time intervals where the real quotations are available and can be used as a reference value for comparison reasons. In particular the validation procedure starts from the first quarters of 2012, 2013, and 2014. As a consequence, a suitable forecast horizon is chosen according to the availability of real CO quotations up to February, 2015. Contrary to modified-Chevallier's model in Equation (35b), the whole OPEC-based model described by Equations (40-45-46-50-51-54) apparently does not contain any random functions or stochastic shock terms that justify the creation of distinct price and input variable scenarios. However, the model is created in order to catch possible variations of CO prices and input variables (*i.e.* demand, inventories, OPEC production, production capacity, and USA production) by means of the *Matlab©* function *rand*, which returns pseudorandom values drawn from the standard uniform distribution on the open interval between 0 and 1 and accounts for both the moment and the magnitude of the variations. Section 5.3 provides more details about the frequency of these variations, and major/minor increases/decreases of CO prices. Figure 102, Figure 103, and Figure 104 show 3000 distinct scenarios (*i.e.* a sufficiently high number to guarantee general conclusions) that were simulated from 2012, 2013, and 2014, respectively. The computation of these 3000 scenarios by a conventional laptop with a 4GB RAM required 0.996 seconds, while the creation of 10000 would take 3.153 seconds. Apart from this elapsed time difference, the representation of 3000 scenarios is a good compromise between adequate generality and useless graphic details.

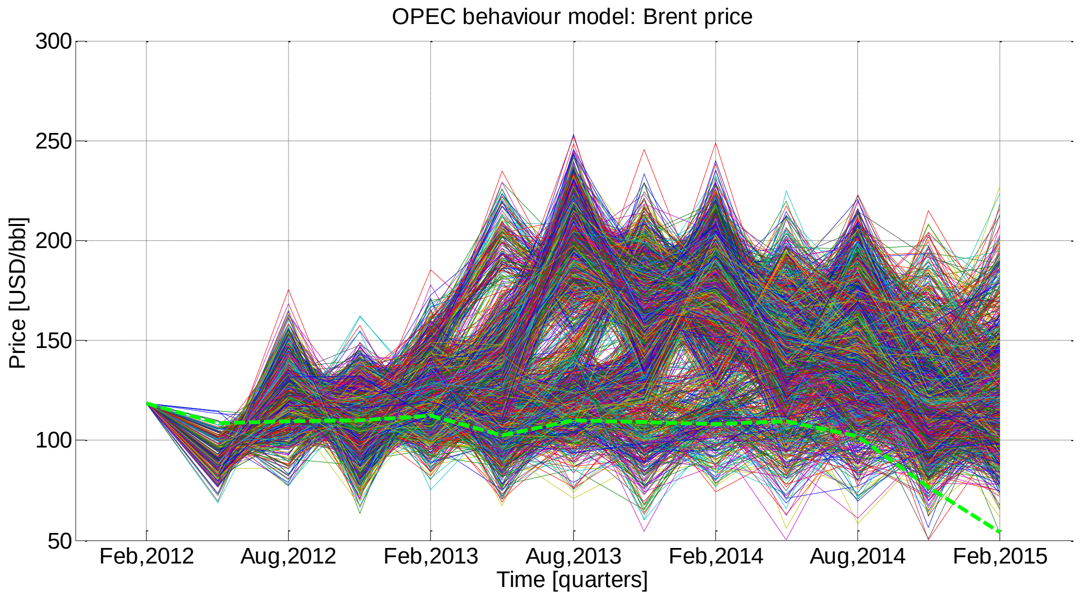


Figure 102 - Fully-predictive scenarios of Brent quarterly prices from Feb, 2012 to Feb, 2015. The green dotted line represents the real quotations (real data from EIA).

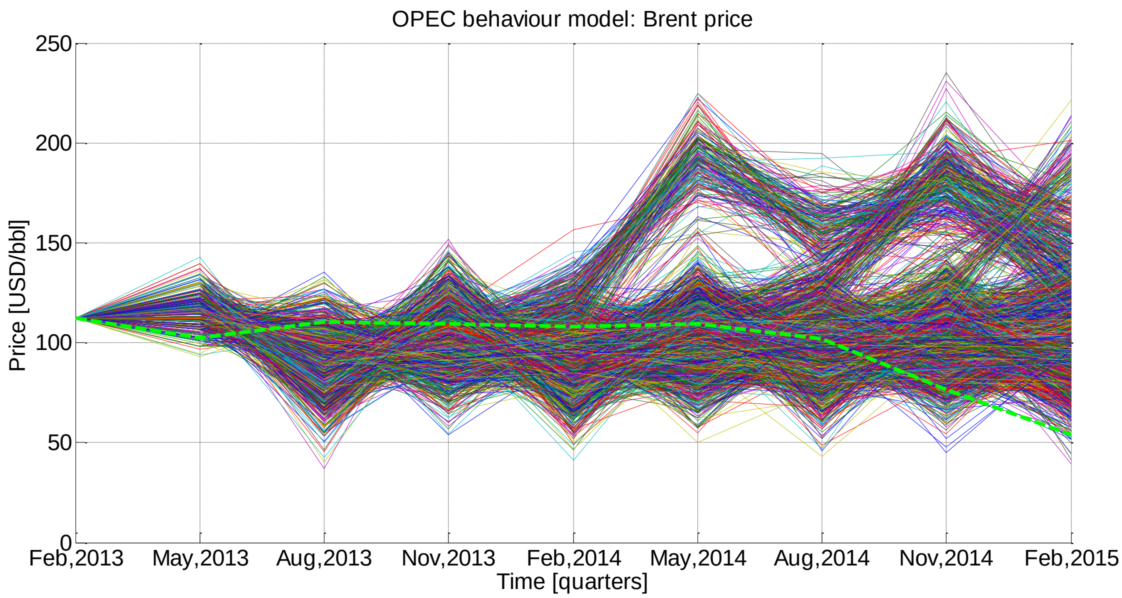


Figure 103 - Fully-predictive scenarios of Brent quarterly prices from Feb, 2013 to Feb, 2015. The green dotted line represents the real quotations (real data from EIA).

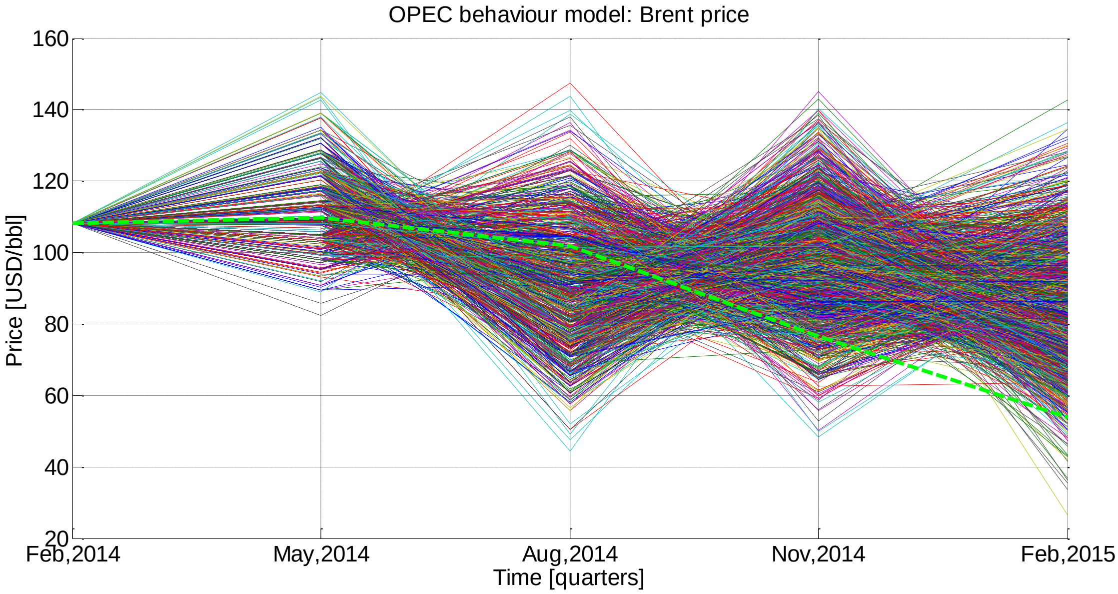


Figure 104 - Fully-predictive scenarios of Brent quarterly prices from Feb, 2014 to Feb, 2015. The green dotted line represents the real quotations (real data from EIA).

It is worthy observing that past prices (green dotted line) are nearly included in the simulated scenarios bounds throughout the simulation time span, except for the period between Nov, 2014 and Feb, 2015 when starting from 2012 (see Figure 102). Moreover, the distribution of these scenarios is not probabilistically uniform, because very high CO prices (*i.e.* higher than 150 USD/bbl) rather rarely occur. In particular, the validation procedure shows that 117 (3.9%), 1 (0.03%), and no one scenarios fall into the price range between 150 and 250 USD/bbl after one year in Figure 102, Figure 103, and Figure 104, respectively. For this reason, the simulation procedure calls for the creation of fan-charts with prediction-value ranges where the prices assume different probabilities to occur, but their creation is postponed in Section 5.3.

Since one-step-ahead simulations are not interesting as far as PSE/CAPE activities are conceived (except for real-time optimization) and one year-long forecasts are too short for PSE purposes, the new model can be applied for two years-long time-horizon, *i.e.* for short- and middle-term problems/activities. Frequently, PSE/CAPE applications involve time intervals of up to five years, because in the fields of Chemical Engineering and Dynamic Conceptual Design the time scale considered is rather large as a plant runs for several years (*e.g.*, fifteen-twenty years). Conversely, the realm where this model can operate correctly and consistently is scheduling and planning, which respectively cover short- and medium-term horizons. These problems deal with manufacturing management processes, by which raw materials and production capacity are optimally allocated to meet demand. This approach is conceptually simple but quite challenging and well suited, as the model really adapts to changes in demand, resource capacity, and material availability. Section 5.4 shows how to create bullish, bearish, and conservative-trend scenarios according to several considerations about input variables, which take into account both supply-and-demand sides.

In Figure 103, the model trend appears noisy and a price range seems not covered by any scenarios. It is worth observing in Figure 105 that respect the total number of 3000 simulated scenarios, the price curves that in May, 2014 and Nov, 2014 went past the maximum historical CO price of 150 USD/bbl (red dash-dot line) are only 190 (*i.e.* 6.3%) and 335 (*i.e.* 11.2%), respectively, and 180 scenarios are the same ones that overcome 150 USD/bbl in the both quarters. The fan-charts in Section 5.3 show how process designers/managers and chemical engineers can take into account the probability distribution of prices in their feasibility studies of chemical plants, and further clarify that CO prices in the upper range (*i.e.* higher than 150 USD/bbl) are unlikely to occur.

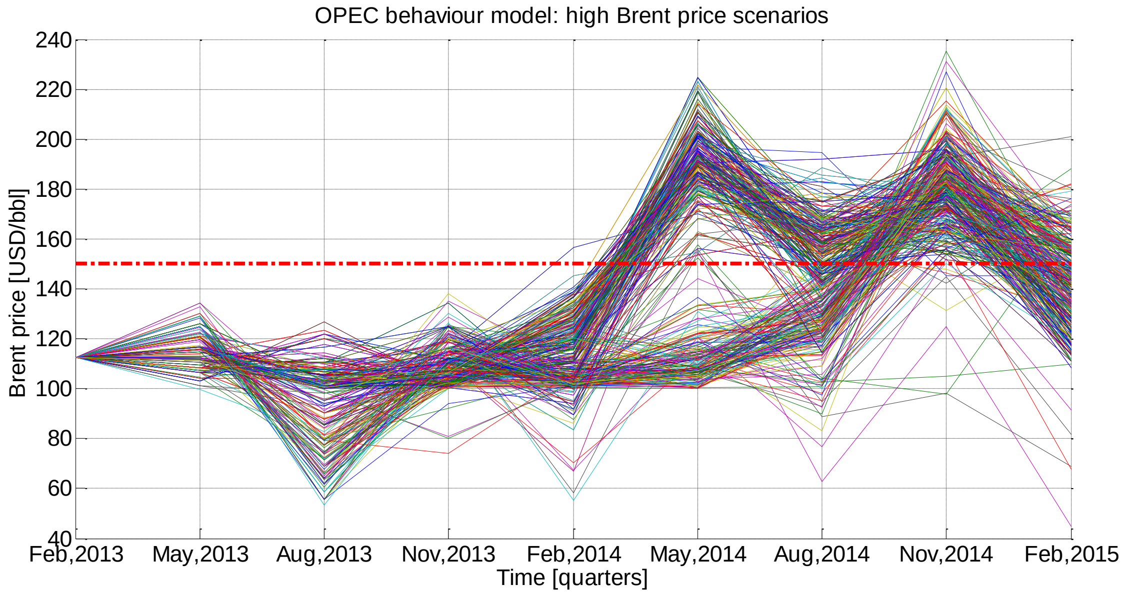


Figure 105 - Fully-predictive scenarios of Brent quarterly prices that go past the 150 USD/bbl threshold in May, 2014 and Nov, 2014. The red dash-dot line represents the price level of 150 USD/bbl which is practically the maximum value of CO quotations and took place in July, 2008.

It is important to verify that also the past real values of the input variables (red dotted line) are contained in the distribution of simulated scenarios (see Figure 106, Figure 107, Figure 108, Figure 109, and Figure 110). Hence, the choice of two year-long time-horizons is suitable for input variables, too.

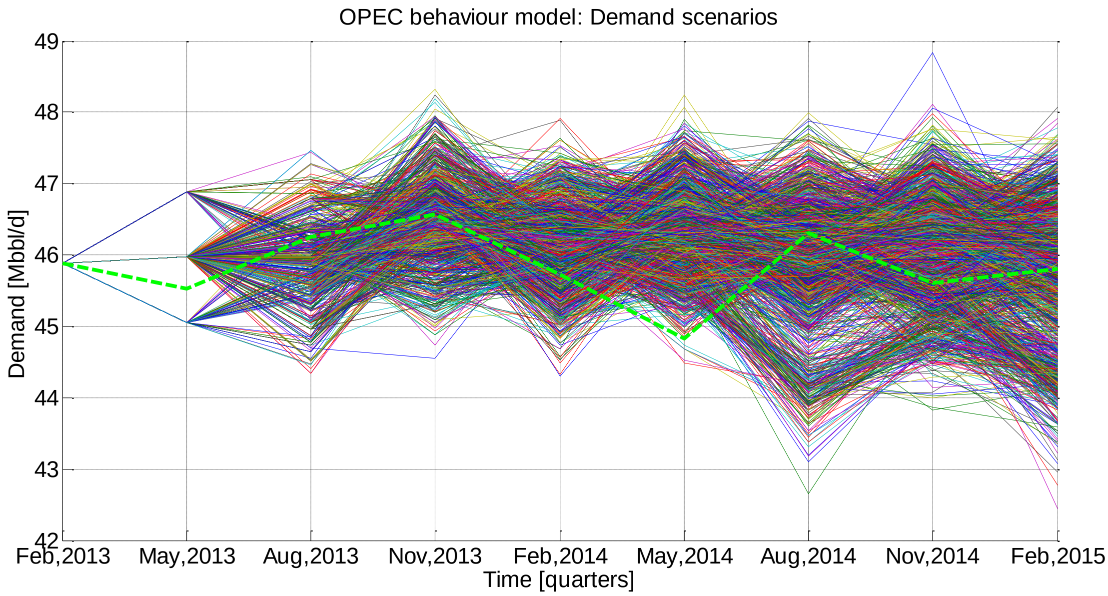


Figure 106 - Fully-predictive scenarios of OECD quarterly demand from Feb, 2013 to Feb, 2015. The green dotted line represents the real demand (real data from EIA).

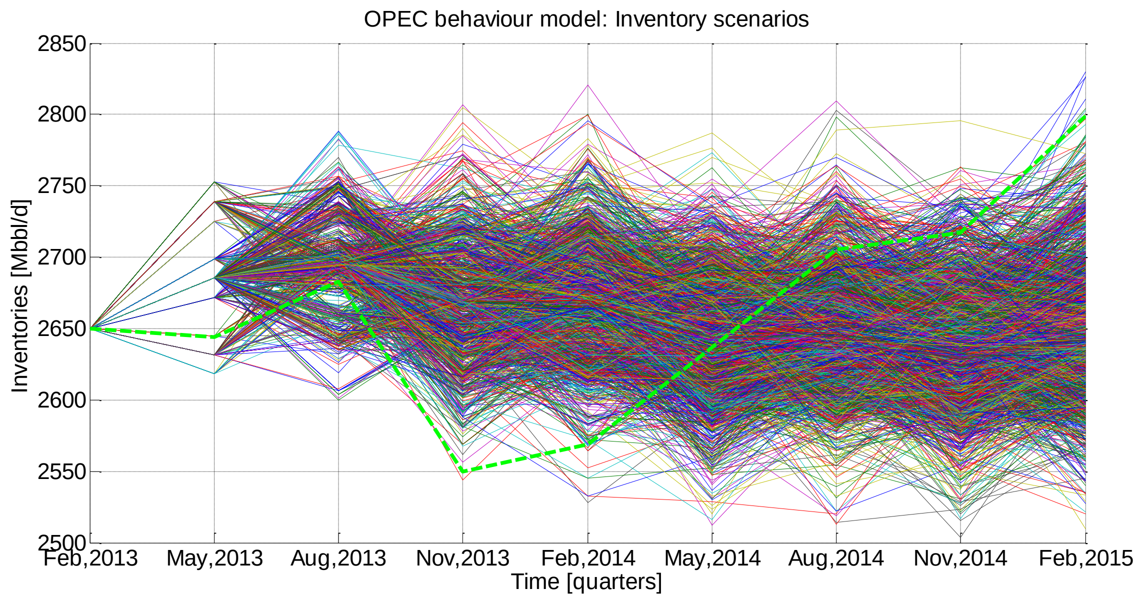


Figure 107 - Fully-predictive scenarios of OECD quarterly inventories from Feb, 2013 to Feb, 2015. The green dotted line represents the real inventories (real data from EIA).

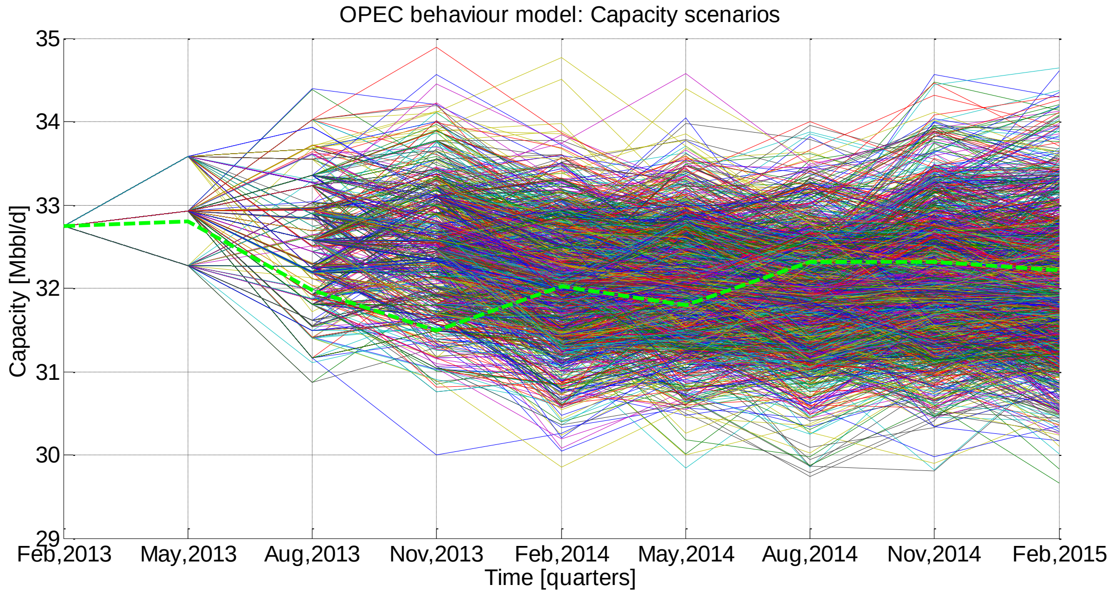


Figure 108- Fully-predictive scenarios of OPEC production capacity from Feb, 2013 to Feb, 2015. The green dotted line represents the real capacity (real data from EIA).

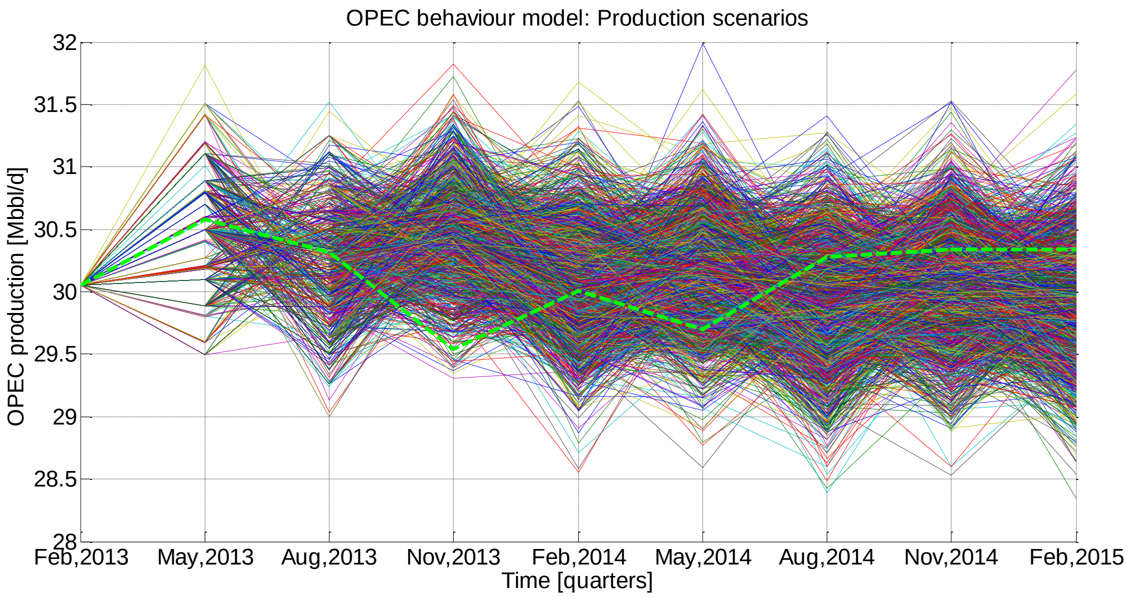


Figure 109 - Fully-predictive scenarios of OPEC production from Feb, 2013 to Feb, 2015. The green dotted line represents the real OPEC production (real data from EIA).

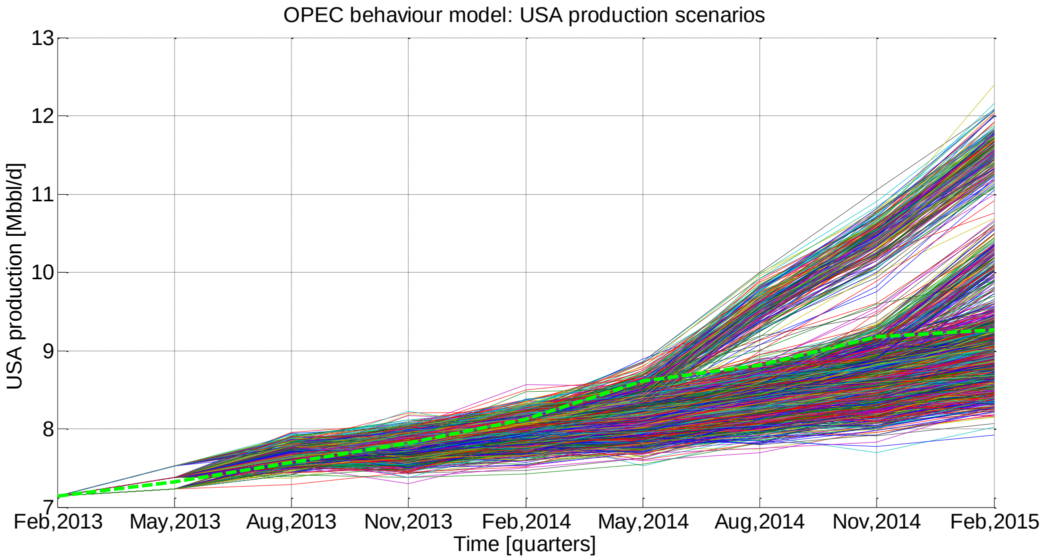


Figure 110 - Fully-predictive scenarios of USA production from Feb, 2013 to Feb, 2015. The green dotted line represents the real USA production (real data from EIA).

5.3 Model simulations: the creation of fully predictive scenarios

In order to use the model to forecast future CO prices according to the geographical position of the chemical plant/refinery that a process designer/manager wants to either schedule or plan, this Section employs the CO price model to simulate fully-predictive scenarios over two-year time span in the future (*i.e.* from the second quarter of 2015 to the second quarter of 2017 where no experimental (*i.e.* real) values of CO quotations are available as, at the time of finalizing this thesis, such a time interval is actually in the future).

First of all, it is worth underling how the model can simulate possible variations of both supply-and-demand variables. Furthermore, the recent historical events and market trends (see Chapter 1) have influenced the analysis conducted in the present work. For instance, the transitory geopolitical situation has the highest short-term risk to US economy and CO price, according to 40% of economists (a percentage higher than 19% since March 2015, according to Liesman, 2015). The model contains neither war premium nor terror premium variables, but takes into account the stochasticity that is intrinsic in these events by means of some standard deviation considerations. Since aim of PSE/CAPE forecasting activities is to get a trend of the future prices according also to different geopolitical and economic scenarios, it is worth underlining that the frequency of political events (*e.g.*, wars, tensions, local crisis) and the standard deviation of petroleum prices should be taken into account when shaping the stochastic features of the model. The variation contribution is calculated by means of the coefficient of variation, which can be estimated by using the ratio of the sample standard deviation to the mean of the data set. In particular, by analyzing the CO trend, it is possible to quantify that the standard deviation (σ) of WTI CO price from the first quarter of 2011 (*i.e.* Feb, 2011) to the second quarter of 2014 (*i.e.* May, 2014) was 5.6%, while Brent CO price had a σ of 3.8%. In order to catch the more recent variations of CO market, and the political and historical events, the coefficient of variation until the last quarter of 2014 (*i.e.* Nov, 2014) was 8.4% for WTI and 9% for Brent. These are the values that are considered in the fully-predictive simulations. By analyzing the historical time series of CO quotations and the periodicity of the occurrences influencing prices, it was decided to take different frequencies for major/minor increases/decreases of CO prices due to technical, economic, and historical events of different extents. At our knowledge, scientific literature does not report this information. The magnitude of increases and decreases depends on the normal-shock assumption that was demonstrated in Chapter 3 (see Section 3.2). Indeed, major increases/decreases are assumed equal to $\pm 3\sigma$ because 99.73% of the normal distribution values lie within three standard deviations of the distribution center (*i.e.* the mean) and $\pm 3\sigma$ values represent major shocks that occur exceptionally. Instead, minor increases/decreases are more frequent than major raises/falls and assume arbitrarily $\pm 1.5\sigma$ values by observing historical data series of CO quotations. Major increases of the price ($+3\sigma$) were due to major variations (*e.g.*, war, attack, crash) and were characterized by a 10-year period (*i.e.* 40 quarters). Minor increases of the price ($+1.5\sigma$) were due to minor variations (*e.g.*, tension or growth of China) and were characterized by a 3-year period (*i.e.* 12 quarters). Minor decreases of the price (-1.5σ), due to minor variations (*e.g.*, economic or social issues or technical defects), had a frequency of about 7 years (*i.e.* 28 quarters). Major price reductions (-3σ), due to major economic/financial instabilities, were rather rare and had a frequency of about 10 years (*i.e.* 40 quarters). The frequency of variations of the input variables is included in the model, too. By analyzing past values of the input variables it was decided to take a 8-quarter periods of variations for demand, 8 quarters for inventories, 7 quarters for production, 7 quarters for production capacity, and 16 quarters for USA production. Figure 111 and Figure 112 show the results of 3000 simulations of future Brent and WTI trends for a 2-year horizon. These simulations couple the deterministic contribution of the terms present in Equation (40), which are individually modeled by Equations (45-46-50-51-54), and the abovementioned stochastic contributions. The same initial condition (*i.e.* first quarter of 2015) characterizes both Brent and WTI trends. A bullish trend of the GDP is supposed, with a 2% annual increase. CO prices are not limited to any ranges, but if the lower/upper thresholds of 40 USD/bbl and 100 USD/bbl are violated for some quarters (*e.g.*, two consecutive quarters) the forecast values are suitably corrected by changing quotas to keep the future quotations within that expected interval. Indeed, OPEC bets on stabilization for several reasons. A too low price of CO (less than 40 USD/bbl) would produce serious consequences on the producing countries (OPEC) that need to maximize profits from their investments. On the other side, an excessive cost of the barrel would penalize the consumers (OECD), decrease the demand, and favor the exploitation of other sources of energy. For the sake of correctness, it is worth underling that the lower threshold was modified recently (from the original value of 80 USD/bbl) after the OPEC decision, taken in December 2014, of not cutting members' production even if quotations have been under the breakeven point of some OPEC nations such as Iran and Venezuela.

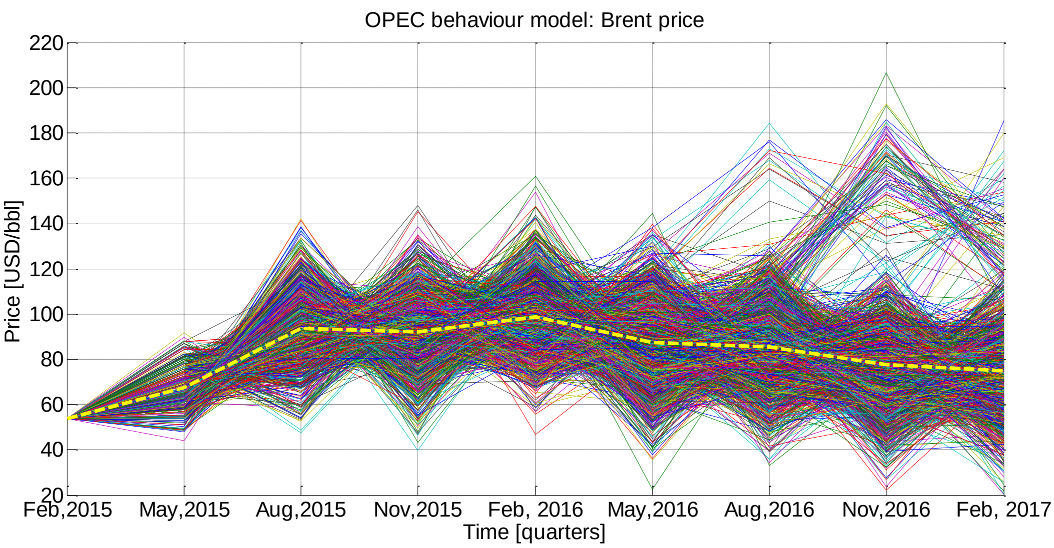


Figure 111 - Fully-predictive scenarios of Brent quarterly quotations from Feb, 2015 to Feb, 2017 (3000 simulations over 8 quarters). The yellow dotted line is the average forecast price.

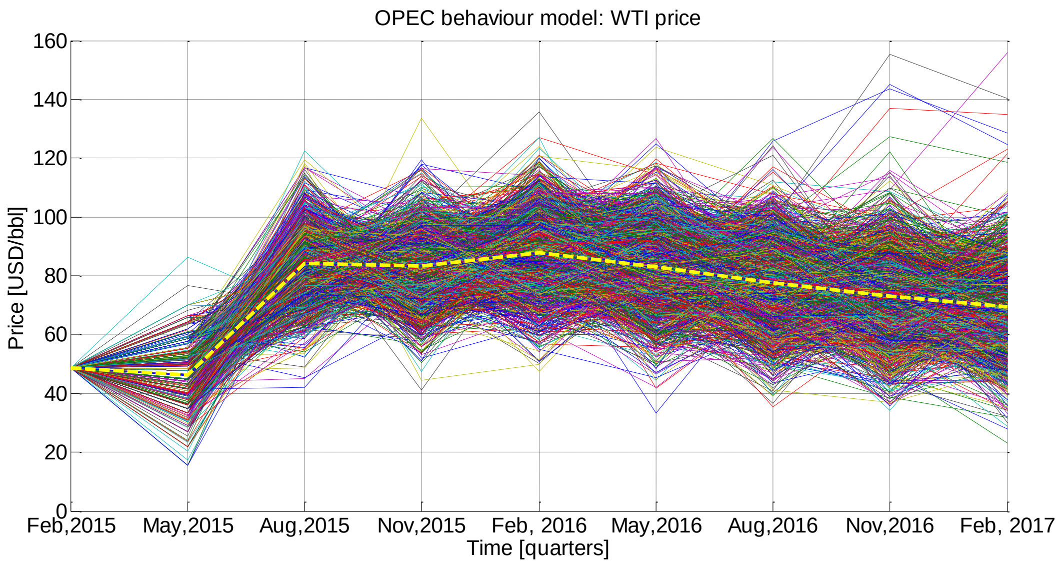


Figure 112 - Fully-predictive scenarios of WTI quarterly quotations from Feb, 2015 to Feb, 2017 (3000 simulations over 8 quarters). The yellow dotted line is the average forecast price.

Figure 113, Figure 114, Figure 115, Figure 116, and Figure 117 show the input variable scenarios on both supply and demand sides (for the of conciseness, only Brent simulations are reported).

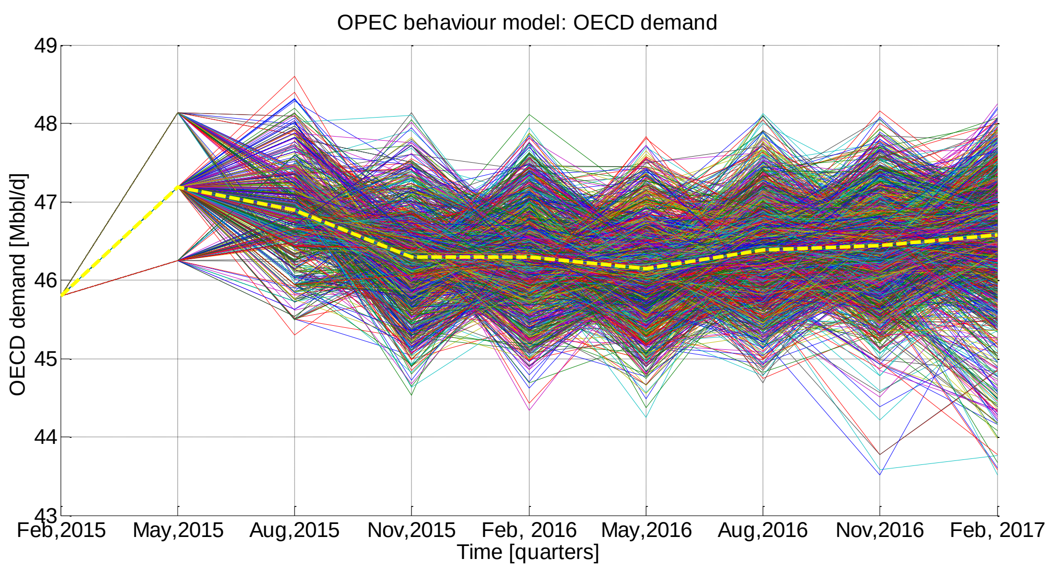


Figure 113 - Fully-predictive scenarios of OECD quarterly demand from Feb, 2015 to Feb, 2017 (3000 simulations over 8 quarters). The yellow dotted line is the average forecast demand.

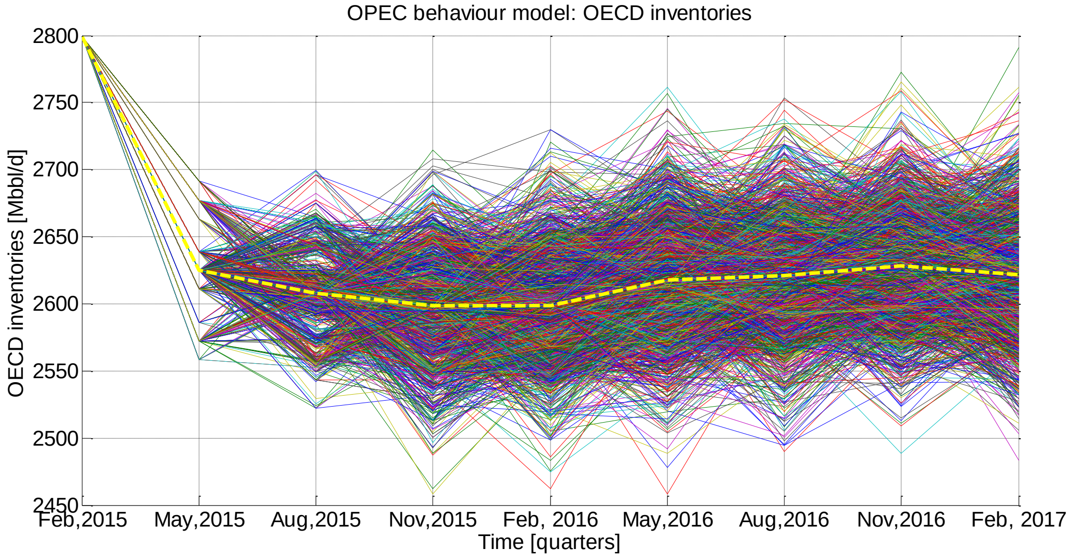


Figure 114 - Fully-predictive scenarios of OECD quarterly inventories from Feb, 2015 to Feb, 2017 (3000 simulations over 8 quarters). The yellow dotted line is the average forecast inventories.

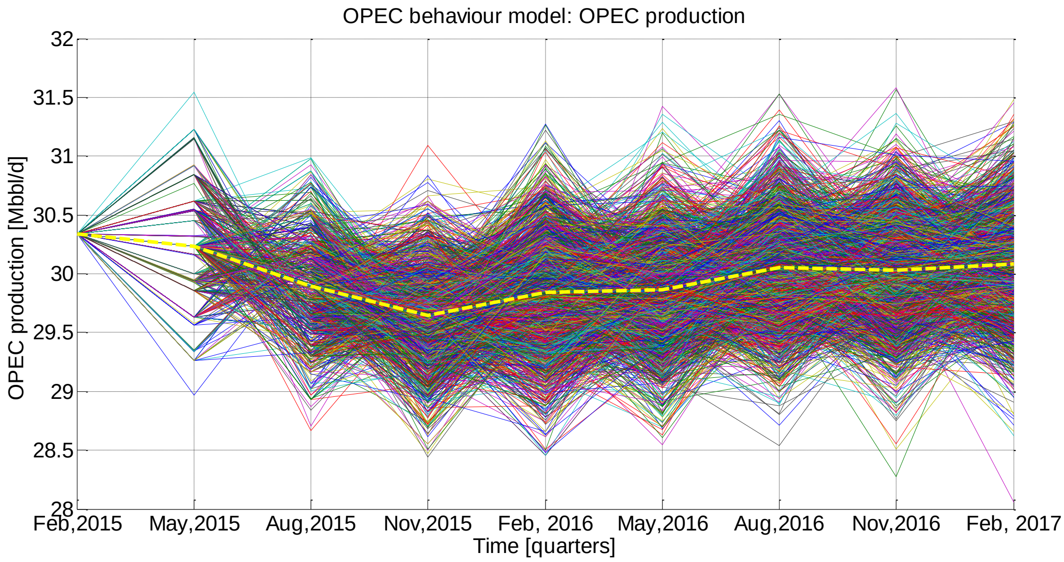


Figure 115 - Fully-predictive scenarios of OPEC quarterly production from Feb, 2015 to Feb, 2017 (3000 simulations over 8 quarters). The yellow dotted line is the average forecast production.

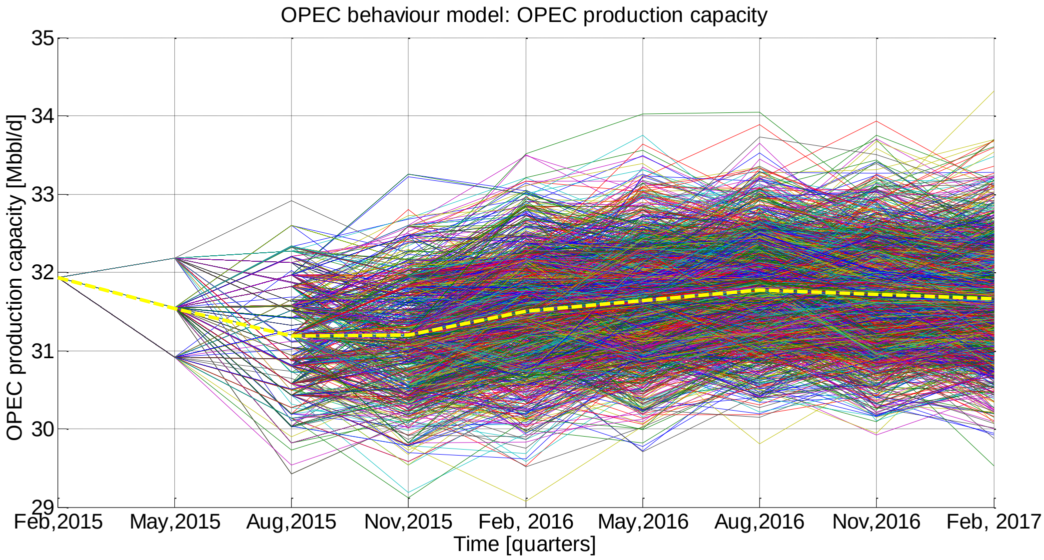


Figure 116 - Fully-predictive scenarios of OPEC quarterly production capacity from Feb, 2015 to Feb, 2017 (3000 simulations over 8 quarters). The yellow dotted line is the average forecast production capacity.

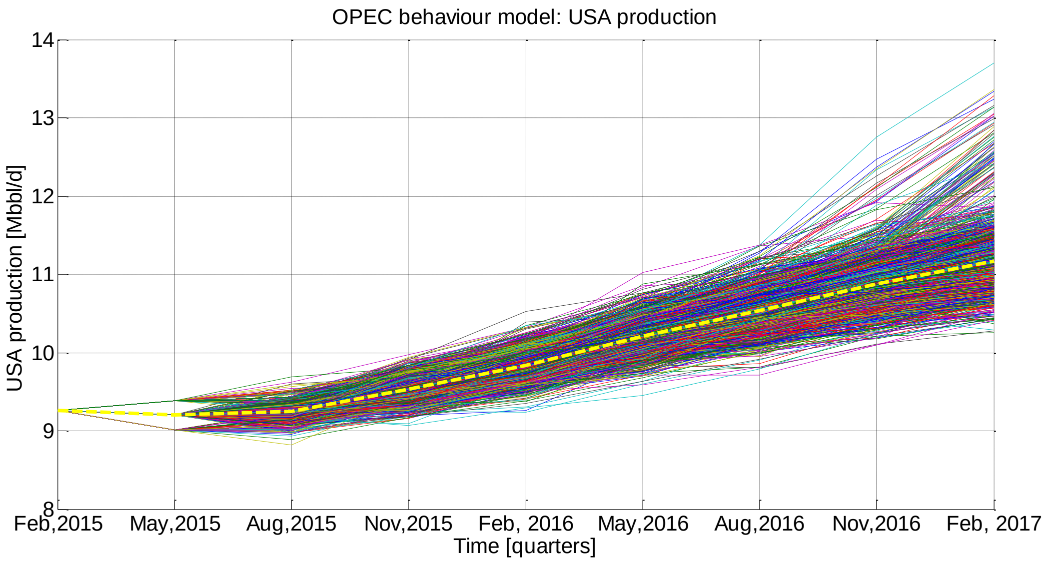


Figure 117 - Fully-predictive scenarios of USA quarterly production from Feb, 2015 to Feb, 2017 (3000 simulations over 8 quarters). The yellow dotted line is the average forecast production.

With reference to Figure 111 and Figure 112, the global trend of petroleum prices seems to stabilize between 30 and 110 USD/bbl for WTI and between 20 and 160 USD/bbl for Brent. Indeed, the forecast values are probabilistically higher for Brent than WTI, as it happens for the average scenario quotations (Figure 118). Again, it is worth underlining that the scenarios characterized by very high prices are rather few and consequently have a reduced impact on the distribution of possible future highly bullish trends.

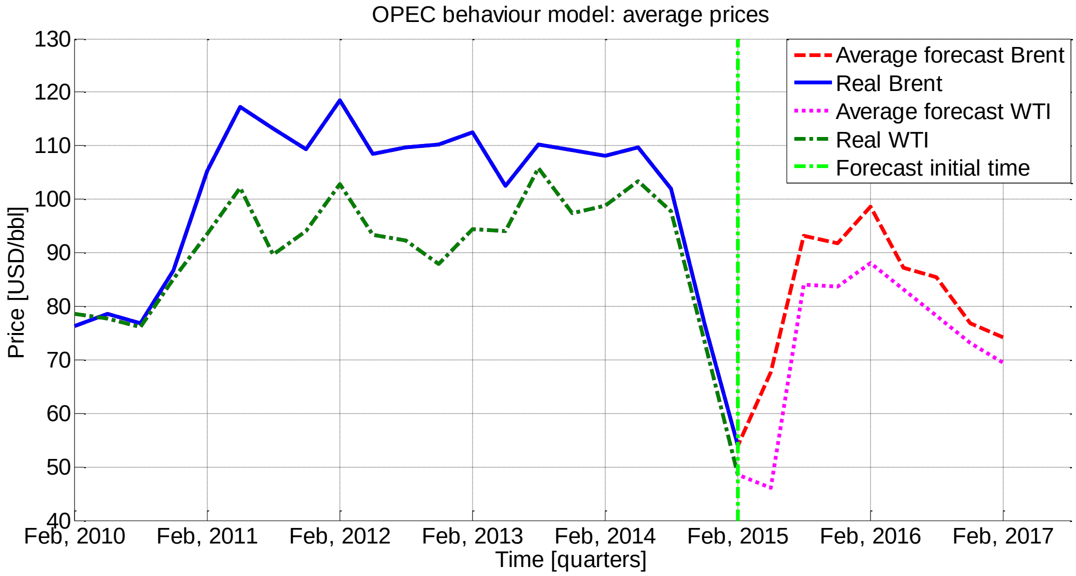


Figure 118 - Brent and WTI average forecast prices start from the vertical dashed green line (Feb, 2015).

As already discussed in Section 3.2, fan-charts show ranges for possible values of future price data (see Figure 119 and Figure 120). As predictions become increasingly uncertain going deeper into the future, those ranges of forecast scenarios spread out and create the so-called *river of blood*. The various future intervals are pictorially represented by varying the shades (*i.e.* darker near the center of the range, fainter near the tails) according to the probability associated with that specific price range. The confidence intervals chosen are 0.1%, 5%, 10%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 90%, 95%, and 99.9%. The last percentage is intentionally selected very high in order to show that quotations higher than 90 and 100 USD/bbl are probabilistically not significant (*i.e.* negligible). Furthermore, Table 19 shows the number of scenarios that in the first quarter of 2017 finish with a quotation in a specific price range. The same scenarios are shown in Figure 121 and Figure 122 for Brent and WTI, with different colors according to the price interval that characterizes the last forecast quotation. Brent quotations are more volatile, as it can be anticipated by the coefficient of variation.

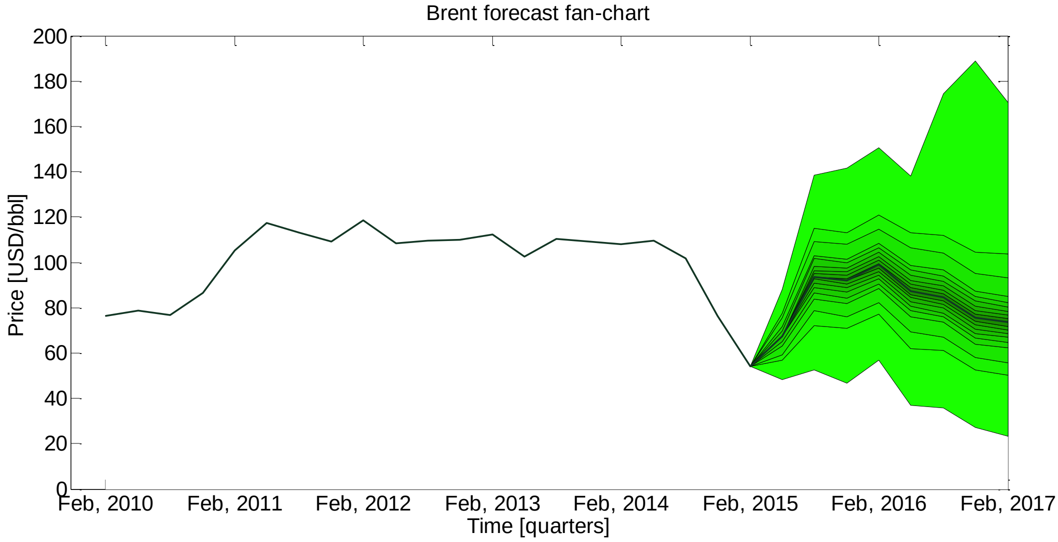


Figure 119 - Fan-chart of WTI quarterly prices from Feb, 2010 to Feb, 2017: the period from Feb, 2015 to Feb, 2017 represents the forecast range (3000 different simulations) with probability from 0.1% (dark green) to 99.9% (light green).

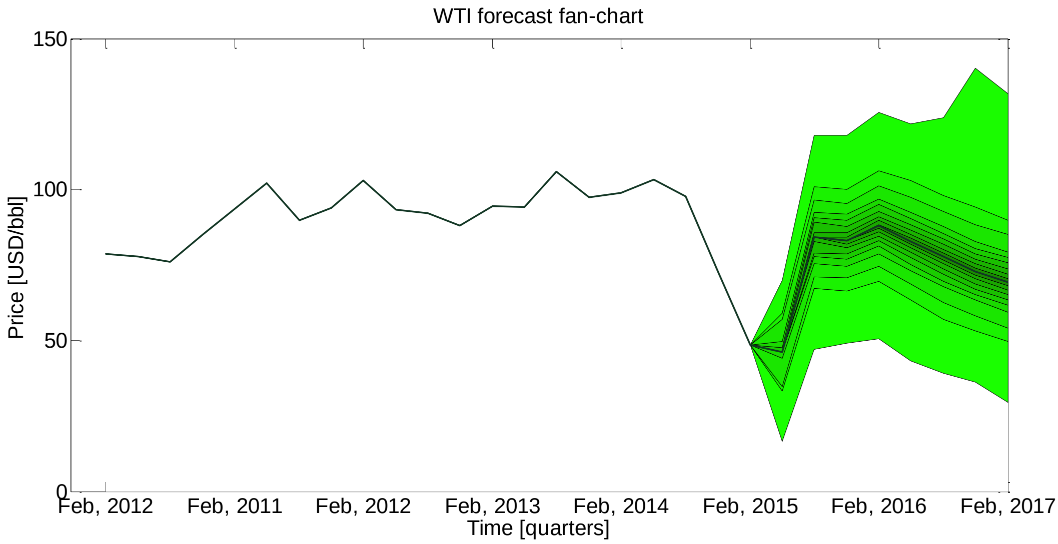


Figure 120 - Fan-chart of WTI quarterly prices from Feb, 2010 to Feb, 2017: the period from Feb, 2015 to Feb, 2017 represents the forecast range (3000 different simulations) with probability from 0.1% (dark green) to 99.9% (light green).

Table 19 - Number of scenarios belonging to different price ranges and their relative percentages for both Brent and WTI forecast quotations.

Price range [USD/bbl]	Scenarios	
	Brent	WTI
$P \geq 150$	18 (0.6%)	1 (0.03%)
$120 \leq P < 150$	62 (2.07%)	6 (0.2%)
$90 \leq P < 120$	315 (10.5%)	141 (4.7%)
$60 \leq P < 90$	2127 (70.9%)	2216 (73.87%)
$30 \leq P < 60$	470 (15.67%)	632 (21.07%)
$P < 30$	8 (0.27%)	4 (0.13%)

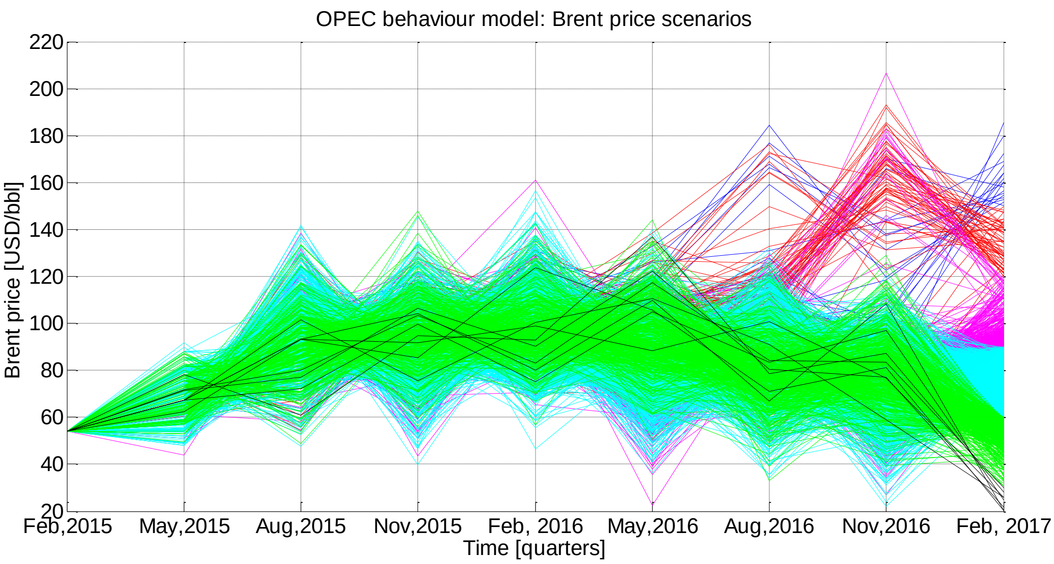


Figure 121 - Brent quarterly prices from Feb, 2015 to Feb, 2017 that finish in the different price ranges described in Table 11. The price ranges are in ascending order for black, green, cyan, magenta, red, and blue lines.

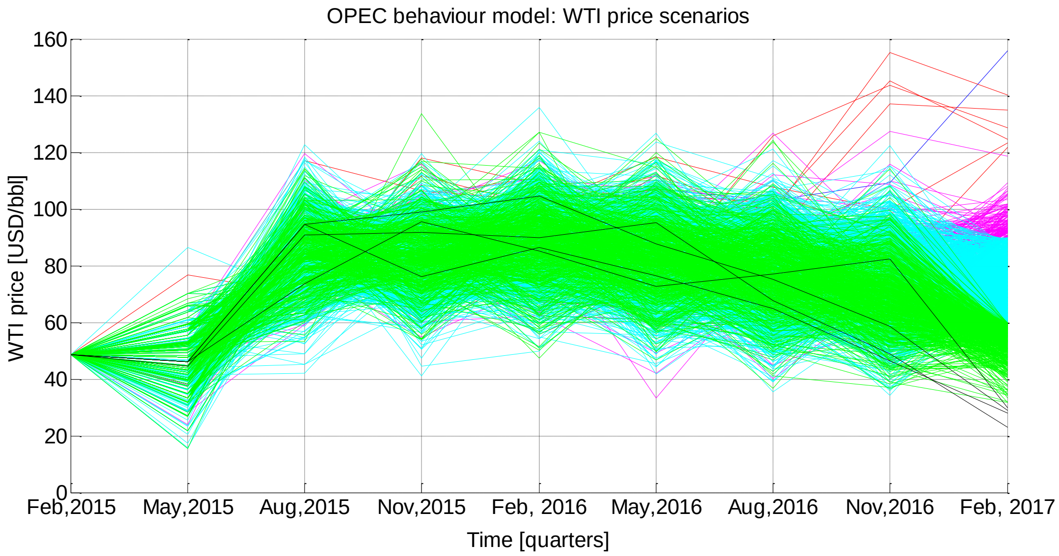


Figure 122 - WTI quarterly prices from Feb, 2015 to Feb, 2017 that finish in the different price ranges described in Table 11. The price ranges are in ascending order for black, green, cyan, magenta, red, and blue lines.

Despite the unpredictable events that can influence market quotations and force to often update the model, a PSE/CAPE-oriented approach remembers past bull-and-bear periods and tends to stabilize the trend of CO prices in the range between 30 and 120 USD/bbl.

5.4 Model manipulation: bullish, bearish, and "neutral" scenarios

This Section is dedicated to create several price scenarios with an overall bullish, bearish, and conservative (*i.e.* neutral) behavior. Indeed, by means of a sensitivity analysis and the considerations about adaptive parameter signs of the input variables, it is possible to manipulate the new CO price model for manufacturing and management purposes. The OPEC-based model described by Equations (40-45-46-50-51-54) simulate the fluctuations in CO market by means of supply-and-demand variables that account for real market demand, global supply, possible offer disruption or oversupply conditions. The user (*e.g.* process designer/manager) can modify the shock values, variation frequencies, or values of input variables to simulate the conditions where a plant/refinery likely will be scheduled or planned. This is the main advantage of the proposed model respect to econometric models and existing economic ones. In particular, bullish scenarios are created by considering a quarterly constant increase of 0.5% of demand and OPEC production, and a 0.5% decrease of OECD inventories, OPEC production capacity, and USA production. On the contrary, a decrease of demand and OPEC production, and an increase of inventories, production capacity, and USA production cause CO prices to fall (*i.e.* bearish scenarios). An additional positive feature of the proposed model consists in manipulating the involved variables in order to create possible future scenarios that take into account economic development, crisis (by means of GDP and demand), supply disruption (based on OPEC production, production capacity, and OECD inventories), or growth of other sources of oil (*e.g.* shale oil that contributes to the USA production).

Figure 123, Figure 124, and Figure 125 show the collections of fully-predictive scenarios featuring average trends (yellow dotted line) that are respectively bullish, bearish, and neutral. A bullish or bearish trend is characterized respectively by on average upward or downward trends, even if locally key reversal points may exist.

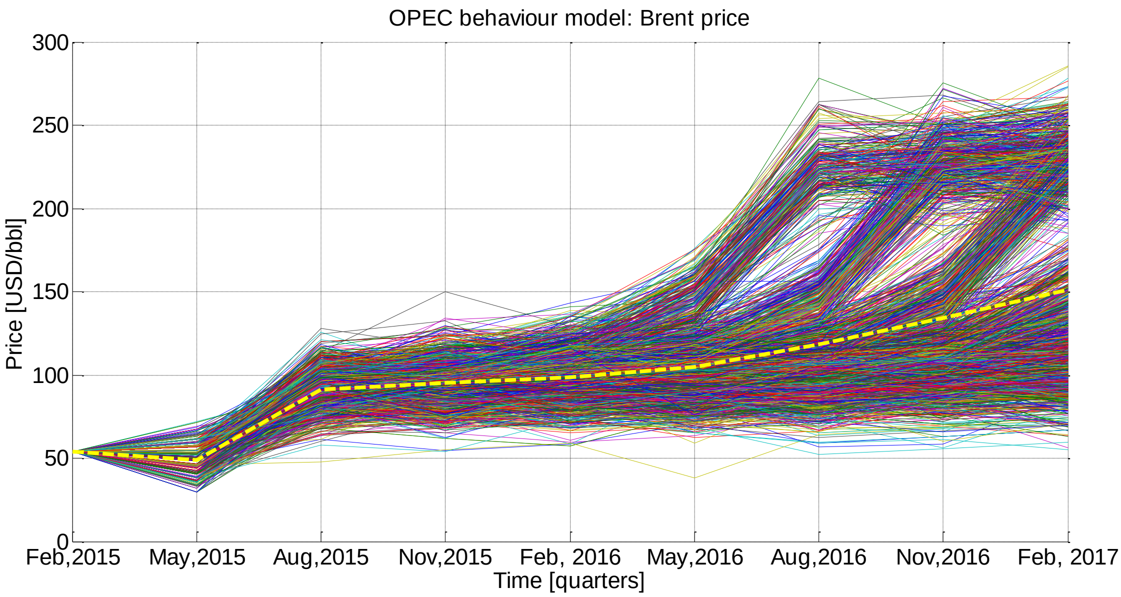


Figure 123 - Fully-predictive scenarios of Brent quarterly quotations from Feb, 2015 to Feb, 2017 with an overall bullish trend (3000 simulations). The yellow dotted line represents the average forecast price.

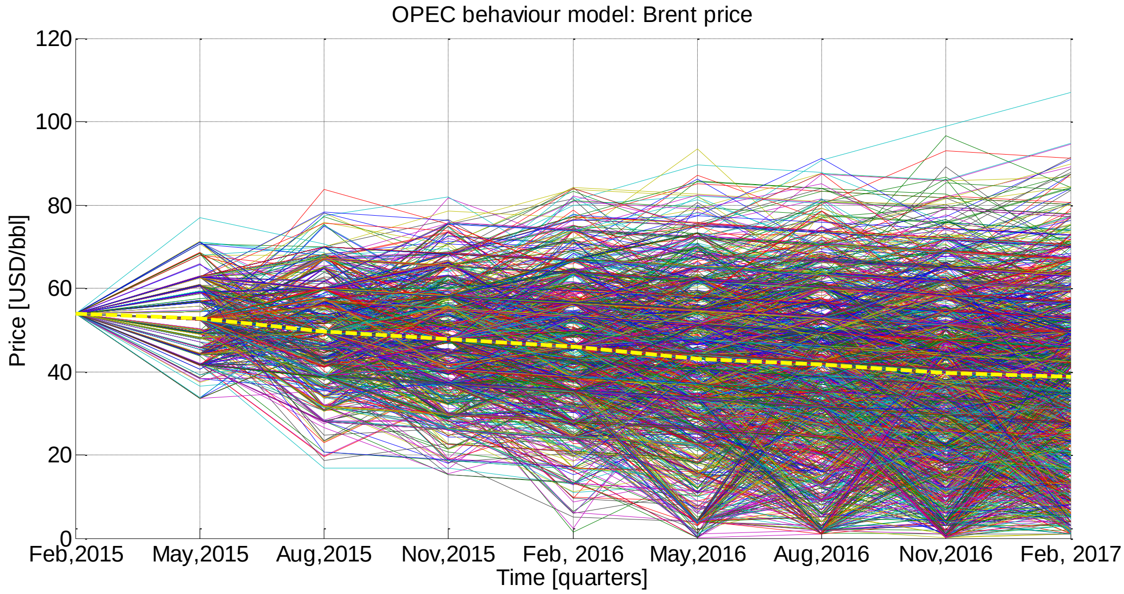


Figure 124 - Fully-predictive scenarios of Brent quarterly quotations from Feb, 2015 to Feb, 2017 with an overall bearish trend (3000 simulations). The yellow dotted line represents the average forecast price.

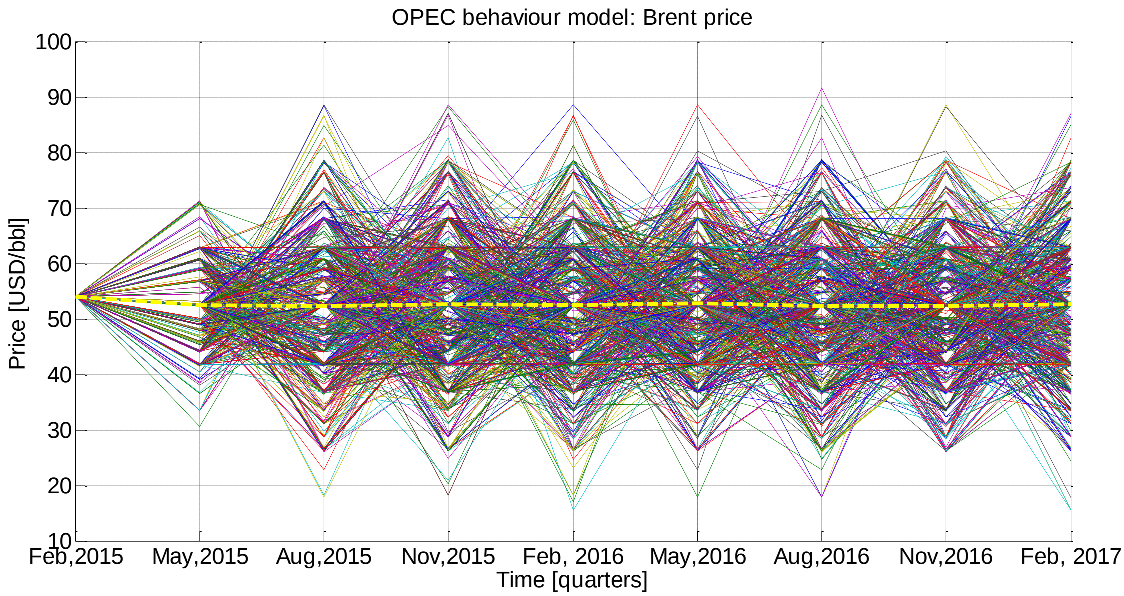


Figure 125 - Fully-predictive scenarios of Brent quarterly quotations from Feb, 2015 to Feb, 2017 with an overall constant trend (3000 simulations). The yellow dotted line represents the average forecast price.

It is worthy observing that these scenarios are created in order to show the power of the proposed model for commercial purposes and industrial applications. On one hand, it can be used to forecast future CO prices based on the current supply-and-demand condition (*i.e.* by means of the adaptive parameters), and on the other hand, it can be manipulated by the process designers/managers to simulate precise situations (*e.g.* economic assessment, feasibility study, revamping, retrofitting, scheduling, planning, supply chain). Figure 123 shows also scenarios that are probabilistically quite rare (*i.e.* higher than 150 USD/bbl), but they have just the purpose of showing the average-upward trend. In addition, the first-step simulation appears to be still affected by the recent bearish prices trend.

Another interesting remark can be made about the impact of GDP, which influences the demand and has an impact on the long-term forecasts, but does not affect significantly either short- or middle-term horizons. Figure 126 allows maintaining that the lower the GDP the slightly-higher the price average.

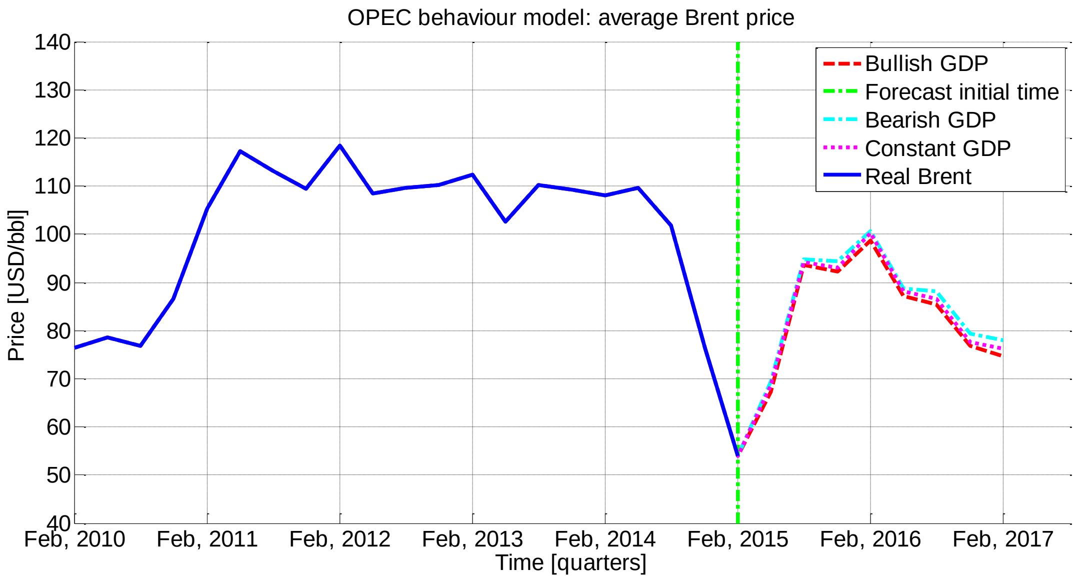


Figure 126 - Brent and WTI average forecast prices start from the vertical dashed green line (Feb, 2015). The red, magenta, and cyan line correspond to 2% GGDP annual bullish variation, GGDP constant annual trend, and 2% GGDP annual bearish variation, respectively.

Chapter 6 Development and simulation of crude oil price hybrid model

This Chapter describes a new procedure that allows forecasting the CO price by combining the supply-and-demand law, involved in OPEC-based model, with a new econometric model. The econometric model works with monthly prices and features a moving-average approach to quotations. The difference of time-granularity between the economic and econometric models is overcome by linear interpolation of the data set provided in the economic model. The new hybrid model allows creating different price scenarios with a background noise that characterizes market prices. This is a rather promising tool for the optimal management of production sites as a function of the real market demand, global supply, market uncertainties, and historical balances/imbalances.

6.1 Why a *hybrid* model

The term *hybrid* comes from the Latin word *hibrida* and designates something that is obtained from mixing two or more different things. In this thesis, *hybrid* signifies a model that derives from the combination of econometric models with economic ones, and that can simulate the trend originated by supply-and-demand law in combination with the stochastic fluctuations of CO quotations.

The creation of the new hybrid model calls for two model classifications proposed in Chapter 2. First of all, the hybrid model combines both econometric and economic features. As already discussed, these model categories are different for both the structure/mechanism used to predict the prices and the type of fundamentals provided about their trend (*i.e.* physical, economic, and financial features for economic model and statistical analysis of past price shocks for econometric models). In particular, the main advantage of economic models (such as the OPEC-based one) is that they involve supply-and-demand variables, which can be manipulated in order to create future scenarios, with an overall bullish or bearish trend, and simulate possible demand crisis, situations of oversupply, or economic and technological developments. These characteristics make the OPEC-based model interesting also for commercial purposes, because it can be implemented for scheduling, planning, and feasibility studies under market uncertainties and fluctuating factors. The feature of the OPEC-based model that can be improved is the time-granularity of the input variables and so of the predicted scenarios. Both price and input data of the OPEC-based model have a quarterly time granularity, as this is the frequency of political decision about OPEC quotas and the availability of supply-and-demand variables in the EIA and IEA databanks. Conversely, econometric models may feature daily, weekly, or monthly discretizations. Even if they are not intended to follow the forces that cause price fluctuations, the econometric models can catch the oscillations that characterize CO quotations. The graphs collected in Chapter 5 do not have the swinging trend that instead the reader can observe in common trading websites. The OPEC-based model can provide bullish-, bearish-, and conservative-trend scenarios, with frequent key reversal points between adjacent quarters. At the same time, an economic model can predict data that are linked to supply-and-demand variations, *i.e.* that category of models is based on real, objective, and economic/financial/political indicators. This new forecast data-set is fed to the econometric model, which can reproduce (*i.e.* simulate) also the fluctuations inside the quarters, with either a weekly or more commonly monthly granularity.

At our knowledge, this procedure is innovative and there are not any hybrid models in the literature capable to combine the stochastic fluctuations to the supply-and-demand fluctuations. In particular, Barzaghi and Conte (2015) led a deeper analysis of Brent and WTI price shocks in their thesis and developed a new econometric model to forecast the CO quotations. That econometric model is used in this Chapter and is concisely described in the next Section. For sake of brevity, the simulation results are reported only for Brent prices (which are good for industrial processes/plants to be operated in Europe).

6.1.1 Econometric model description

The analysis of price shocks is extensively analyzed and discussed in Barzaghi and Conte (2015). Indeed, if one performs an analysis of the relative price variations between a time unit (*e.g.* one month) and the next one, a set of values that are representative of the price volatility is obtained. Fini and Oliosi (2010) demonstrated that these variations are normally distributed and can be assimilated to a stochastic variable by means of the autocorrelogram analysis. Fini and Oliosi (2010) showed that these attributes characterize a typical Markovian process (whose status does not depend from previous historical quotations but varies stochastically starting from the last price). On the contrary, Barzaghi and Conte (2015) arrived to opposite deductions. Indeed, they demonstrated that the moving average trend of CO prices is not a Markovian process. As a matter of fact, Fini and Oliosi (2011) worked on real spot values of CO prices, whilst Barzaghi and Conte (2015) used moving-averaged values (with four-month time spans). By doing so, they eliminated most of the high-frequency fluctuations of CO prices that can be assimilated to a background noise, which, on its turn, is responsible for the Markovian nature of a time series. Actually, Barzaghi and Conte (2015), by analyzing the autocorrelogram of CO price shocks, concluded that CO prices depend on the quotations of the two previous months (Figure 127).

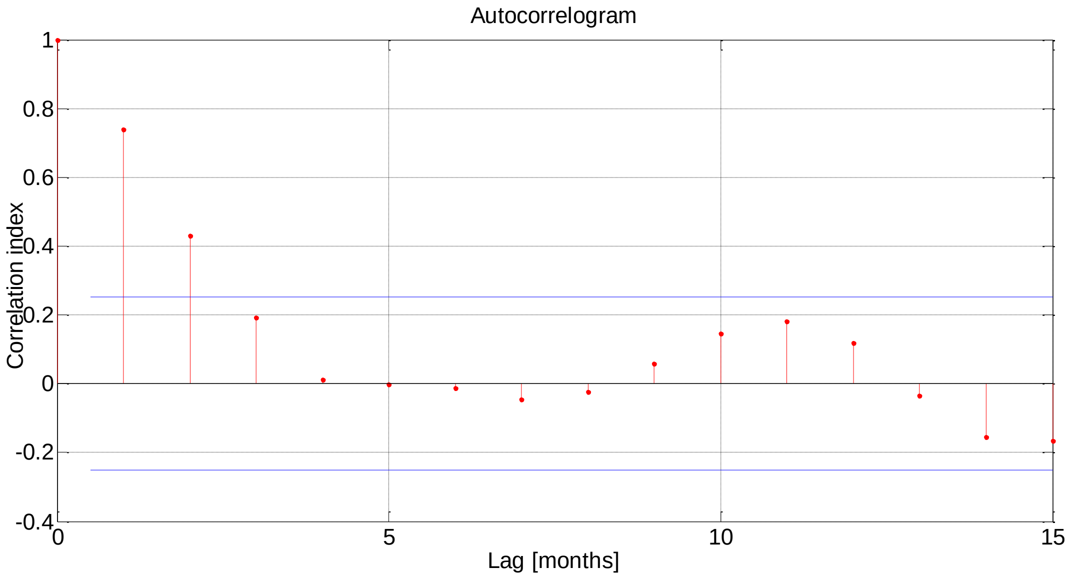


Figure 127 - Autocorrelogram of Brent moving average shocks on monthly basis (data from Jan, 2010 to Jan, 2015).

Based on the autocorrelogram reported in Figure 127, the corresponding econometric model can be formulated as follows:

$$PRICE_t = A + B \cdot PRICE_{t-1} + C \cdot PRICE_{t-2} \tag{67}$$

where A, B, and C are adaptive coefficients that are calculated by means of linear regression procedure, while $PRICE_t$, $PRICE_{t-1}$, and $PRICE_{t-2}$ are the CO moving-averaged prices. The characteristic feature of the proposed econometric model is that it can recover the background variations, which characterize CO prices, by means of a deeper analysis of CO shocks as discussed in Barzaghi and Conte (2015). In particular, the shock analysis can be supplemented with the study of variation frequencies of historical data. Figure 128 shows the comparison between moving-averaged Brent prices and Brent prices predicted by Equation (67). As can be seen, the moving-average operator makes smooth the quotation trend and eliminates the background noise respect to real CO quotations. The model follows rather well the moving-averaged prices. The red dotted line is created by means of one-step-ahead model simulation.

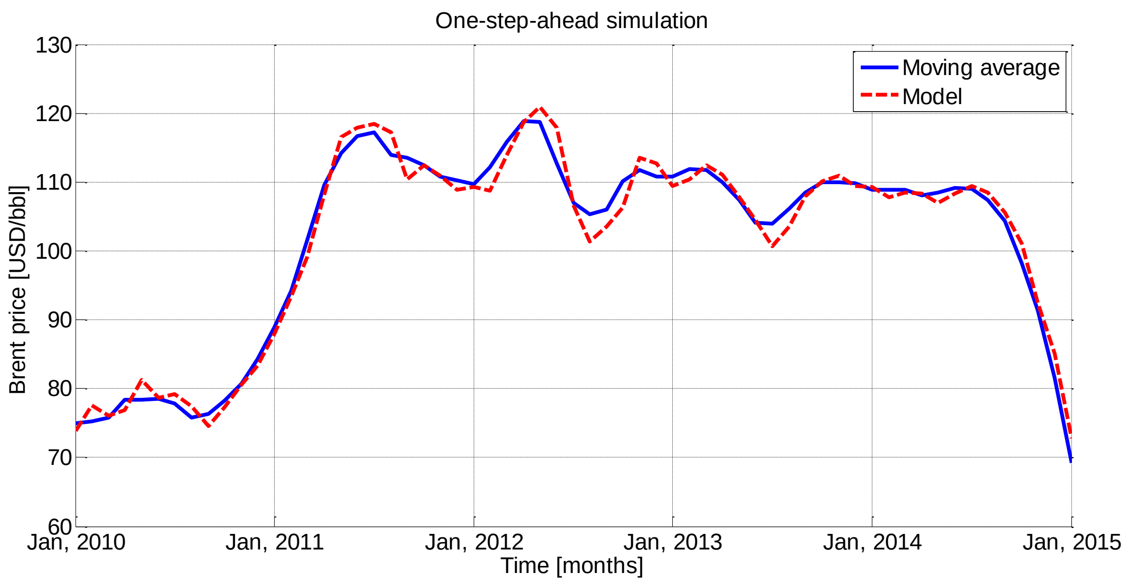


Figure 128 - One-step-ahead simulation of Brent monthly prices from Jan, 2010 to Jan, 2015 and comparison with moving average quotations.

6.2 The Hybrid model

6.2.1 Model description and validation

The new hybrid model combines the OPEC-based model given by Equations (40-45-46-50-51-54) with the econometric model of Equation (67). These models differ in time-granularity (*i.e.* monthly-based for the econometric model and quarterly-based for the economic one). Since the economic model has a quarterly time-granularity, the quotations provided by that model are equaled to the mid-quarter monthly prices (*i.e.* February, May, August, and November), while the remaining three quotations are linearly interpolated. The quotations could also be interpolated by means of parabolas or cubic splines, but such curves would raise the question of which is the best representative of price trend.

As the proposed economic model takes into account the reality by means of the supply-and-demand variables, the OPEC-based model provides pseudo-real quotations to the econometric models, which can simulate price fluctuations. The pseudo-real data provided by the OPEC-based model and linear interpolation are fed to the econometric model that allows determining the adaptive parameters reported in Table 20. These adaptive parameters are only for the average price of the 3000 scenarios collected in Section 5.3 and are calculated by means of the *Matlab*© minimization function *fminsearch*.

Table 20 - Adaptive parameters in Equation (67) for Brent quotations.

Parameter	Value
A	5.116351
B	1.666626
C	-0.73675

Figure 129 shows the comparison between the real moving-averaged Brent prices and the Brent price provided by Equation (67), whilst Figure 130 compares the Brent price predictions simulated with the econometric model in Equation (67) and the hybrid model predictions featuring the background noise. It is evident that the hybrid model adduces fluctuations to quotation trend and follows real price trend (Figure 131).

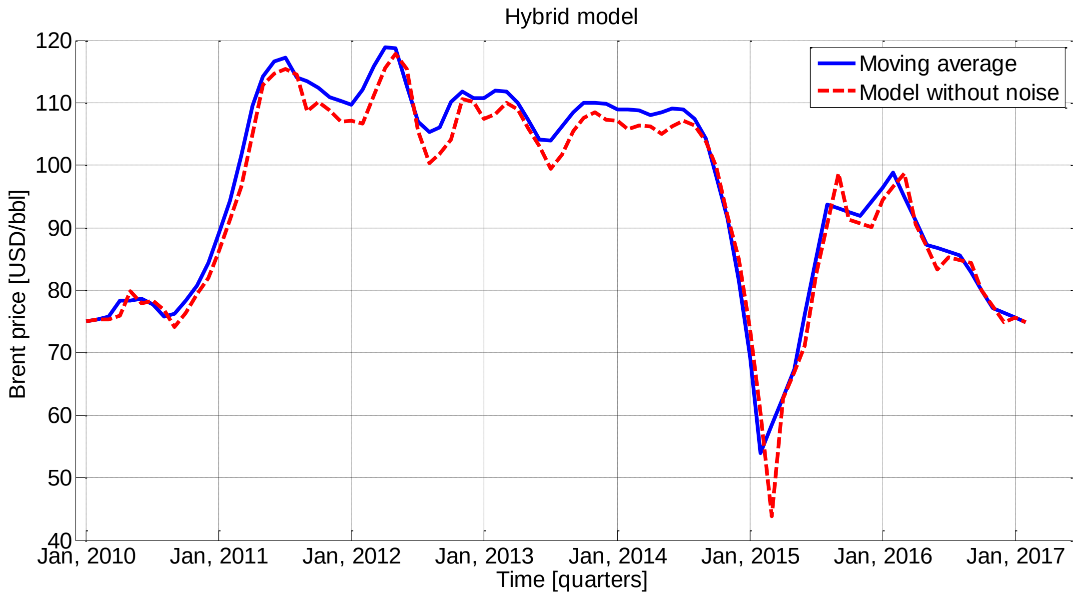


Figure 129 - One-step-ahead simulation of Brent monthly prices from Jan, 2010 to Feb, 2017 without background noise and comparison with moving-averaged quotations. The period between Feb, 2015 and Feb, 2017 collects averaged forecast prices of 3000 simulations.

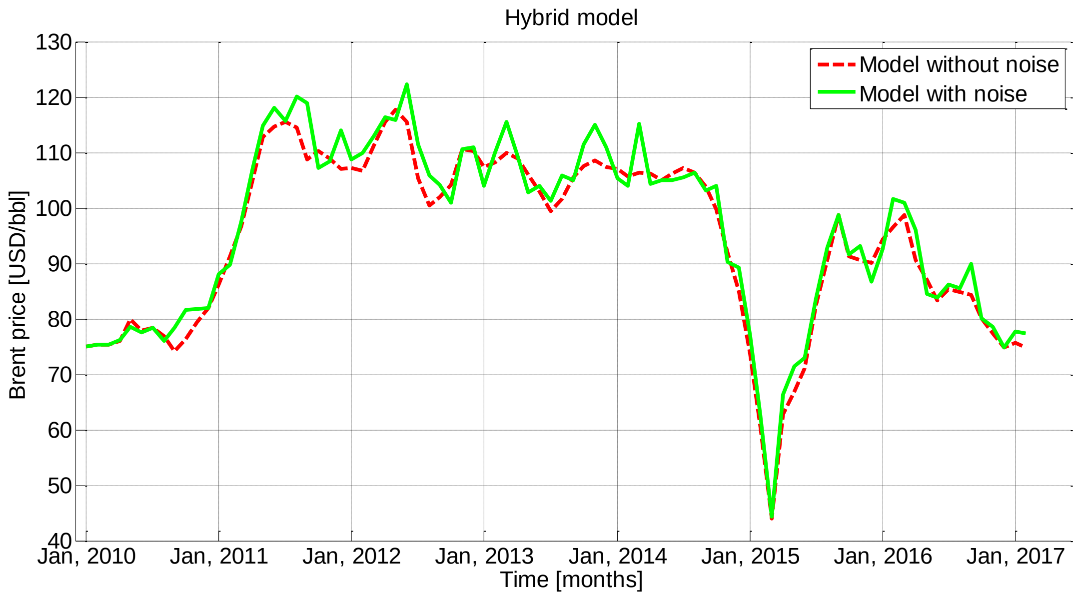


Figure 130 - One-step-ahead simulation of Brent monthly prices from Jan, 2010 to Feb, 2017 without background noise and comparison with the hybrid model. The period between Feb, 2015 and Feb, 2017 represents the forecast horizon.

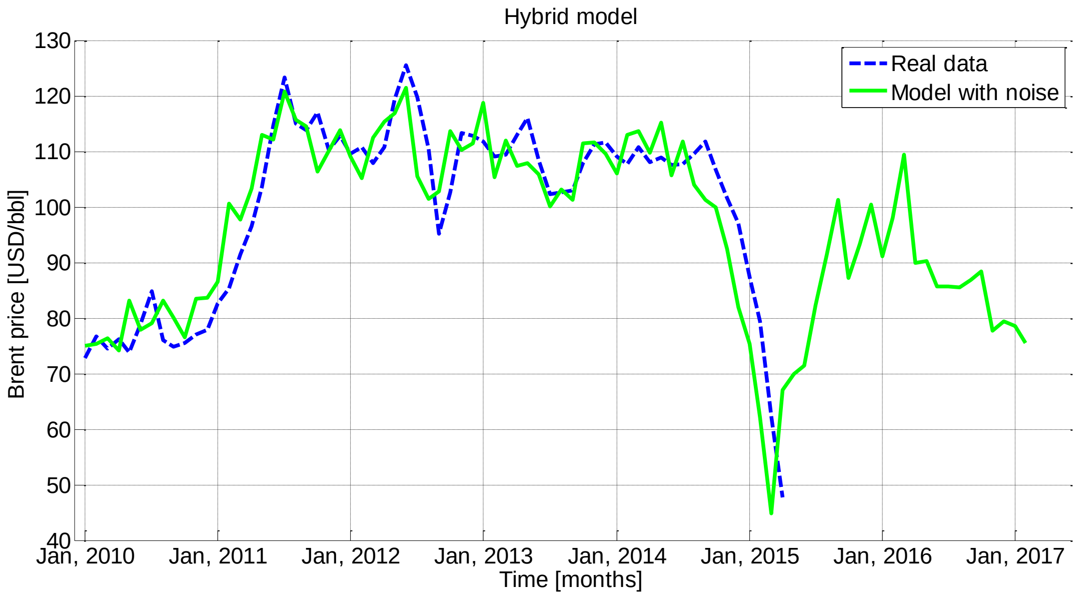


Figure 131 - One-step-ahead simulation of Brent monthly prices from Jan, 2010 to Feb, 2017 by the hybrid model and comparison with real quotations from Jan, 2010 to Feb, 2015. The period between Feb, 2015 and Feb, 2017 represents the forecast horizon.

6.2.2 Model simulation: fully-predictive scenarios

Similarly to the procedure adopted for the economic model, this Section adopts the hybrid model for the creation of a distribution of fully-predictive scenarios over a time horizon of two years in the future (*i.e.* from the second quarter of 2015 to the second quarter of 2017) for either scheduling or planning purposes. The economic model provides a distribution of 3000 different scenarios (see also Chapter 5 for further details). It was decided to apply the same procedure described in Section 6.2.1 for each scenario. Figure 132 allows observing that the price scenarios of the hybrid model have acquired a background noise that was originally missing in the quarterly data of the OPEC-based economic model.

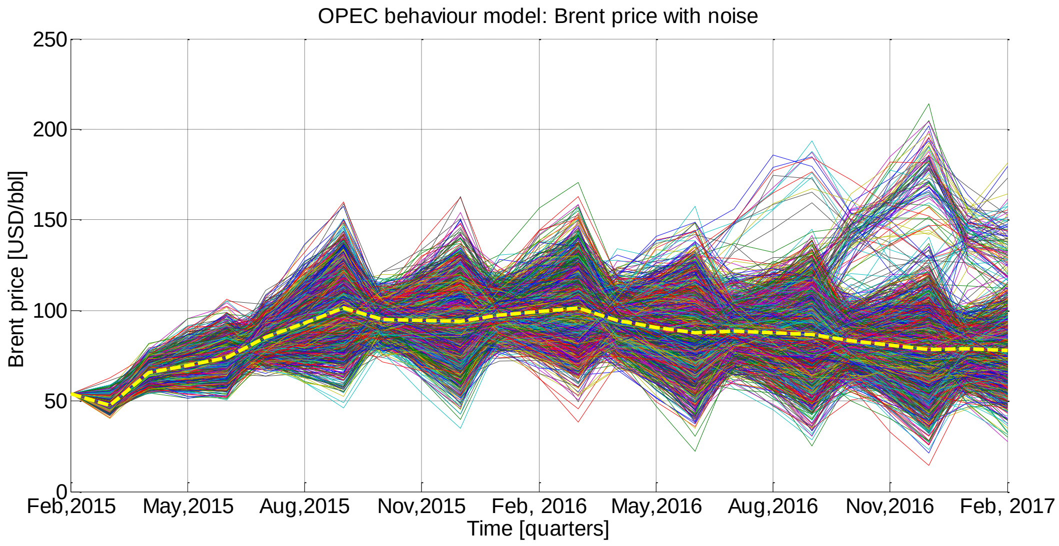


Figure 132 - Fully-predictive scenarios of Brent monthly quotations from Feb, 2015 to Feb, 2017 by the hybrid model (3000 simulations). The yellow dotted line is the average forecast price.

Eventually, Figure 133 shows the fan-chart of probability distribution of future values belonging to a certain price range.

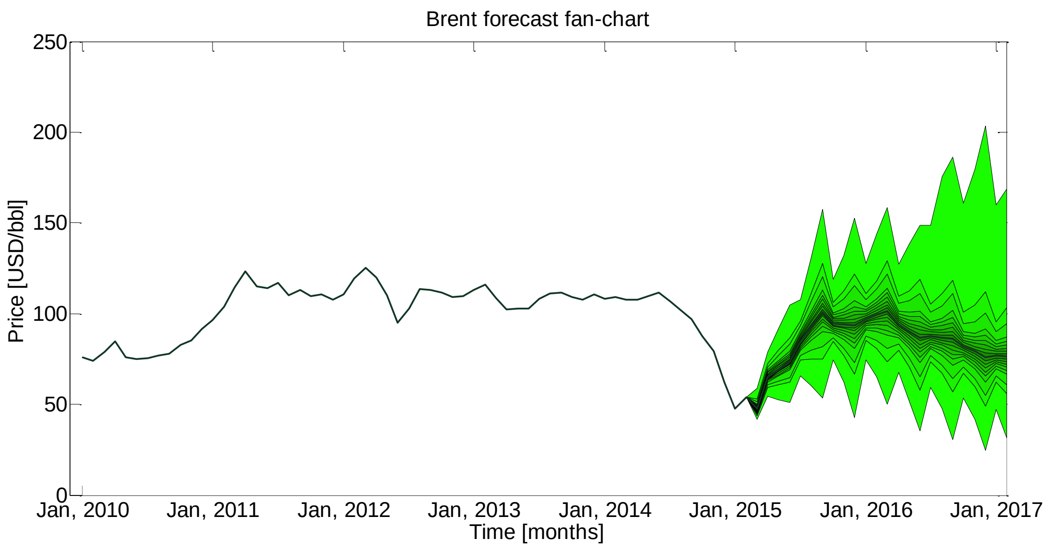


Figure 133 - Fan-chart of Brent monthly prices from Feb, 2010 to Feb, 2017 calculated with the hybrid model. The period from Feb, 2015 to Feb, 2017 is the forecast interval (3000 different simulations) with probability from 0.1% (darker green) to 99.9% (lighter green).

Also in this case the probabilistic thresholds are 0.1%, 5%, 10%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 90%, 95%, and 99.9%. These thresholds are more swinging than the ones provided by the economic model, because the hybrid model features the shock analysis of CO prices. In addition, it is worth observing that future prices belong to the 50-130 USD/bbl interval with a 95% confidence interval.

The same considerations drawn for the OPEC-based model can now be applied to the new hybrid model, as it is an enhanced version of the economic model. The hybrid model can be applied for two-year time-horizons, *i.e.* for scheduling and planning of chemical plants under market uncertainties about supply, demand, and historical unbalances.

Conclusions and future developments

Aim of the present work of thesis was the development of a new CO model for PSE/CAPE applications, such as planning, scheduling, and feasibility studies of chemical plants, under market and historical uncertainties.

Starting from a number of econometric and economic models proposed in literature, the thesis described the dynamic models capable to forecast the evolution of CO quotations over different time horizons, and mainly focused on short- and medium-time intervals (*i.e.* a few years). During these periods, several historical events can influence CO quotations. For instance, shale oil spread in the USA, Greece crisis, Iran withdrawal of embargo, and China crisis of stock exchanges are just a few examples that were presented and discussed.

It is worth observing that past studies mainly focused on one-step-ahead models, which allow predicting the variable of interest for just the following time-step. On the contrary, PSE/CAPE applications are more complex and usually based on short-, medium, and long-term fully-predictive time horizons. The call for creating several distinct scenarios, which correspond to the different historical or technological situations, comes from the necessity to identify possible distributions of economic conditions to answer the typical question of PSE/CAPE applications about the feasibility of products and processes. The proposed models (*i.e*. economic and hybrid models) are used to simulate a number of stochastic scenarios. As economic scenarios are based on a stochastic contribution, it is important to understand possible causes of random variations, which the OPEC-based economic model can assess. This model can produce a global trend for a middle-term horizon (*i.e.* two years).

As discussed in Chapter 5, OPEC decisions about quota and capacity utilization have a significant and immediate impact on oil price, and act as a control on prices, even if the cartel does not have the same power that had in the past (*i.e.* in the 70's of last century), when it could set almost independently the CO prices. The new economic model involves a rather simple structure and a reduced number of parameters. The power of the model consists in including the CO price contributions from both the producers and consumers in the assessment of the price dynamics. Equally, the economic background is changed and is more complex than the one of 5-10 years ago, as OPEC and OECD do not include the so-called BRIC countries, which are considered to be at a similar stage of newly advanced economic development, and emerging markets such as Argentina and South Africa. Furthermore, the increasing concern about oil sands in both the USA and Canada has given a new power to the American economy that could become independent of OPEC decisions about quotas and production. The introduction of a new economic variable that takes into account the recent global oversupply due to the increasing CO production by the USA allowed improving the predictive capability of the OPEC-based model respect to the other literature ones. Another advantage of the proposed economic model is its capability to account for the physical factors that influence CO quotations on both supply-and-demand sides. The process designer can manipulate the input variables in order to create bullish-, bearish-, and conservative-trend scenarios, and take into account economic developments, crisis, supply disruption, or the growth of other sources of oil. The combination of the OPEC-based model with the econometric model proposed by Barzaghi and Conte (2015) allowed creating a hybrid model that has not been reported in the literature yet. This model can simulate the trend proposed by supply-and-demand law, but in combination with the stochastic fluctuations of CO quotations.

Both the economic and hybrid models can be implemented in commercial plant simulators (*e.g.*, UniSim®) or other programs in view of the optimal management of storage capacity of a production site as a function of the global supply and real market demand.

At the same time, it is necessary to highlight the disadvantage of this procedure. The economic model needs to be updated often because of all the unpredictable events that may influence market quotations. Today, more than ever, CO market is seeing the emergence of new countries, the advanced economic development of other ones, and the revival of interest and investment in non-conventional petroleum reservoirs. In addition, the historical price analysis was conducted critically, but without a deeper knowledge of the theory of economic dynamics and other disciplines, such as financial mathematics.

In future, it could be interesting to study if the stochastic fluctuations that are simulated by the hybrid model are relevant for scheduling and planning problems. Furthermore, the costs analysis could be extended by considering the influence of time granularity on the forecast horizon and the price/cost volatility respect to equipment transients. This approach can be used for the design of chemical plants and extended to revamping and retrofitting of existing processes/plants.

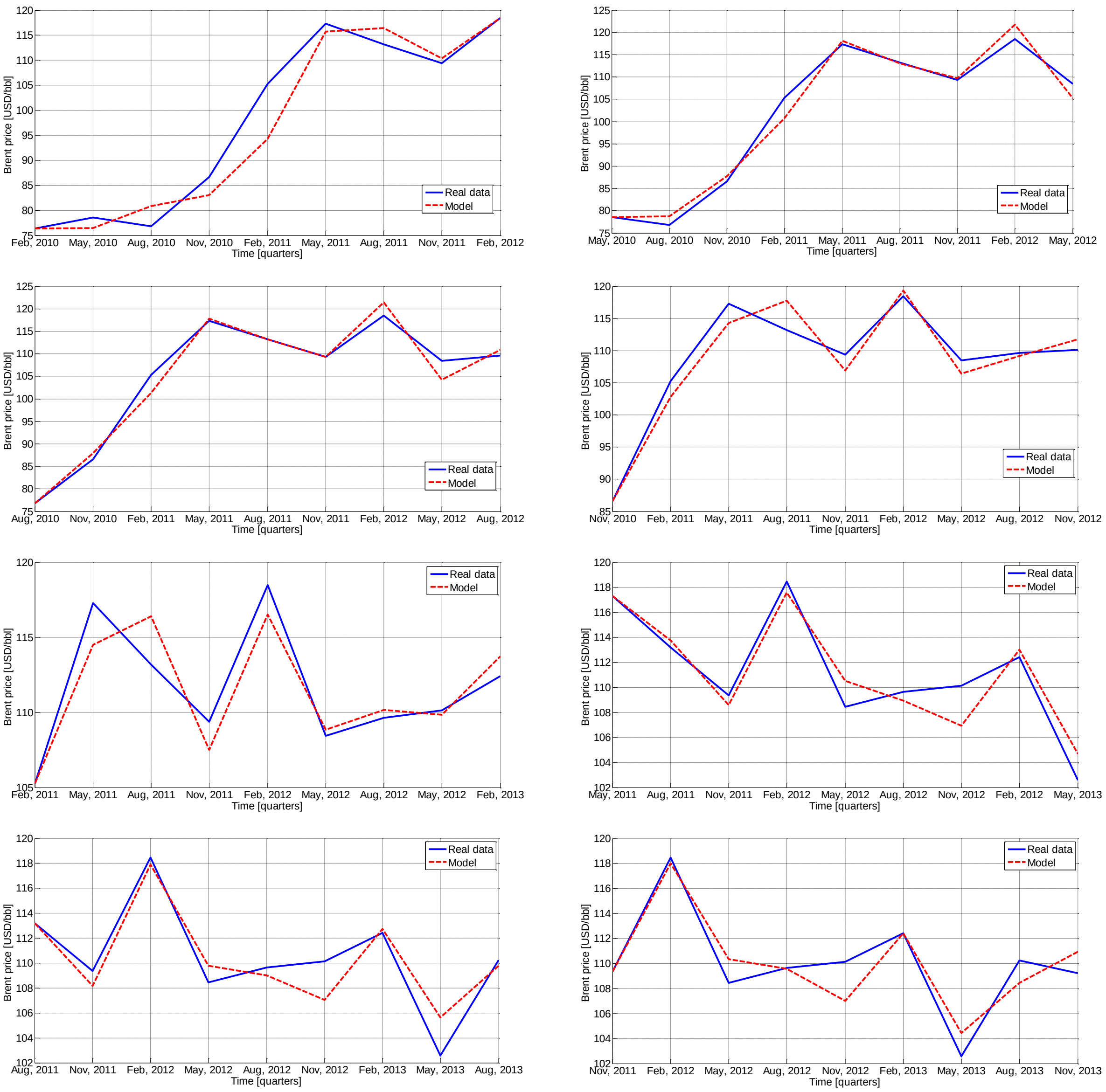
Another important point that could be applied in future to the economic analysis is the Principal Component Analysis (PCA). PCA is a multivariate statistical technique that transforms the original set of correlated variables into a reduced set of uncorrelated variables. This idea can be applied to determine other factors or combination of factors that play a major role on variations of CO prices.

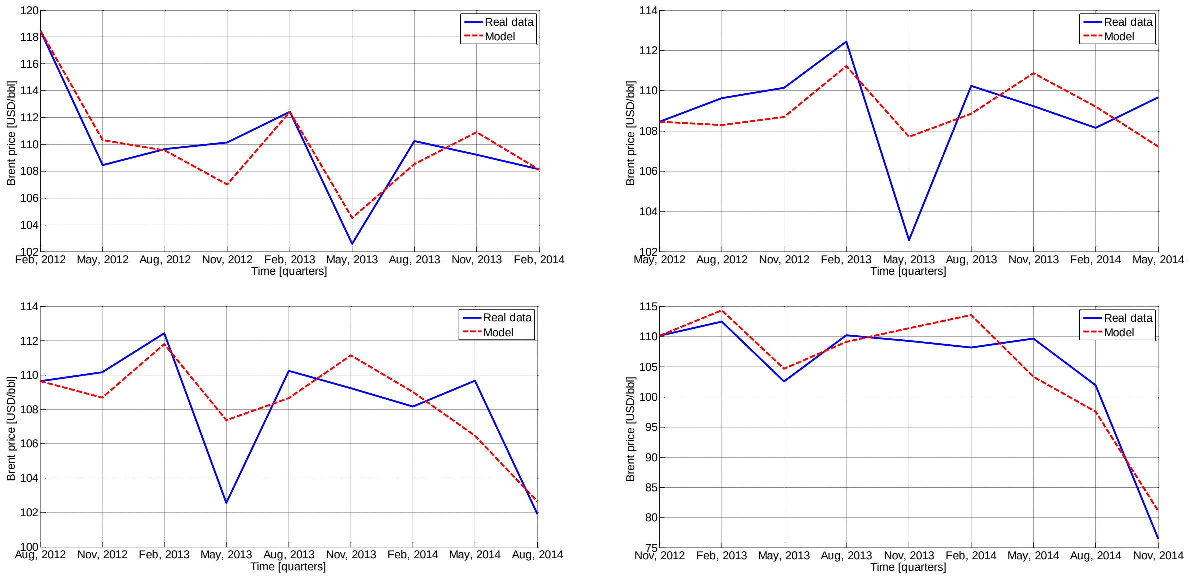
Hopefully, the future research may use these results as a basis for further assessments of price modeling and new optimal management of chemical plants.

Appendix A

A.1 Brent price

This Section reports the graphs of the eight-quarter long one-step-ahead simulations between Feb, 2010 and Nov, 2014 for the Brent prices provided by Equation (40).

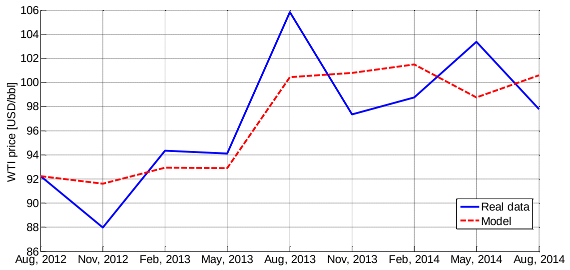


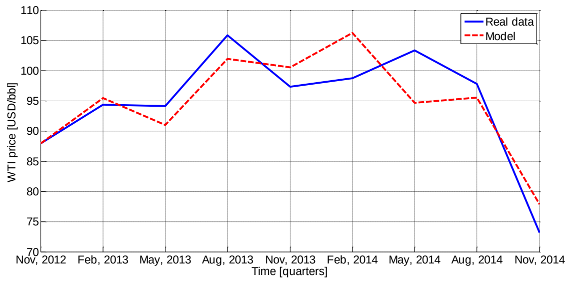


Real data
Model
Brent price [USD/bbl]
Time [quarters]
Feb, 2012
May, 2012
Aug, 2012
Nov, 2012
Feb, 2013
May, 2013
Aug, 2013
Nov, 2013
Feb, 2014
May, 2014
Aug, 2014
Nov, 2014

A.2 WTI price

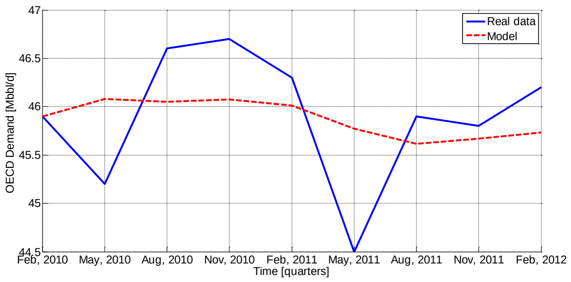
This Section reports the graphs of the eight-quarter long one-step-ahead simulations between Feb, 2010 and Nov, 2014 for the WTI prices provided by Equation (40).

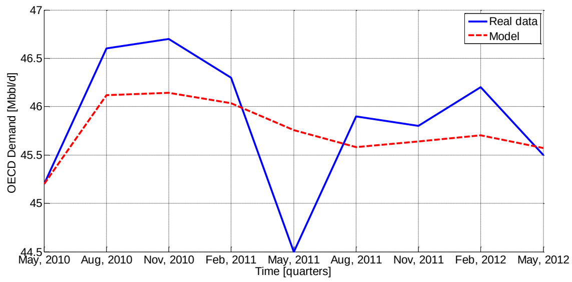

WTI price [USD/bbl]
Real data
Model
Aug, 2012
Nov, 2012
Feb, 2013
May, 2013
Aug, 2013
Nov, 2013
Feb, 2014
May, 2014
Aug, 2014

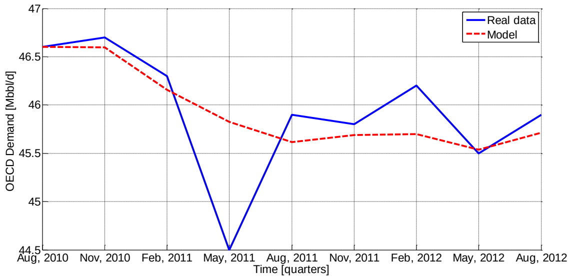

WTI price [USD/bbl]
Real data
Model
Nov, 2012
Feb, 2013
May, 2013
Aug, 2013
Nov, 2013
Feb, 2014
May, 2014
Aug, 2014
Nov, 2014
Time [quarters]

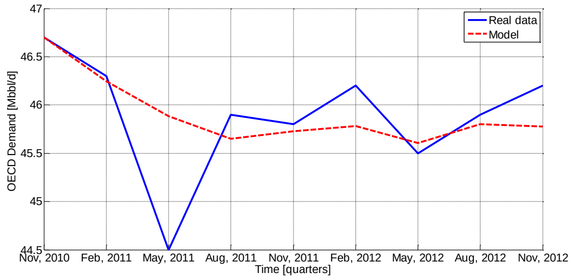
A.3 OECD demand

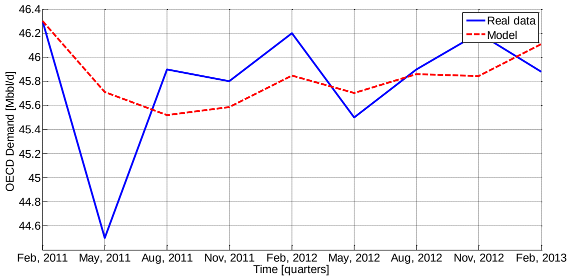
This Section reports the graphs of the eight-quarter long one-step-ahead simulations between Feb, 2010 and May, 2014 for the OECD demand provided by Equation (45).

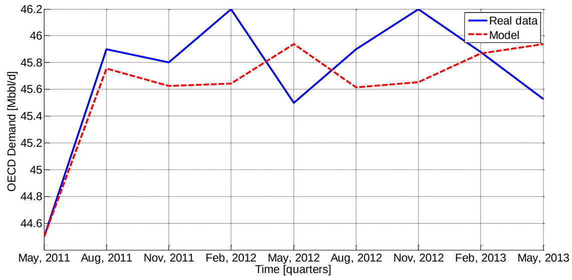


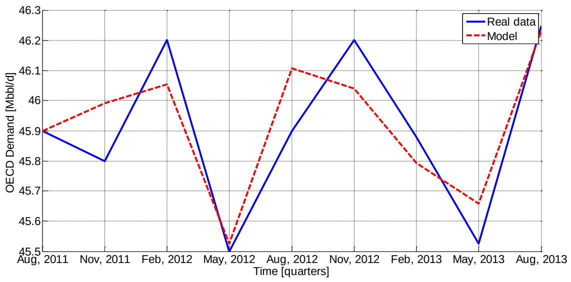


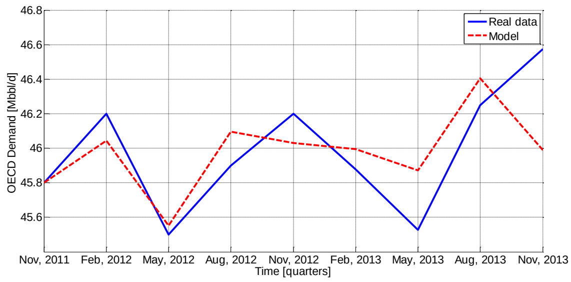


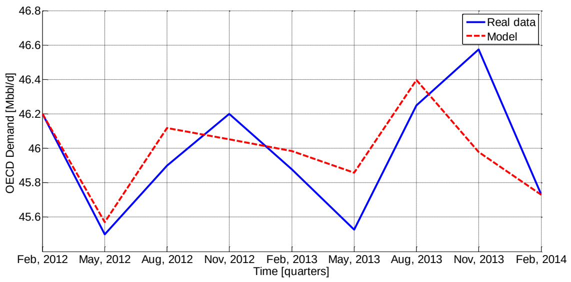


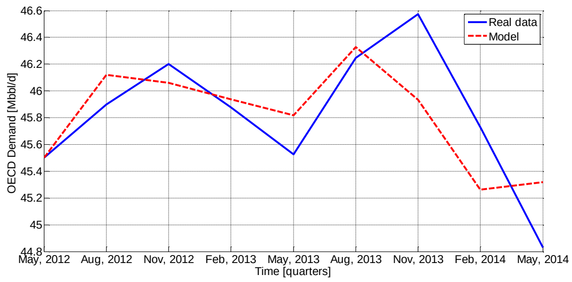










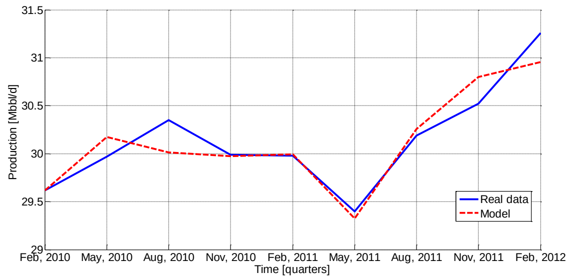


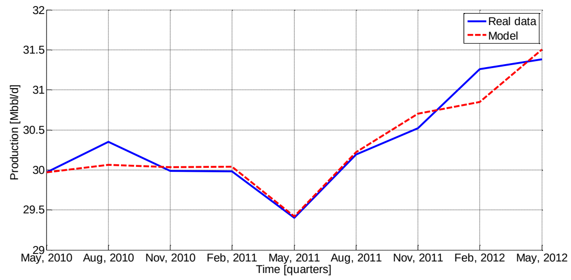
A.4 OECD inventories

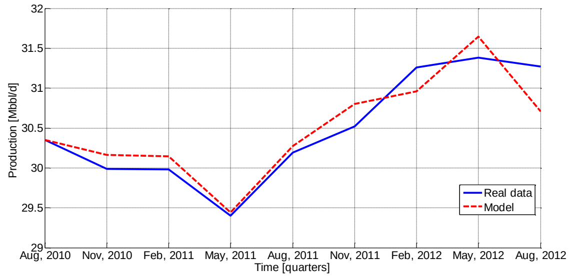
This Section reports the graphs of the eight-quarter long one-step-ahead simulations between Feb, 2010 and May, 2014 for the OECD inventories provided by Equation (46).

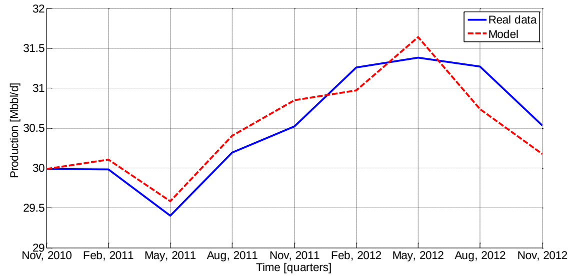
A.5 OPEC production

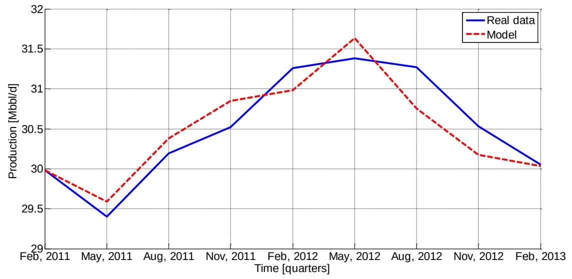
This Section reports the graphs of the eight-quarter long one-step-ahead simulations between Feb, 2010 and May, 2014 for the OPEC production provided by Equation (50).

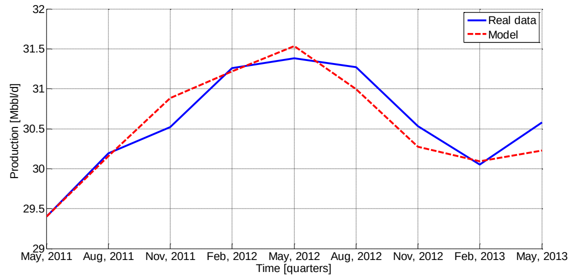


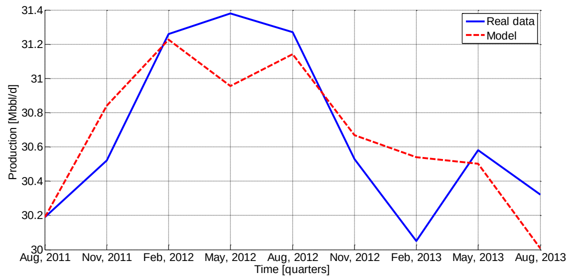


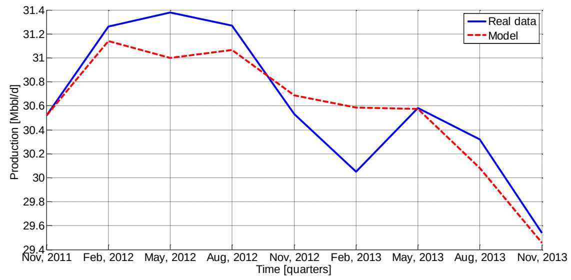


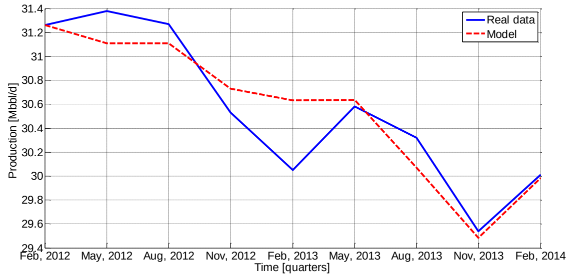


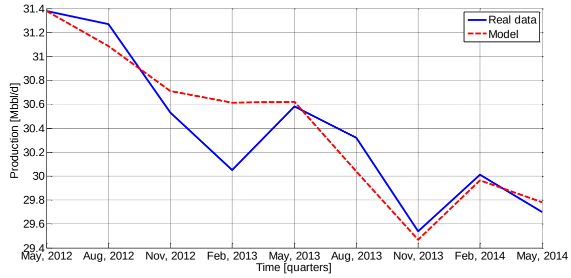






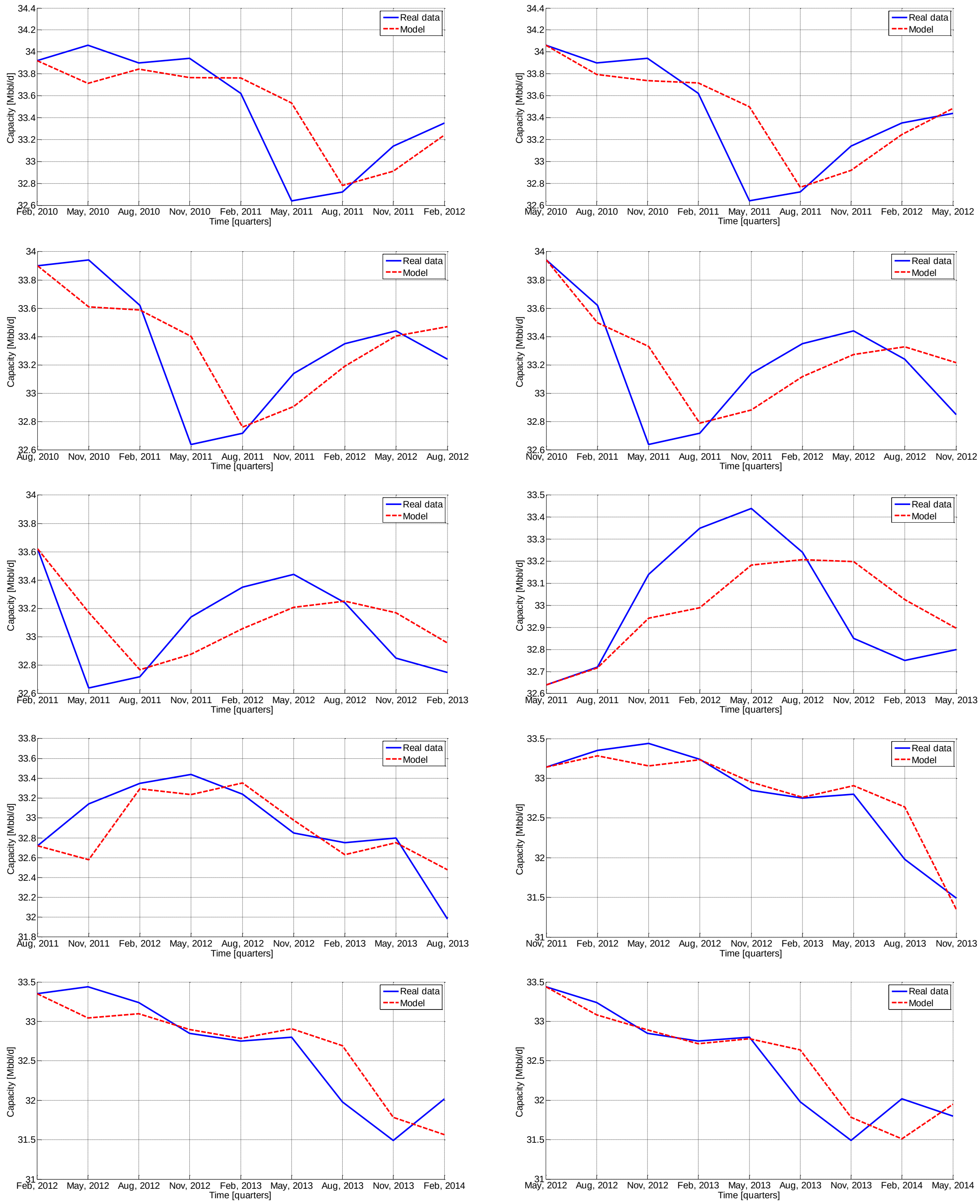






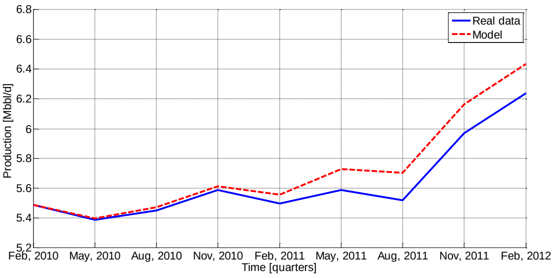
A.6 OPEC production capacity

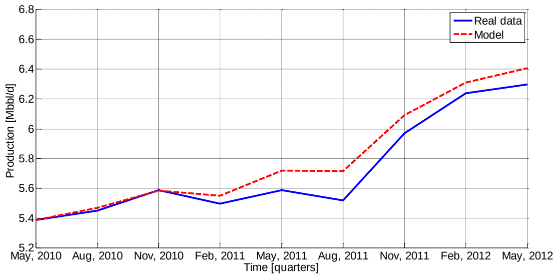
This Section reports the graphs of the eight-quarter long one-step-ahead simulations between Feb, 2010 and May, 2014 for the OPEC production provided by Equation (51).

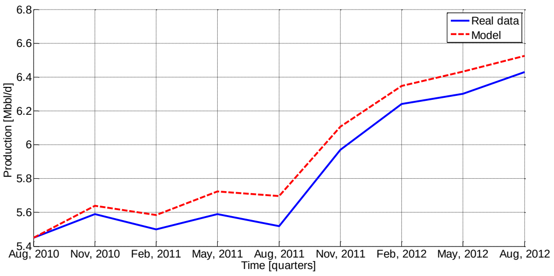


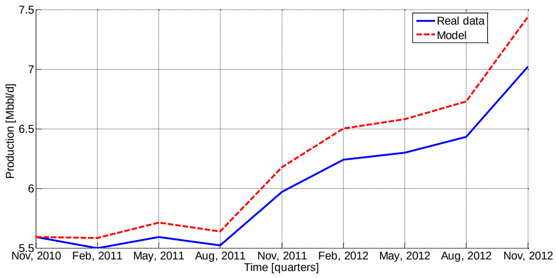
A.7 USA production

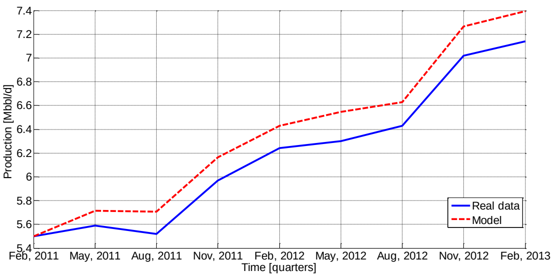
This Section reports the graphs of the eight-quarter long one-step-ahead simulations between Feb, 2010 and May, 2014 for the USA production provided by Equation (54).

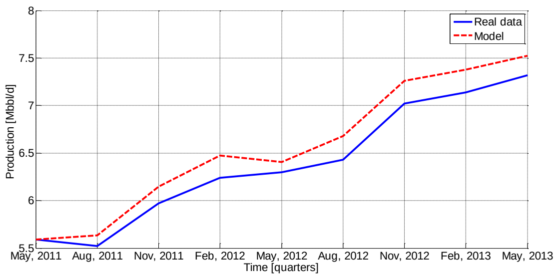


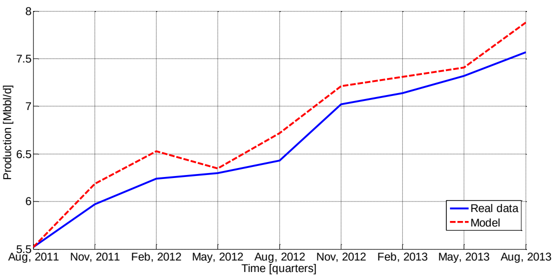


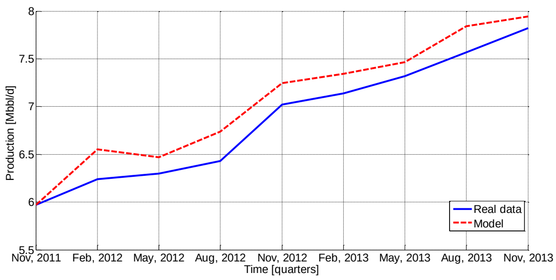


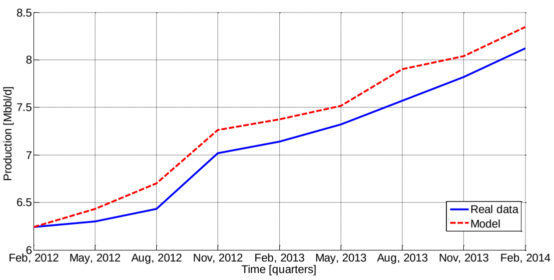


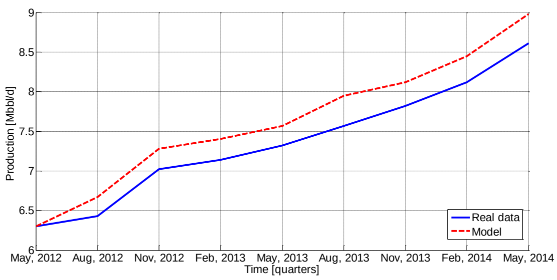








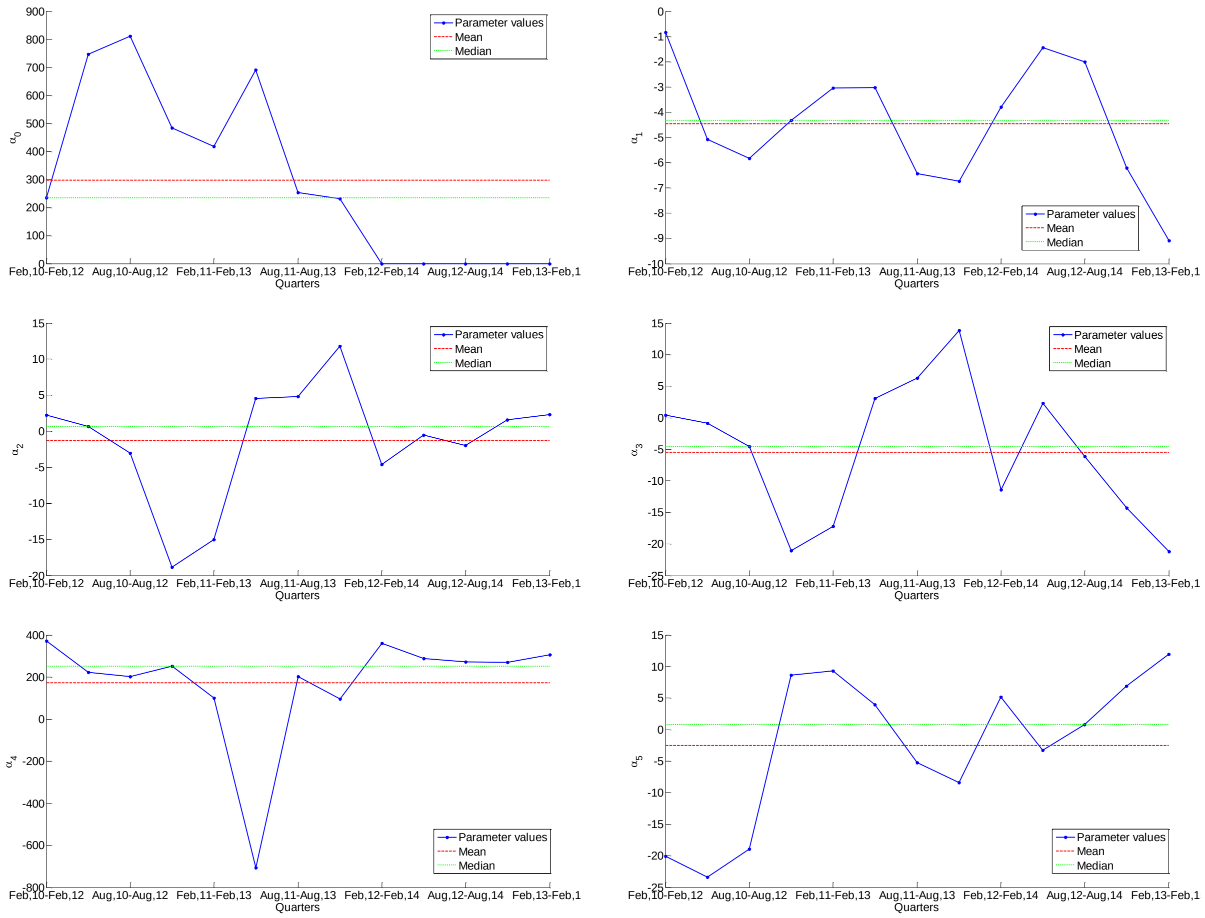




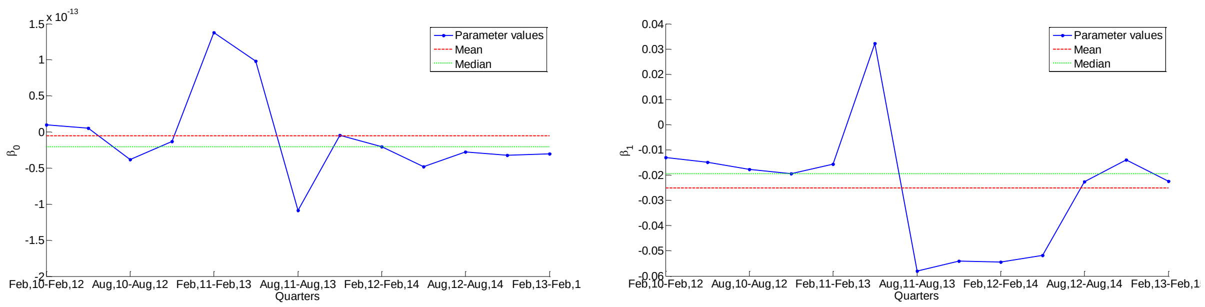
A.8 Model parameters

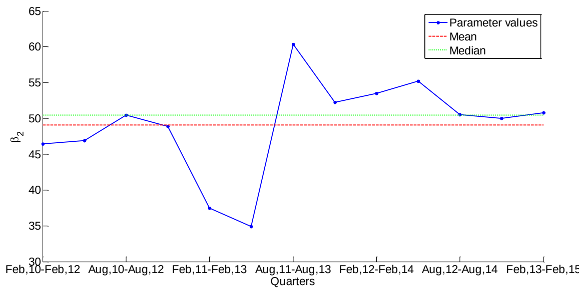
This Section reports the graphs of adaptive parameter trends in Equations (40-45-46-50-51-54).

$$WTI_t = \alpha_0 + \alpha_1 \cdot Days_t + \alpha_2 \cdot Quotas_t + \alpha_3 \cdot Cheat_t + \alpha_4 \cdot Caputil_t + \alpha_5 \cdot Delta_t$$

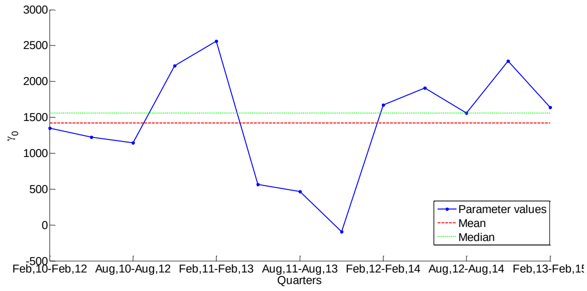


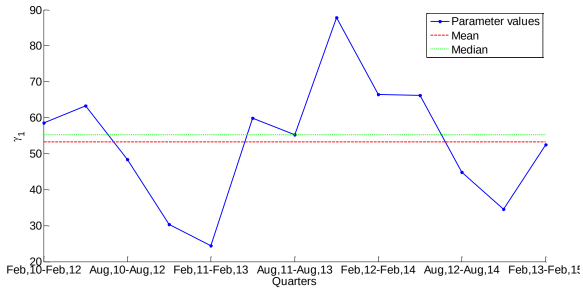
$$Demand_{t+1}^{OECD} = \beta_0 GDP_{t+1} + \beta_1 Price_t + \beta_2$$

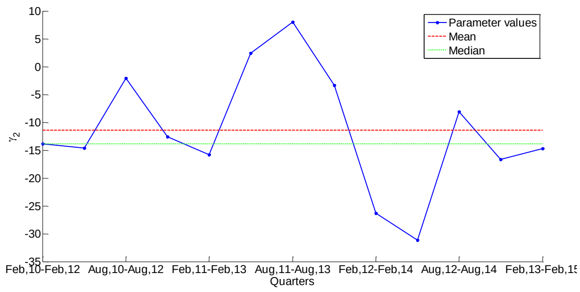




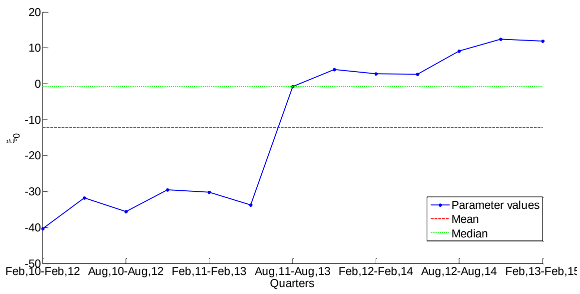
$$Inventory_{t+1}^{OECD} = \gamma_0 + \gamma_1 Capacity_t^{OPEC} + \gamma_2 Demand_{t+1}^{OECD}$$

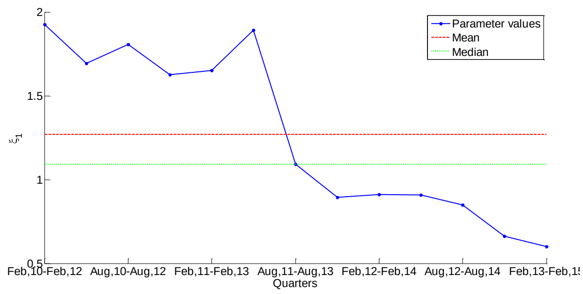


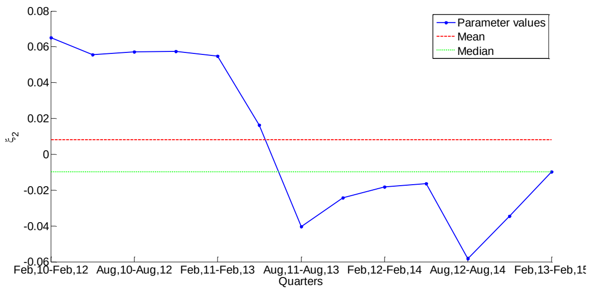




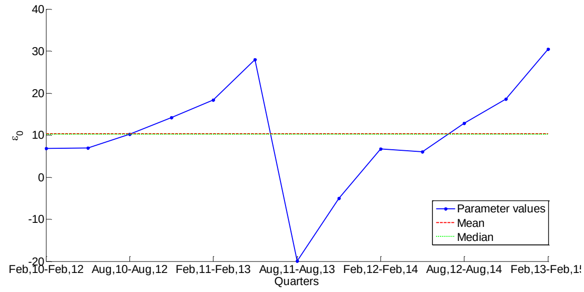
$$Production_{t+1}^{OPEC} = \xi_0 + \xi_1 Capacity_{t+1}^{OPEC} + \xi_2 Price_t$$

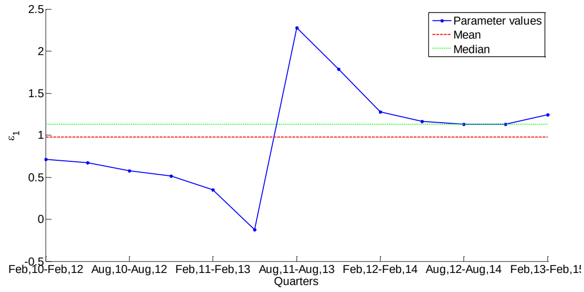


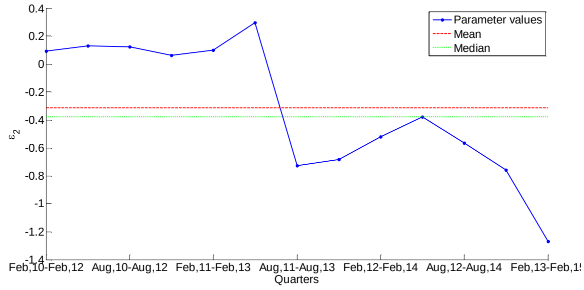




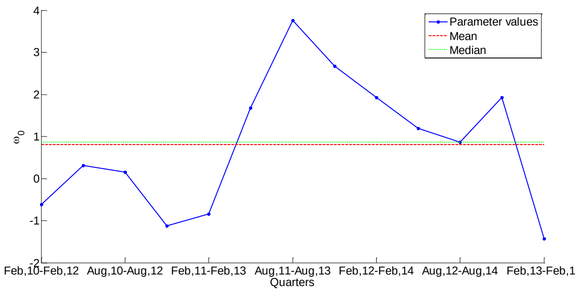
$$Capacity_{t+1}^{OPEC} = \varepsilon_0 + \varepsilon_1 Capacity_t^{OPEC} + \varepsilon_2 Production_t^{OPEC}$$

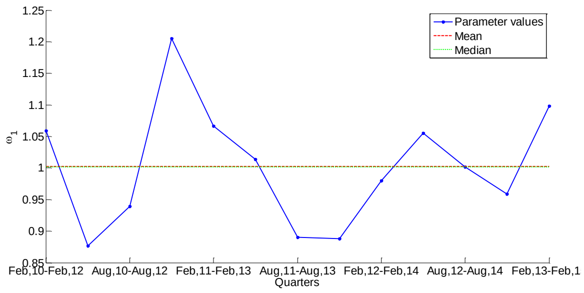


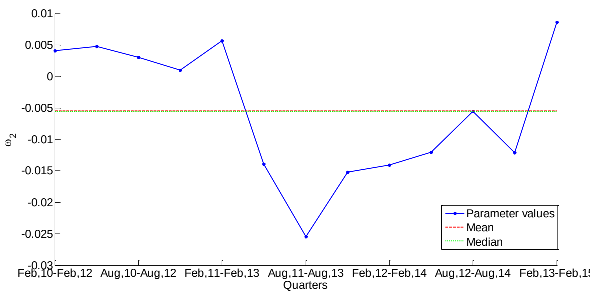




$$Production_{t+1}^{USA} = \omega_0 + \omega_1 Production_t^{USA} + \omega_2 Price_t$$







Appendix B

B.1 Fisher information matrix

This Section reports the code to be inserted by the user in the *Matlab©* script that generates and analyzes the Fisher Information Matrix and the Covariance Matrix. The comments to the code start with a % symbol.

```
%% Declaration of the symbolic variables, i.e. model parameters
syms alfa0 alfa1 alfa2 alfa3 alfa4 alfa5

%% Reading of the observed data of the input variables
Capacity=xlsread('dataRevisione','Sheet1','Z12:AA12');
Demand=xlsread('dataRevisione',' Sheet1','Z8:AA8');
Quota=xlsread('dataRevisione',' Sheet1','Z11:AA11');
USA=xlsread('dataRevisione',' Sheet1','Z17:AA17');
Inventory=xlsread('dataRevisione',' Sheet1','Z16:AA16');
Production=xlsread('dataRevisione',' Sheet1','Z10:AA10');

%% Declaration of the model
for i=1:length(Inventory)
    P(i)=alfa0+alfa1*(Inventory(i)/Demand(i))+alfa2*Quota(i)+alfa3*
        (Production(i)-
Quota(i))+alfa4*(Production(i)/Capacity(i))+alfa5*
        (Production(i)-USA(i));
end

%% First-order derivative calculation
% The diff function allows calculating the approximate derivative.
dPdalfa0=diff(P,alfa0,1)
dPdalfa1=diff(P,alfa1,1)
dPdalfa2=diff(P,alfa2,1)
dPdalfa3=diff(P,alfa3,1)
dPdalfa4=diff(P,alfa4,1)
dPdalfa5=diff(P,alfa5,1)
dPdalfa=[dPdalfa0;   dPdalfa1;   dPdalfa2;   dPdalfa3;   dPdalfa4;
dPdalfa5];

% The subs function allows substituting the symbolic parameters with
% their fiducial values.
ds=subs(dPdalfa,{alfa0,alfa1,alfa2,alfa3,alfa4,alfa5},[629.7461,-
7.2994,
-46.546,-52.1525,768.8437,25.7076])

%% FIM construction and analysis
```

```
H=size(dPdalfa,1) % parameter number

for j=1:H
    for k=1:H
        A1(k,j)=ds(k,1)*ds(j,1);
        A2(k,j)=ds(k,2)*ds(j,2);
        F=A1+A2
    end
end

R=rank(F) % matrix rank
c=cond(F) % matrix condition number
D=det(F) % matrix determinant


%% Covariance matrix construction
[L,U,P] = lu(F) % LU factorization
C=inv(U)*inv(L)*P % covariance matrix
```

Appendix C

C.1 DAISY input file

This Section reports in lower-case the source code to be inserted by the user in the input file. The fixed structure is upper-case. The language used is REDUCE and the comments to the code start with the % symbol.

```
WRITE "OPEC MODEL COMPLETE"$

% B_ is a reserved name used to denote the vector of the input, output, and state variables. Note that the components should be ordered as follows: first the input and then the output followed by the states (in the OPEC-based model there are no states). Names of the variables can be freely chosen by the user.

B_:={PRICE,DEMAND,INVENTORY,PRODUCTION,CAPACITY,USA,QUOTAS}$

% The following instruction defines the components of vector B_ as discrete time-depending variables. Constant inputs must not be listed in vector B_, but directly included in the model equations.

FOR EACH EL_ IN B_ DO DEPEND EL_,T$

% B1_ is a reserved name used to indicate the vector of unknown parameters.

B1_:={ALFA0,ALFA1,ALFA2,ALFA3,ALFA4,ALFA5}$

% If there are constraints relating the parameters or some parameters are known, instructions such as “LET” can be used and the user must delete from vector B1_ the known or constrained parameters.

% NU_, NY_ and NX_ are reserved to indicate the number of inputs, outputs and states included in vector B_. Thus the number of the B_ components should be equal to NU_ + NY_ + NX. If the model has no input just write NU_:=0.

NU_:=0$

NY_:=1$

NX_:=6$

% C_ is a reserved variable name used to indicate the system of POLYNOMIAL or RATIONAL first order differential equations describing the model.
```

```
C_:={DEMAND=46.7,
INVENTORY=2708,
PRODUCTION=30.11,
CAPACITY=32.22,
USA=8.82,
QUOTAS=30,
PRICE=ALFA0+ALFA1*INVENTORY/DEMAND+ALFA2*QUOTAS+ALFA3*(PRODUCTION+
-QUOTAS)+ALFA4*PRODUCTION/CAPACITY+ALFA5*(PRODUCTION-USA)}$
% Note that algebraic (i.e. non differential) equations are allowed.
In this case the additional algebraic variables have also to be
included in the B_ vector and the number NX_ has to be correctly
incremented.

% Choose an integer value (seed_) bigger than the number of unknown
parameters. The subroutine “random” will choose, in a random way in
the interval [1, seed_], the numerical values corresponding to each
component (model unknown parameter) of vector B1_.
SEED_:=10000$

% Declare the procedure that calculates the characteristic set.
DAISY()$

% Complete the input file.
END$
```

C.2 DAISY output file

The results provided by DAISY are reported below. For the sake of brevity, the whole file is not reported. Just the relevant results for identifiability test are reported.

```
OPEC MODEL COMPLETE

NUMBER OF EQUATIONS$
n_:=7$

VARIABLES VECTOR$
b_:={price, demand, inventory, production, capacity, usa, quotas}$

UNKNOWN PARAMETER(S) VECTOR$
b1_:={alfa0, alfa1, alfa2, alfa3, alfa4, alfa5}$
```

RANKING AMONG THE VARIABLES$

```
bb_:={price, df(price,t), df(price,t,2), df(price,t,3),
df(price,t,4), df(price,t,5), df(price,t,6), demand, inventory,
production, capacity, usa, quotas, df(demand,t), df(inventory,t),
df(production,t), df(capacity,t), df(usa,t), df(quotas,t)}$
```

NUMBER OF INPUT(S)$

```
nx_:=6$
```

NUMBER OF OUTPUT(S)$

```
ny_:=1$
```

MODEL EQUATION(S)$

```
c_:={demand=467/10,inventory=2708,production=3011/100,capacity=1611/
50,
usa=441/50,quotas=30,price=(((production-usa)*alfa5*capacity+alfa4*
production+(production-
quotas)*alfa3*capacity+alfa2*capacity*quotas)*
demand+alfa1*capacity*inventory+alfa0*capacity*demand)/(capacity*dem
and)}$
```

RANDOMLY CHOSEN NUMERICAL PARAMETER(S) VECTOR$

```
b2_:={alfa0=8708,alfa1=1191,alfa2=3428,alfa3=3520,alfa4=4775,alfa5=7
167}$
```

MODEL PARAMETER SOLUTION(S)$

```
G_:=GROESOLVE(FLIST_,B1_)$
```

```
g_:={{alfa0=8708,alfa1=1191,alfa2=3428,alfa3=3520,alfa4=4775,alfa5=7
167}}$
```

MODEL GLOBALLY IDENTIFIABLE$

References

Abledu, G. K., Agyemang, B., Reubin, S., 2013. *Forecasting demand of petroleum products using time series models*. Journal of Economics and Sustainable Development, 17(4), 129-142.

Agnoli, S., 2014. *Petrolio sotto quota 90: la reazione di Eni & Co*. Corriere della Sera. Available online at http://www.corriere.it/economia/finanza_e_risparmio/notizie/petrolio-sotto-quota-90-quanti-guai-big-reazione-eni-co-ff7cc6e8-6411-11e4-8b92-e761213fe6b8.shtml.

Al-Bisharah, M., Al-Fattah, S., Nashawi, I. S., Malallah, A., 2009. *Forecasting OPEC crude oil supply*. Society of Petroleum Engineers, SPE Middle East Oil & Gas Show and Conference, Bahrain, Working paper 120350.

Al-Qahtani, A., Balistreri, E., Dahl, C., 2008. *Literature review on oil market modeling and OPEC's behavior*. Available online at https://www1.ethz.ch/cepe/education/EnergyPolicy/LitReviewOPEC.pdf.

Alquist, R., Kilian, L., 2010. *What do we learn from the price of crude oil futures?* Journal of Applied Econometrics, 25(4), 539-573.

Alquist, R., Kilian, L., Vigfusson, R. J., 2011. *Forecasting the price of oil*. Board of Governors of the Federal Reserve System, International Financial Discussion paper 1022, Washington, DC.

Aprea, M., Catovic, D., Reale, F., Talhi, A., 2005. *I modelli teorici per la determinazione del prezzo del petrolio*. Available online at http://www.eni.com/attachments/lavora-con-noi/pdf/Modelli-teorici-di-determinazione-del-prezzo-del-petrolio.pdf.

Bardi, U., 2009. *Peak oil: the four stages of a new idea*. Energy, 34, 323-326.

Bardi, U., Lavacchi, A., 2009. *A simple interpretation of Hubbert's model of resource exploitation*. Energies, 2, 646-661.

Bardi, U., Yaxley, L., 2005. *How general is the Hubbert curve? The case of fisheries*. 4th International ASPO Conference, Lisbon, Portugal.

Barzaghi, R., Conte, A., 2015. *Effetto della variabilità di mercato sulla progettazione di impianti chimici*. Politecnico di Milano, Tesi di Laurea.

Bassett, E., 2014. *Few in US oil, gas realm very concerned about possible oil price collapse: survey*. Platts.com, McGraw Hill Financial, November 18, Houston, TX.

Bellomo, S., Negri, A., 2014. *Petrolio, l'OPEC: "Manteniamo ferma la produzione". Prezzo del greggio ancora a picco*. Sole 24 Ore, November 28, Vienna.

Bellu, G., Saccomani, M. P., Audoly, S., D'Angiò, L., 2007. *DAISY: a new software tool to test global identifiability of biological and physiological systems*. Computer Methods and Programs in Biomedicine, 88, 52-61.

Bezruchko, B. P., Smirnov, D. A., 2010. Extracting knowledge from time series. Springer Series in Synergetics, Berlin.

Bollerslev, T., 1986. *Generalized autoregressive conditional heteroskedasticity*. Journal of Econometrics, 31(3), 307-327.

BP, 2014. *Statistical Review of World Energy June 2014*. Available online at http://www.bp.com/statisticalreview.

BP, 2015. *Statistical Review of World Energy June 2015*. Available online at http://www.bp.com/statisticalreview.

Campbell, C J., 1997. *The coming oil crisis*. Multi-Science Publishing Company and Petrocunsultants, Essex.

Chatfield, C., 2000. *Time-series forecasting*. Chapman & Hall/CRC, London.

Chatfield, C., 2003. *The analysis of time series: an introduction, Sixth edition*. Chapman and Hall/CRC, London.

Chen, S. S., 2014. *Forecasting crude oil price movements with oil-sensitive stocks*. Economic Inquiry, 52(2), 830-844.

Chevallier, J., 2014. *The 2008 oil price swing or the quest for a 'smoking gun'*. Crude Oils, Chapter 3, 39-69.

Chis, O. T., Banga, J. R., Balsa-Canto, E., 2011. *Structural identifiability of systems biology models: a critical* comparison of methods. PLoS ONE, http://dx.doi.org/10.1371/journal.pone.0027755.

Ciffarelli, G., Paladino, G., 2010. *Oil price dynamics and speculation. A multivariate financial approach*. Energy economics, 32, 363-372.

Clifford, K., 2014. *US oil prices fall below $80 a barrel*. The New York Times, November 03, Houston TX.

Commandeur, J. J. F., Koopman, S. J., 2007. *An introduction to state space time series analysis, First edition*. Oxford University Press, Oxford.

Cooper, J.C.B., 2003. *Price elasticity of demand for crude oil: estimates for 23 countries*. Organization of the Petroleum Exporting Countries, Glasgow.

Dahmani, A., Al-Osaimy, M. H., 2001. *OPEC oil production and market fundamentals: a causality relationship*. Organization of the Petroleum Exporting Countries, Vienna.

Davis, N., Fleming, N., 2014. *The oil price crash and what it means for the petrochemical industry*. ICIS, December 08.

DeCarlo, L. T., 1997. *On the meaning and use of kurtosis*. Psychological Methods, 2, 292-307.

Dees, S., Karadeloglou, P., Kaufmann, R., Sanchez, M., 2007. *Modelling the world oil market: Assessment of a quarterly econometric model*. Energy Policy, 35(1), 178-191.

Deffeyes, K. S., 2001. *Hubbert's peak: the impending world oil shortage*. Princeton University Press, NJ.

Dietrich, F., Raatz, A., Hesselbach, J., 2010. *A-priori Fisher information of nonlinear state space models for experiment design*. IEEE International Conference on Robotics and Automation, Anchorage, AK.

Ding, Z., Granger C. W. J., Engle R. F., 1993. *A long memory property of stock market returns and a new model*. Journal of Empirical Finance 1, 83-106.

Douglas, J. M., 1988. *Conceptual Design of chemical processes*. McGraw Hill Book Company, New York.

Dowling, M., Cummins, M., Lucey, B. M., 2014. *Psychological barriers in oil futures markets*. Energy Economics. Article in press.

Dvir, E., Rogoff, K., 2009. *Three epochs of oil*. National Bureau of Economic Reaserch. Working paper 14927.

Dvir, E., Rogoff, K., 2014. *Demand effects and speculation in oil markets: theory and evidence*. Journal of International Money and Finance, 42, 113-128.

Engle, R., 1982. *Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation*. Econometrica, 50(4), 391-407.

Eisenberg, M. C., Hayashi, M. A. L., 2014. *Determining identifiable parameters combinations using subset profiling*. Mathematical Biosciences, 256, 116-126.

Elliott, S., 2015. *European oil majors cut CAPEX, eye output growth*. Platts.com, McGraw Hill Financial, February 24, London.

Fadhil, M., 2015. *Iran nuclear deal to alter Mideast polymer trade flows*. ICIS, July 15, Abu Dhabi.

Fadhil, M., Dennis, J., 2015. *Iran nuclear deal reached: crude oversupply concerns heighten*. ICIS, July 14, Abu Dhabi.

Fini, A., Oliosi, M., 2010. *Dynamic Conceptual Design: studio di fattibilità di un impianto chimico in funzione della variabilità di mercato*. Politecnico di Milano, Tesi di Laurea.

Fisher, A.C., 1981. *Resources and environmental economics*. Cambridge University Press, Cambridge.

Fong, W. M., See, K. H., 2002. *A Markov switching model of the conditional volatility of crude oil futures prices*. Energy Economics, 24, 71-95.

Fortunati, S., Gini, F., Greco, M., Farina, A., Graziano, A., Giompapa, S., 2012. *An identifiability criterion in the presence of random nuisance parameters*. 20th European Signal Processing Conference, Bucharest, Romania.

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

Gilbert, R. O., 1987. *Statistical methods for environmental pollution monitoring*. Van Nostrand Reinhold Company, New York.

Glosten, L., Jagannathan, R., Rukle, D., 1993. *On the relation between the expected value and the volatility of the normal excess return on stocks*. The Journal of Finance, 48(5), 1779-1801.

Goldman Sachs, 2014. *Unlocking the economic potential of North America's energy resources*. Global Markets Institute, Goldman Sachs Global Investment Research, New York.

Grana, R., Manenti, F., Manca, D., 2009. *Towards Dynamic Conceptual Design*. Computer Aided Chemical Engineering, 26, 665-670.

Griffin, J. M., 1985. *OPEC behavior: a test alternative hypotheses*. American Economic Review, 10, 209-230.

Hamilton, J. D., 1983. *Oil and the macroeconomy since World War II*. The Journal of Political Economy, 228-248.

Hamilton, J. D., 2005. *Oil and the macroeconomy*. The New Palgrave Dictionary of Economics. Palgrave Macmillan, New York.

Hines, K. E., Middendorf, T. R., Aldrich, R. W., 2014. *Determination of parameter identifiability in nonlinear biophysical models: a Bayesian approach*. Journal of General Physiology, 401-416.

Horn, M., 2004. *OPEC's optimal crude oil price*. Energy Policy, 32, 269-280.

Hoyos, C., 2007. *The new Seven Sisters: oil and gas giants dwarf western rivals*. Financial Times, March 12.

Hubbert, M. K., 1956. *Nuclear energy and fossil fuels*. American Petroleum Institute Drilling and Production Practice, 7-25, New York.

Hubbert, M. K., 1962. *Energy Resources: a report to the committee on natural resources*. National Academy of Sciences, Washington, DC.

ICIS, Trusted market intelligence for the global chemical, energy, and fertilizer, 2015. *www.icis.com*.

International Energy Agency, 2015. *www.iea.org*.

Kang, S. H., Yoon, S. M., 2013. *Modeling and forecasting the volatility of petroleum futures prices*. Energy Economics, 36, 354-362.

Kao, C. W., 2012. *Price discount, inventories and the distortion of WTI benchmark*. Energy Economics, 34, 117-124.

Kaufmann, R. K., 1995. *A model of the world oil market for Project LINK: integrating economics, geology, and politics*. Economic Modeling, 12(2), 165-178.

Kaufmann, R. K., Dees, S., Karadeloglou, P., Sanchez, M., 2004. *Does OPEC matter? An econometric analysis of oil prices*. Energy Journal, 25(4), 67-90.

Kaufmann, R. K., Ullman, B., 2009. *Oil prices, speculation, and fundamentals: interpreting causal relations among spot and futures prices*. Energy Economics, 31, 550-558.

Kaufmann, R.K., Karadeloglou, P., Di Mauro, F., 2008. *Will oil prices decline over the long run?* Occasional Paper Series 98, European Central Bank.

Kilian, L., 2008. *Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil market*. Unpublished manuscript, University of Michigan, Ann Arbor, MI.

Laherrère, J., 2011. *OPEC quotas and crude oil production*. Available online at http://www.theoildrum.com/node/7363.

Li, L., 2014. *Market review: China oil market*. ICIS, C1 Energy, November 30.

Liesman, S., 2015. *Why Is the Economy Always So Weak in the First Quarter? Nobody Really Knows*. CNBC, April 22.

Liu, W. M., Schultz, E., Swieringa, J., 2014. *Price dynamics in global crude oil markets*. The Journal of Futures Markets, 0, 1-15.

Ljung, L, Glad, S. T., 1994. *On global identifiability for arbitrary model parameterizations*. Automatica, 30(2), 265-276.

Manca, D., 2013. *Modelling the commodity fluctuations of OPEX terms*. Computers and Chemical Engineering, 57, 3-9.

Manca, D., Depetri, V., Boisard. C., 2015. *A crude oil economic model for PSE applications*. Computer Aided Chemical Engineering, 37, 491-496, http://dx.doi.org/10.1016/B978-0-444-63578-5.50077-3.

Manca, D., Fini, A., Oliosi, M., 2011. *Dynamic Conceptual Design under market uncertainty and price volatility*. Computer Aided Chemical Engineering, 29, 336-340.

Manca, D., Grana, R., 2010. *Dynamic Conceptual Design of Industrial Processes*. Computer and Chemical Engineering, 34 (5), 656-667.

Mardia, K. V., 1970. *Measures of multivariate skewness and kurtosis with applications*. Biometrika, 57, 519-530.

Masarotto, G. (2005). *Analisi delle serie temporali*. Università di Padova, Italy.

Mazzetto, F., Ortiz-Gutiérrez, R. A., Manca, D., Bezzo, F., 2013. *Strategic design of bioethanol supply chains including commodity market dynamics*. Industrial & Engineering Chemistry Research, 52, 10305-10316.

Medlock, K. B., Jaffe, A. M., 2009. *Who is in the oil futures market and how has it changed?* Baker Institute Study, Rice University, Houston, TX.

Miller, R. G., Sorrell, S. R., 2013. *The future of oil supply*. Philosophical Transactions of Royal Society, 372, rsta.royalsocietypublishing.org.

Moebert, J., 2007. *Crude oil price determinants*. Applied Research in Economics. Working paper 186.

Molchanov, P., 2003. *A statistical analysis of OPEC quota violations*. Duke University, Durham, NC.

Morana, C., 2001. *A semiparametric approach to short-term oil price forecasting*. Energy Economics, 23, 325-338.

Murphy, D. J., Hall, C. A. S., 2011. *Energy return on investment, peak oil, and the end of economic growth*. Ecological Economics Review, Annals of the New York Academy of Sciences, 1219, 52-72.

Nashawi, I. S., Malallah, A., Al-Bisharah, M., 2010. *Forecasting world crude oil using multicyclic Hubbert model*. Energy Fuels, 24, 1788-1800.

Nelson, D., 1991. *Conditional heteroskedasticity in asset returns: a new approach*. Econometrica, 59(2), 347-370.

Oil & Gas Journal, 2015. *US petroleum product exports rise for 13th consecutive year*. March 3, Houston, TX.

Ollivier, F., 1990. *Le problème de l'identifiabilité structurelle globale: étude thèorique, mèthodes effectives et bornes de complexité*. PhD thesis, École Polytéchnique, Paris.

Organization of Petroleum Exporting Countries, 2012. *OPEC Statute*. Vienna, Austria.

Organization of Petroleum Exporting Countries, 2015. *OPEC Monthly Oil Market Report*. April 16, Vienna.

Organization of Petroleum Exporting Countries, 2015. *www.opec.org*.

Rasello, R., Manca, D., 2014. *Stochastic price/cost models for supply chain management of refineries*. Computer aided Chemical Engineering, 33, 433-438.

Saccomani, M. P., Audoly, S., Bellu, G., D'Angiò, L., 2010. *Examples of testing global identifiability of biological and biomedical models with the DAISY software*. Computers in Biology and Medicine, 40, 402-407.

Saccomani, M. P., Audoly, S., D'Angiò, L., 2003. *Parameter identifiability of nonlinear systems: the role of initial conditions*. Automatica, 39, 619-632.

Salisu, A. A., Fasanya, I. O., 2013. *Modeling oil price volatility with structural break*. Energy Policy, 52, 554-562.

Sandrea, R., 2003. *OPEC's next challenge: rethinking their quota system*. Oil and Gas Journal, 101.29, July 28.

Sen, A., 2012. *Oil benchmarks in international trade*. Oxford Energy Forum, 87, 19-22.

Sole 24 Ore, 2014. Petrolio, *OPEC: niente tagli a produzione, anche se scenderà a 40 $ al barile*. December, 14.

Sterman, J. D., 2000. *Business dynamics: systems thinking and modeling for a complex world*. Irwin/McGraw Hill, Boston, MA.

Suratman, N., 2015. *OPEC to welcome back Indonesia as member after seven-year absence*. ICIS, July 14, Singapore.

U.S. Energy Information Administration, 2014. *International Energy Outlook*. Washington DC, USA.

U.S. Energy Information Administration, 2015. *www.eia.gov*.

Weiqi, L., Linwei, M., Yaping, D., Pei, L., 2011. *An econometric modeling approach to short-term crude oil price forecasting*. 30th Chinese Control Conference, Yantai, China.

Wittman, D., 2015. *Fisher matrix for beginners*. Available online at http://www.physics.ucdavis.edu/~dwittman/Fisher-matrix-guide.pdf.

Yao, K., 2014. *China third-quarter GDP growth seen at five-year low of 7.3 percent, more stimuli expected*. Reuters, October 10, Beijing, China.

Ye, M., Zyren, J., Blumberg, C.J., Shore, J., 2009. *A short-run crude oil forecast model with Ratchet effect*. Atlantic Economic Journal 37, 37-50.

Ye, M., Zyren, J., Shore, J., 2005. *A monthly crude oil spot price forecasting model using relative inventories*. International Journal of Forecasting, 21, 491-501.

Ye, M., Zyren, J., Shore, J., 2006. *Forecasting short-run crude oil price using high- and low-inventory variables*. Energy Policy, 34, 2736-2743.

Zagaglia, P., 2010. *Macroeconomic factors and oil future prices: a data-rich model*. Energy economics, 32, 409-417.

Zhang, X., Lai, K. K., Wang, S. Y., 2008. *A new approach for crude oil price analysis based on empirical mode decomposition*. Energy Economics, 30(3), 905-918.

Ringraziamenti

A questo punto del lavoro sono doverosi i ringraziamenti istituzionali. Il primo ringraziamento va al Prof. Davide Manca per gli insegnamenti, i consigli e la disponibilità dimostrata durante l'attività di tesi. La ringrazio per l'efficienza e la rapidità con cui ha corretto ogni minima parte di questo testo, dispensandomi innumerevoli consigli lessicali, grammaticali e stilistici relativi alla lingua inglese, che mi saranno sicuramente utili nell'immediato futuro. Non posso che essere orgogliosa di aver portato il mio contributo al PSE-Lab su un argomento così interessante e *challenging*.
Ringrazio anche Roberto per aver condiviso con me il suo ufficio e per aver sempre fornito risposte esaurienti a tutti i miei interrogativi.

È giunto il momento di ripercorrere la mia vita al Poli. Non è la fine della tesi. Non è la fine dell'università. È giunta la fine di qualcosa di più, la fine di una fase della vita, forse la più bella. I cinque anni a Milano mi hanno segnata e ora questa città la sento più mia di quando sono partita quella domenica pomeriggio di settembre per iniziare questa nuova esperienza. E tutto questo affetto è merito delle persone che mi sono state accanto negli ultimi cinque anni. Molte anche da tanto tempo prima.
La prima citazione più che un ringraziamento è una dedica di questo lavoro. Ai miei genitori. Non so come ringraziarvi per il supporto che mi avete sempre dato, la fiducia dimostrata, la vostra presenza in questi ventiquattro anni. Il traguardo raggiunto non è solo merito dei miei sforzi, ma anche vostro.
Un ringraziamento a mio fratello Giulio. Anche se non te l'ho mai detto, sei stato sempre per me un esempio. Grazie per l'affetto dimostrato, per il supporto (non ultimo l'aiuto nei compiti di matematica al liceo e per le incitazioni quando non ero convinta della strada intrapresa) e per avermi spianato la strada da Ingegnere. Un enorme grazie anche a mia cognata Erika e alla piccola Delia, che tra qualche anno leggerà le parole piene di affetto della zia. Nonostante gli impegni universitari mi abbiano spesso allontanata da voi (la mole di questo lavoro è una spiegazione!) ci tengo a scrivere quanto vi voglio bene.

Un ringraziamento a tutti i parenti, ai miei zii e cugini, che condividono con me la gioia di questo traguardo. Vorrei ricordare anche le mie nonne con cui non ho potuto condividere i successi e insuccessi di questi ultimi anni, ma so che sarebbero orgogliose di me.
Nonostante sia contenta di essere arrivata fino in fondo, ora come non mai vorrei rivivere uno a uno i giorni trascorsi dietro i banchi del Poli. Un ringraziamento ai miei amici di università, che più di tutti gli altri hanno condiviso con me la gioia e i sacrifici di quei giorni, ma sappiamo anche quanto è stato bello trascorrerli insieme. Sono convinta che questo percorso ci regalerà grandi soddisfazioni nel futuro, soddisfazioni che spero continueremo a condividere (davanti ad un aperitivo, perché no?!?). Meritano di essere ricordate le fatiche per organizzare le nostre uscite e il divertimento delle serate passate insieme, lo studio matto e disperato nella biblioteca di chimica e le pause ai divanetti, il nervosismo prima degli esami e la gioia per ogni successo che ci avvicinava al traguardo. Ognuno di voi merita un ringraziamento per i meravigliosi ricordi che mi ha lasciato. Anna Sofia per essere sempre stata un'amica sincera dal primo momento in cui ci siamo conosciute nell'aula C.1.1 e per avermi sopportata negli ultimi tre anni trascorsi da coinquiline in Via Pacini. Francesca per essermi stata vicina con i tuoi consigli e sempre presente quando era il momento di uscire e far festa. Fabio per tutti i consigli che mi hai dispensato e per avermi allietato ogni giorno con innumerevoli messaggi, video, foto su qualsiasi social immaginabile. Sono sopravvissuta al Poli nonostante i nostri buoni propositi all'inizio di ogni semestre andassero sistematicamente in fumo dopo una settimana circa. O forse sono sopravvissuta proprio per quello? Mattia per aver sempre avuto un buon consiglio (e gli appunti!) da condividere; non posso che augurarti il meglio per la carriera che intendi intraprendere. Paolo per essere stato sempre pronto a portare una ventata di ottimismo durante la preparazione degli esami (a parte per Complementi di Analisi Matematica). Martino per essere stato la prima persona a parlarmi cinque anni fa e per aver allietato le interminabili ore di Analisi 1. Ivan per essere stato presente nelle uscite più indimenticabili della mia (anzi, nostra) avventura milanese. Ci sono tanti motivi per ringraziarvi, ma ho voluto ricordarvi con questi pensieri.
Ringrazio tutti gli altri miei compagni di corso che, in un modo o nell'altro, sono entrati a far parte della mia vita. È stato un onore condividere questi cinque anni con voi. Vorrei ringraziare anche Stefano e Guido, che per tre anni hanno condiviso con me le emozioni, le gioie, i sacrifici da studenti del Poli. Vi auguro il meglio.

Voglio però ringraziare anche chi ha condiviso con me non l'intero percorso ma questi ultimi mesi: gli amici dell'Athens. Daniele perché con il tuo carisma e la tua simpatia sei diventato il mio milanese preferito. Melo per essere stato sempre pronto a organizzare una serata come si deve (a cui puntualmente dovevo rinunciare causa impegni di studio o tesi, ma non era colpa tua!).

Nonostante i giorni trascorsi a Milano, un posto speciale della mia vita e dei miei week end è stato occupato dalla compagnia de "I soliti" di Cremona. Grazie per le serate tranquille o movimentate che abbiamo trascorso insieme da molto prima che iniziassi l'università. Bume per essere stato il coinquilino con cui ho condiviso i primi due anni a Milano. Non dimenticherò mai i viaggi (spesso interminabili!) da Cremona a Milano Dateo, i racconti delle nostre giornate universitarie e i discorsi pseudo-filosofici che accompagnavano le cene in Via Ceradini. Claudia per l'amica che sei stata, sei e sarai, sempre presente dal liceo fino a oggi, nonostante la distanza che ci separa. Grazie anche a Giulia, Lally, Roby, Simo e Trio. Nonostante abbiamo intrapreso strade diverse che ci hanno allontanato, grazie per avermi regalato delle serate fantastiche.

Infine, modestia a parte, un ringraziamento anche a me stessa che sono arrivata fino a qua. Visto Vale?!? Nonostante i sacrifici, gli sforzi e i dubbi, ne è valsa la pena.

Rileggendo i ringraziamenti mi sono accorta che ho usato soprattutto il *passato* per parlare delle persone che mi hanno accompagnata in questa avventura. A tutti voi: desidero che siate parte anche del mio futuro.

Non so dove questo percorso mi porterà, ma sono sicura che vorrei condividere con ciascuno di voi ogni traguardo e nuovo inizio della mia vita.

Milano, 30.09.2015 *Valentina*