3-2007

Predicting Argentine Jet Fuel Prices

Juan A. Salaverry

PREDICTING ARGENTINE JET FUEL PRICES

THESIS

Juan A. Salaverry, Lieutenant Colonel, Argentine Air Force

AFIT/GLM/ENC/07M-01

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

APROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense or the United States Government.

AFIT/GLM/ENC/07M-01

PREDICTING ARGENTINE JET FUEL PRICES

THESIS

Presented to the Faculty

Department of Mathematics and Statistics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Logistic Management

Juan Angel Salaverry, BSE

Lieutenant Colonel, Argentine Air Force

March 2007

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

AFIT/GLM/ENC/07M-01

PREDICTING ARGENTINE JET FUEL PRICES

Juan Angel Salaverry

Lieutenant Colonel, Argentine Air Force

Approved:

//SIGNED//

Edward D. White III (Chairman)	date

//SIGNED//

Michael J. Hicks (Member)	date

//SIGNED//

Bradley E. Anderson (Member)	date

AFIT/GLM/ENC/07M-01

Acknowledgments

First and foremost, I dedicate this to my wife, my daughters, and relatives for knowing that my job is not an easy one, and doing everything they can to make it easier. Special thanks to my parents, who can always be depended on for help.

I'm also very grateful to my thesis committee for their help during this research: Dr. Edward D. White III, Dr. Michael J. Hicks and Lt Col Bradley E. Anderson for keeping the leash long and letting me work through the hard stuff and to discover what it means to do research.

I would like to thanks the Argentine Air Force officers who have spent their valuable time providing their professional expertise in the logistic field: Brigadier D. Gabriel Rodino, Comodoro D. Guillermo Piuzzi, Comodoro D. Jorge Guarnieri, and Comodoro (RE) Carlos Sierra. I would like to express my sincere gratitude to my good friends, Vicecomodoro D. Daniel Kolarovic and Major D. Luis Canton, and their families which whom we have shared wonderful moments of our life.

I will not forget Major Robert Anderson, his wife, and their sons which whom we share a beautiful friendship. Finally, I recognize my classmates, who were capable of looking at me as another member of the class without considering me as a foreign student.

Juan Angel Salaverry

AFIT/GLM/ENC/07M-01

Abstract

Oil distillates are considered important elements to accomplish the missions of the Argentine Air Force (AAF). Of all oil products consumed by the AAF, jet fuel is the resource with highest demand and at the end of the day the most expensive support item procured by the Argentine Air Force. Accurate predictions of Argentine jet fuel prices are necessary to improve AAF financial and logistics planning. This thesis presents a systematic, statistical regression approach to forecast Argentine jet fuel prices. This methodology has allowed us to obtain a very useful model that utilizes information available on the internet to produce forecasting with average percentage absolute errors lower than 3%. An adjusted R^2 higher than 0.99 allows us to conclude that the model presents an excellent goodness of fit. Mathematically, the model (after some rounding for display purposes only) can be expressed as:

$\hat{y} = 0.034 + 0.425 \times JFP(L1) + 0.01 \times WTI + 0.00062 \times IPP(O\,\&\,G) + 0.1995 \times Dummy$,

where $\hat{y}$ represent our prediction of Argentine jet fuel price expressed in Argentine pesos per liter, JPF (L1) is the Argentine jet fuel price lagged one month in the same unit of measure, WTI is the West Texas Intermediate in US Dollar per Barrel lagged one month, IPP (O&G) is the Price Index of Argentine-Produced Wholesale Goods for natural gas and oil also lagged one month, and the dummy variable takes the value of 1 for calculations from February 2006 and zero otherwise.

Table of Contents

Page

List of Figures

List of Tables

List of Appendices

PREDICTING ARGENTINE JET FUEL PRICES

1. The Problem and its Setting

Background

Oil distillates are considered important elements to accomplish the missions of the Argentine Air Force (AAF). Of all oil products consumed by the AAF, jet fuel is the resource with highest demand and at the end of the day the most expensive support item procured by the Argentine Air Force. The AAF consumes more than 12 million gallons each year and spends almost 35% of its total material budget in the acquisition of this resource (Argentine Air Force, 2006). High consumption rates, volatility of the prices, and limited storage capacity are only some of the aspects that affect budget prediction of this item.

Crude oil is the main element in the production of jet fuel. Especially during recent years, crude oil price instability has brought additional problems to budget and logistics planning. Inaccurate forecasts over fuel prices can cause major problems in the AAF budget. High jet fuel price predictions result in the AAF receiving more funds than required for this concept, resources that otherwise could be used to meet other priorities. In contrast, low jet fuel predictions mean that the received funds are not sufficient to pay for the cost of fuel, prompting the AAF to either request a supplemental appropriation or transfer funds from another account which produces other significant negative effects over the organization.

Accurate oil predictions are also important to improve AAF strategy to face the contractual relationship with its provider. The AAF is tied to a fixed price contract with a clause of adjustment with a unique provider. Each time international and domestic conditions change, the parties meet with each other to agree upon the price adjustment of the product. For this reason, a model that helps the AAF to accurately predict jet fuel prices would provide an invaluable tool to protect taxpayer contributions.

Individual efforts have been attempted in the AAF to solve this issue such as the use of simple regression models, but the results have never been universally accepted in the organization. Not only is there a lack of understanding of the variables that affect the problem, but there also are difficulties in finding the appropriate tools to address this issue.

The Problem and the Research Questions

Accurate predictions of jet fuel prices are necessary to address a variety of budget and logistics problems that affect the AAF. This thesis attempts to analyze and develop a comprehensive model that allows the AAF to make better predictions of jet fuel price to improve the AAF financial and logistic planning. Taking into account this problem, this thesis seeks to answer the following research question:

- How can the Argentine Air Force better predict jet fuel prices to improve financial and logistic planning?

To answer this question, some critical areas should be analyzed. Distinguishing the appropriate factors that exert influence over jet fuel prices, the methods that can be used to predict jet fuel prices, and the data that are necessary and available to build any forecasting model are important components of the problem that have to be considered. The following research subproblems should be addressed to find the solution to the established research question:

- What are the necessary variables to introduce in the model to predict jet fuel price in Argentina?
- What are the necessary data to solve the problem? Are they available?
- Can jet fuel prices be adequately predicted using multiple regression models?
- Would a multiple regression model provide a useful planning and decision aid for the Argentine Air Force?

The answers to these questions would help us to address the purpose of this thesis in a manageable and systematic form.

Summary of Current Knowledge

Predictions of prices have always been a challenge for analysts. This is particularly true in the case of the prediction of oil and its subproducts. Several methods have been used to predict jet fuel prices with varied results over the years. Artificial networks (Kasprzak, 1995), multiple regression models (United States Department of Energy, 2002) and econometric forecasting (Coloma, 1998; Mercuri, 2001) have proved to be effective to forecast oil distillates prices like gas, fuel oil and jet fuel prices. All these models have been developed to forecast the variable of interest in the particular environment of the market of reference. Despite this, the direct application of these models to the particular conditions of the Argentine market to forecast jet fuel prices has brought meager results.

In the same way, the variables used to build comprehensive models to improve oil predictions include an ample range of domestic and international factors depending on the forecaster. The domestic factors we should consider comprise the particular conditions of the market in the analyzed country, including supply and demand relationships, domestic policies, inflation rates, and production capacity; the international factors involve aspects that are related with the international conditions of the oil market and how they affect the domestic oil price or the prices of its distillates. Understanding these domestic and international elements is critical for building and interpreting a prediction model to forecast jet fuel prices.

Assumptions

One of the most important aspects of all problem solving strategies is to establish the assumptions involved with the problem to be solved. Assumptions are propositions taken for granted; they are an integral part of the problem and have to be defined and treated carefully.

First of all, we know that, for its own characteristic, constructing a multiple regression model implies the use of a large amount of data. These data have to be classified and analyzed in an appropriated form to reach positive results. We assume that the required data will be available, accurate and complete. The data provided by the Argentine Secretary of Energy, the Argentine Institute of Statistic and Census (official statements of the Argentine government), and Platts, Corporation (a worldwide provider of oil market information) will allow us to approach this problem with greater probability of success.

The second assumption is related to the Argentina economic policy. The country has suffered from a variety of economic problems. Extremely high inflation rates, continuous changes in economic policies, and modifications of the rules of the game of the market can be considered normal in the history of the country. These elements have made it difficult to introduce models that assist in making domestic predictions about the future of any assets; this fact is especially true for the case of forecasting prices of oil and its subproducts.

In spite of historical instability, during the last five years the country has achieved an economic stability which can be expected to continue in the future. Forecasting oil prices in a chaotic environment can be difficult. For this reason, this work is based on the assumption that current stable Argentine economic conditions will continue.

Finally, independent of the model chosen to predict jet fuel prices, some assumptions, inherent to the model, have to be met. These are going to be described in future chapters when we explain in depth the research methodology.

Scope and Limitations

The scope of this work is limited to forecasting Argentine jet fuel prices. This means that, except where reasonable data for a required variable do not exist, any other response factor will not be forecasted. This limitation does not mean that only variables inside the Argentine environment will be considered. Predicting Argentine jet fuel price will involve analysis of the behavior of some variables in the international sphere and the influence that they exert in the domestic price of jet fuel.

Approach and Methodology

A statistical analysis will be used to answer the research question. The chosen methodology to predict Argentine jet fuel prices is a multiple regression model based on historical data provided by the Argentine Secretary of Energy, the Argentine Institute of Statistics and Census, and a worldwide provider of oil market information: Platts, Co. Later we describe in depth the methodology used to investigate the research question. However, the following paragraphs summarize the methodology.

As a first step, based on the review of applicable models and expertise opinions, a preliminary set of domestic and international variables that are supposed to influence Argentine jet fuel prices will be preliminary selected. In the second step, data provided by the Argentine Secretary of Energy, Argentine Institute of Statistic and Census, and Platt, Co. will be collected to perform a stepwise analysis to determine the variables that are the best predictors to forecast jet fuel prices in Argentina. Only eighty percent, randomly selected data will be used to build a multiple regression model using the least squares approach, reserving the remaining twenty percent of the data to validate the model.

The validated model will permit the AAF to introduce a model to accurately predict jet fuel prices inside the Argentina environment. The appropriate use of this model would allow the AAF to improve its financial and logistics planning. Jet fuel not only represents an important asset to accomplish the Argentine Air Force mission, but it is also the resource with highest demand and the most expensive item procured by the Air Force. Therefore, accurate prediction of its price is a necessity to improve financial and logistic planning.

Understanding the Argentine oil market, the potential predictor of jet fuel prices and different methodologies that have been applied to forecast oil prices and its derivates will help us to overcome our first step in the process; they are the goals of Chapter 2 of this work. Chapter 3 describes in depth the multiple regression techniques, which have been selected as methodology to predict Argentine jet fuel prices. In Chapter 4, the selected methodology is applied to the specific case of Argentina jet fuel market; this chapter shows us the analysis of the data and the model that results from this analysis.

Finally, conclusions, model applications, and limitations, as well as future areas of interest in the addressed topic are included in Chapter 5.

2. Literature Review

Introduction

Jet fuel is a light oil distillate obtained by a chemical process called hydrocraking; it is normally defined as: *"a high-quality kerosene product used primarily as fuel for commercial turbojet and turboprop aircraft engines"* (New York Mercantile Exchange Glossary, 2001:24).

Jet fuel is considered not only an important element to accomplish the mission of the Argentine Air Force (AAF), but it is also responsible for the largest amount of its material budget. The influence that this element exerts on the AAF budget demands accurate prediction of its price. To increase its budget efficiency, the AAF should improve its financial and logistics planning, and to do that the development of a comprehensive model that helps the AAF to predict jet fuel prices is required.

The purpose of this literature review is to increase the understanding of the problem and its importance for the AAF, to analyze the Argentina oil market situation, and finally to introduce the reader to the variables that could be considered in a potential model to predict jet fuel prices in Argentina.

Understanding the Problem and its Importance for the AAF

Of all oil products consumed by the AAF, jet fuel is the element with highest demand and in the long run the most expensive support item procured by the Argentina Air Force. The AAF consumes more than 12 million gallons of jet fuel each year and spends almost 35% of its annual material budget in the acquisition of this resource (Argentine Air Force, 2006). The material budget includes all the funds that are necessary to acquire the required assets to support the flight activity.

Accurate oil price predictions have not been easy to achieve since the oil embargo occurred in 1973, and have been the focus of several international studies (Burke, 2005). Especially during the last decade, jet fuel prices have been extremely volatile, led by the erratic behavior of crude oil prices, the main component in jet fuel production. Figure 2-1 illustrates the erratic behavior of crude oil prices (WTI) and jet fuel prices (JetKero 54) from 1994 to 2006. The data were extracted from Platts, Co., one of the largest companies in the world that provides oil market information.

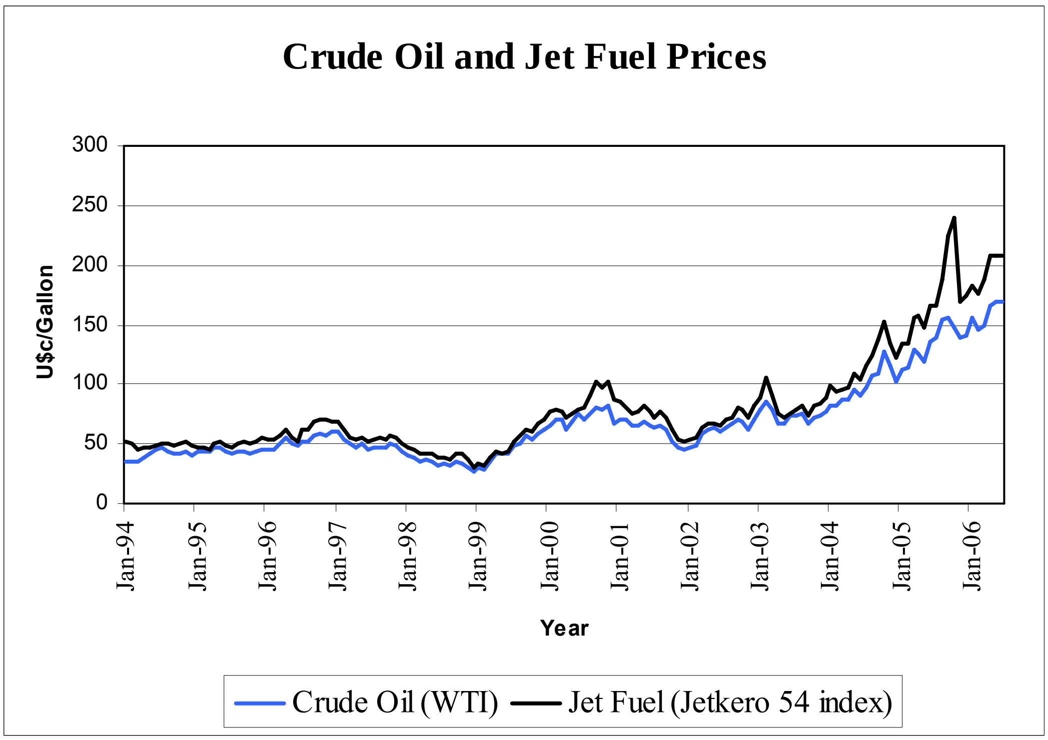

Figure 2-1: Crude Oil and Jet Fuel Prices 1994-2006
(Source Platts Co., 2006)

Volatility of jet fuel prices, high consumption rates, and limited storage capacity are some of the aspects that affect jet fuel budget prediction. The U.S. General Accounting Office (GAO) highlights the importance of better fuel pricing practices to improve budget accuracy (United States General Accounting Office, 2002). In its report of June, 2002, the GAO highlights the problem produced by inaccurate fuel predictions and their consequences in the official budget system. As the document indicates, bad oil price predictions, added to the volatility in crude oil prices, have affected the cash balance flow of the budgeted funds to acquire oil and its derivatives; these facts have also increased the necessity of transferring funds from one account to another increasing the difficulties to provide the rationale for cash movements to the Congress.

The AAF suffers from the same problem with the same consequences. When jet fuel prices are predicted higher than their real value, more funds than required are received thereby diminishing other priorities; on the other hand, when jet fuel prices are predicted lower than their true value, less funds than required are received, which requires the AAF to transfer funds from one account to another or to request supplemental appropriations.

But perhaps this is not the most important issue related to the necessity of good jet fuel price predictions. A long-term, fixed price contract with an adjustment clause ties the AAF to its unique provider, REPSOL-YPF S.A. When international and domestic circumstances change, contractually the parties are called to discuss the required adjustments in jet fuel prices that will apply until the next change in conditions. Accurate jet fuel price prediction will help the AAF to trace a better strategy that would help to protect taxpayer contributions.

Better jet fuel price predictions will help the AAF to improve financial and logistic planning as well as to increase its budget accuracy and to achieve a better efficiency of the contractual relationship with its fuel provider. In the next section, we analyze the characteristics of the Argentine jet fuel market.

The Argentine Jet Fuel Market

An Analysis of the Argentine Oil History and its Consequences

The lack of a vision is undoubtedly the main explanation for individual and corporate failures. It is difficult to think of any person or organization that has sustained some measure of greatness in the absence of goals, value and missions deeply inside the person or organization. In his book *The Fifth Discipline,* Peter Senge observes that: *"When there is a genuine vision, people excel and learn, not because they are told to, but because they want to"* (Senge, 1990:9).

Countries are not different from individuals and organizations in these aspects; they need visions that have to be transformed into objectives and policies by governments to achieve the well-being of their people. The lack of vision restricts the possibility of developing and sharing images of the future they want to create and the principles and practices by which they hope to get there (Senge, et al. 1999:32).

The history of oil in Argentina has suffered from this problem; since its beginning on December 13th, 1907, when Humberto Baghin and Jose Fuchus, drilling for water in the city of Comodoro Rivadavia – Chubut, found oil, the lack of a clear vision has been the principal characteristic of the Argentinean oil policy (Gadano and Sturzenegger, 1998). Since that date, the history of oil in Argentine has been associated with the ups and downs of the Argentine public policy. In practice until 1989, the Argentine oil industry was always under the strong influence of the state, limiting private participation in the sector (Gadano and Sturzenegger, 1998).

The 1989 economic crisis in Argentina found the oil sector in one of its more difficult moments. High foreign debt of the public company and a remarkable incapacity of increasing production rates showed that the ability of the country to maintain oil self-sufficiency was only a dream (Gadano and Sturzenegger, 1998).

To overcome the situation of the energy sector, Argentina initiated that year a series of privatization actions in its oil sector. By 1993, the country had totally privatized its oil production and exploitation. YPF, the main oil company owned by the state, had been transferred to the Spanish company REPSOL (Gadano and Sturzenegger, 1998).

Introducing these changes has not been an easy task; arguments in favor of and against the privatization process can be heard even today, thirteen years after the starting point of the process. Any comparison between the pre- and post- privatization periods has suffered from some partiality in the analysis. Although judging the privatization process is not the goal of this work, we need to evaluate some of the main results of this process if we want to understand the current behavior of the Argentina market.

Figures 2-2 and 2-3 illustrate the evolution, from 1994 to 2005, of two of the most important indicators of the Argentine oil market: crude oil exportation and total oil production. We can observe that the country on the average has almost doubled its oil production. This level of production has allowed the country to achieve self-sufficiency, to respond to the domestic increase in demand, and to increase its level of exportations (Argentine Secretary of Energy, 2006). In that process, jet fuel has followed a similar pattern; since 2002, the country has reached self-sufficiency and today the product is exported to other countries (Argentine Secretary of Energy, 2006). Although some criticism of the privatization process can be made, it is clear that the changes to the Argentina oil sector have begun to provide dividends to the country.

The Current Characteristics of the Market

The new characteristics of the Argentine oil market have established the country as a non-OPEC (Organization of the Petroleum Exporting Countries) producer. Several studies have analyzed price practice models depending on whether the country is considered an OPEC producer or a non-OPEC producer (Dees, et al. undated; Ramcharran, 2002). Independently of the assumptions used to develop the models, generally the authors agree that in contrast to OPEC countries, non-OPEC countries behave as price takers instead of price formers when they sell their product on the international market.

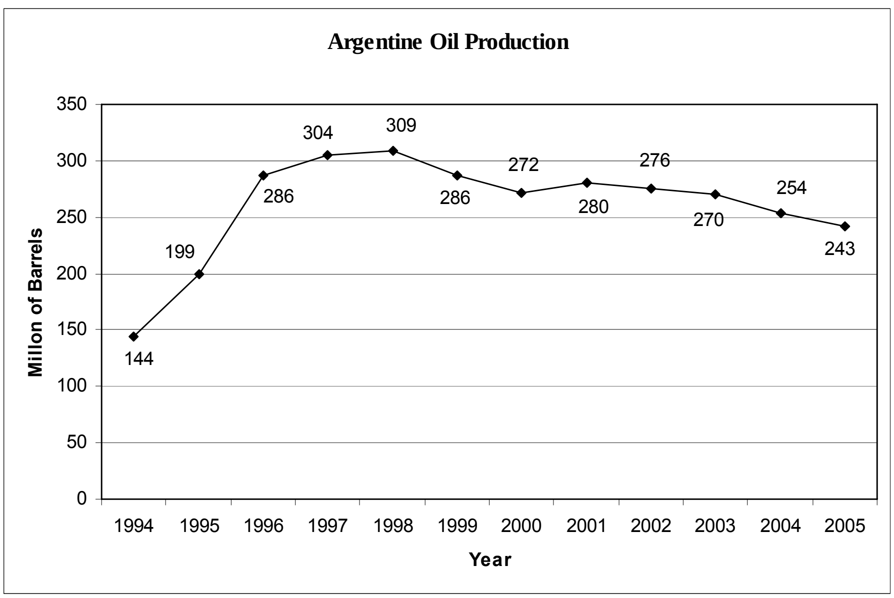

Figure 2-2: Argentine Oil Production, 1994-2005
(Argentine Secretary of Energy, 2006)

With respect to the domestic market, specifically for Argentina, studies were developed to evaluate the behavior of the oil market. Although the studies are exclusively based on analysis conducted on gasoline and diesel fuel, which are the oil derivates with the highest level of consumption rates, they also clearly emphasize that fuel prices are highly correlated with international oil prices (De Dicco, 2004; Mercuri, 2001; Coloma, 1998).

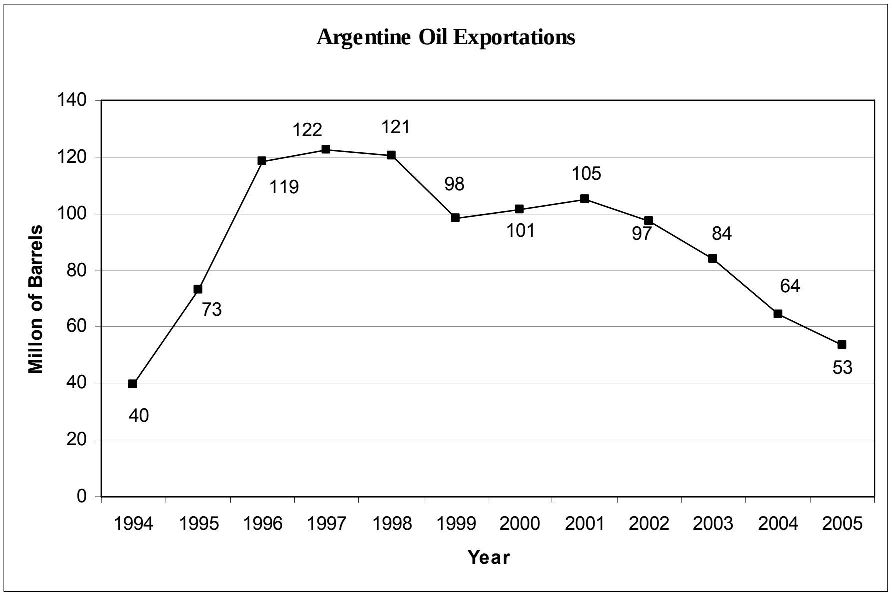

Figure 2-3: Argentine Oil Exportation, 1994-2005
(Argentine Secretary of Energy, 2006)

Some interesting conclusions can be drawn from the De Dicco study. In his work, De Dicco analyzes how domestic oil prices are related with domestic production costs and international oil prices. De Dicco concludes that while production costs (finding, development and lifting cost) in the country have been pretty stable over the last 4 or 5 years (around 7 dollars per barrel), the domestic costs that companies use to price oil derivatives for internal consumption have followed the increase of prices of crude oil in the international market. According to this finding, the behavior of the Argentine jet fuel market and the pricing policies used by the companies reflect fluctuations in the international oil market.

Other factors have been the targets of studies of the Argentine oil market. From these, refining capacity (Coloma, 1998), seasonal variables like cool weather, and variation in domestic and international stock levels (Scheimberg, 1998) have been also indicated as factors that exert some influence over oil prices and the prices of oil derivatives.

One other important factor in analyzing the Argentine oil situation is its market concentration. The total jet fuel supply in the Argentine market is limited to three companies: REPSOL-YPF S.A., ESSO S.A.P.A, and SHELL C.A.P.S.A. Figure 2-4 shows that the Argentina jet fuel market is highly concentrated around REPSOL-YPF, which has dominated the market over time.

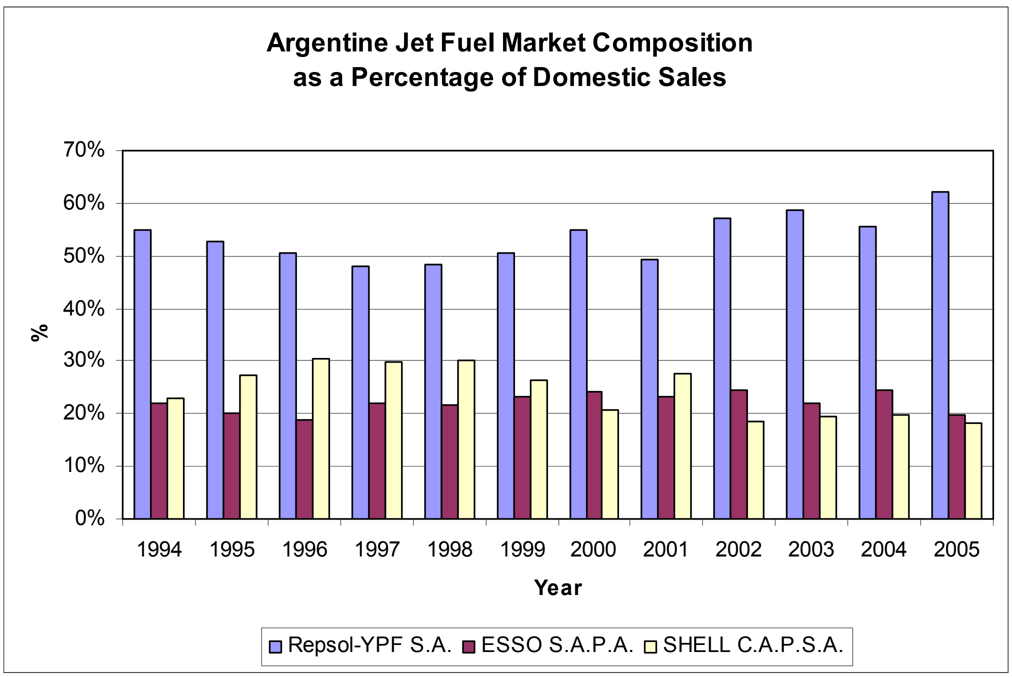

Figure 2-4: Argentine Jet Fuel Market Composition 1994-2005
(Argentine Secretary of Energy, 2006)

While some authors, like Coloma, have developed models that support the existence of competitive behavior inside the oil market and have concluded that market concentration does not exert any influence on fuel prices, other authors, like Mercuri, have concluded that there is not enough evidence to support a competitive behavior in the Argentine oil market (Mercuri, 2001).

In conclusion, since 1993 Argentina has initiated radical changes in its oil policy. The process used to implement these changes is beyond the scope of this study, but an understanding of its consequences is necessary to understand in depth the oil market structure of the country. Some of the characteristics of the Argentina market analyzed here will help us to define what factors should be considered in the future development of a model to predict Argentine jet fuel prices. The next section presents some models and potential predictors used to forecast jet fuel prices.

An Overview of the Models and Predictors Used To Forecast Jet Fuel Prices

Forecasting Models for Oil Prices

After having developed a broader understanding of the Argentina oil market behavior, we now examine some of the models that have been used to predict jet fuel prices and the factors that have been included in their development. At the same time, it is important to realize that we are looking for a comprehensive model to predict jet fuel prices in Argentina. A comprehensive model refers to a model that is easy to understand, practical and useful. As we know, models are only simplifications of the real world, and these simplifications are necessary because otherwise they would be as complex and unwieldy as the natural setting itself (Michalewicz and Fogel, 2004).

Although over time many complex, often intractable models have been created to predict oil prices and its derivatives, artificial neural networks (Kasprzak, 1995), econometric forecasting and intertemporal optimization (Powell, 1990; Gately, 1995), and multiple regression models (United Stated Department of Energy, 2002) have shown very good results when used to forecast jet fuel prices in the United States market. These will be described briefly. Understanding the general idea behind each of the analyzed techniques can help us to understand how to face the problem of developing a comprehensive jet fuel predicting model for the AAF. The list of analyzed models does not pretend to be exhaustive; some of the most known methods have been chosen.

Econometric forecasting is perhaps one of the earliest methods developed to forecast the prices of oil and its derivatives. The technique is based on the use of regression analysis to construct a cause and effect map that helps to predict the analyzed dependent variable. The necessity to find causality forces analysts to choose from a large variety of variables which affect the model's complexity and the number of required equations to predict results. It is important to recall that statistics techniques capture correlation, not causation. Correlation is only one of the elements required to establish a cause and effect relationship between two variables; showing that precedence exists and removing all the other alternative explanations are also necessary conditions (Leedy and Ormrod, 2005:181-182).

For that reason, no single rule exists to build the model; models representing the same phenomenon vary in their forms, involve different variables, and are composed of a varied number of equations. Econometric forecasting has proved to be effective in samples but not to extrapolate out of them (Burke, 2005).

On the other hand, the application of intertemporal optimization to forecast oil prices is based on three assumptions in relation to the owner of oil: perfect knowledge, perfect foresight, and maximum return of investment as a goal; intertemporal optimization is rooted in Hotelling's model of depletable natural resources. The theory behind the model offers a rational explanation of the actors in the model, but its unrealistic assumptions have made difficult its application to solve real world problems (Powell, 1990; Gately, 1995).

Artificial Neural Networks (ANN) is an information processing paradigm that is based on the manner in which biological nervous systems work to process information. ANN is a technique that has been applied to forecast jet fuel prices by Mary Kasprzak in 1995 with results comparable to the National Energy Modeling System. The key element of this model is the existence of a large number of highly interconnected processing elements working in unison to solve specific problems (Stergiou and Siganos, 1996).

As these authors indicate, the utility of artificial neural network models lies in the fact that they can be used to infer a function from observations. This is particularly useful in applications where the complexity of the data or tasks makes the design of such a function by hand impractical, as is the case of oil derivatives. The main drawbacks are: the requirement of specific software packages, high level of training, and unpredictable behavior when the network is poorly designed.

Behavior simulation is a dynamic model that has been developed incorporating system dynamics and the bounded rationality school of thought. Its dynamism permits the model to embrace the uncertainty of the market, which is useful to show how the market changes over time. The U.S. National Energy Modeling System (NEMS) has designed a behavioral simulation model to represent the important interactions of supply and demand in U.S. energy markets. The description of the system establishes that: "NEMS represents the market behavior of the producers and consumers of energy at a level of detail that is useful for analyzing the implications of technological improvements and policy initiatives” (United States Department of Energy, 2003:4). NEMS is composed of several modules, one of which is used to predict the prices of oil derivatives. Jet fuel prices are predicted using the Short-Term Integrated Forecasting System (STIFS), which will be described next.

A Brief Analysis to the U. S. Short-Term Integrated Forecasting System (STIFS)

The U.S Department of Energy through the Energy Information Administration has developed the Short-Term Integrated Forecasting System (STIFS) as a part of its Integrating Module of the National Energy Modeling System. STIFS allows the U.S. Government to generate short-term (up to eight quarters) monthly forecasts of U.S. supplies, demands, imports, stocks, and prices of various forms of energy (United States Department of Energy, 2002).

In a broad sense, the STIFS model comprises more than 300 equations, of which over 100 are estimated. The estimated equations are linear regression equations interrelated to provide a system of forecasting equations. The estimation techniques are generally done on an equation-by-equation basis using the least squares method (United States Department of Energy, 2002).

In the specific case of jet fuel, the price is estimated through the use of the following linear regression model:

$$P_{jetfuel} = \alpha_0 + \alpha_1 P_{jetfuel_{t-1}} + \alpha_2 P_{crudeoil} + \alpha_4 C_{DUM} + \alpha_4 \frac{S_{jetfuel_{t-1}}}{D_{jetfuel_t}} + \alpha_5 I_{t-1}, \qquad (2\text{-}1)$$

where $P_{jetfuel}$ is the average retail price of jet fuel; α_0, α_1, α_2, α_3, α_4, and α_5 are the regression coefficients of the model; $P_{jetfuel_{t-1}}$ is the average retail price of jet fuel lagged one month; $P_{crudeoil}$ is the price of crude oil; C_{DUM} is a dummy variable that represents the period of December 1989 through January 1990, when cold weather caused oil product prices to go up; $S_{jetfuel_{t-1}}$ is the previous month's jet fuel supply; $D_{jetfuel}$ is the projected jet fuel demand for the coming month; and I_{t-1} is the wholesale price index for non-energy products as a measure of inflation. Overall, equation 2-1 calculates the price of jet fuel in a linear regression equation using previous month's jet fuel price, current crude oil price, a relation between previous month's jet fuel supply and current estimated jet fuel demand, and an economic indicator of inflation as predictors.

What Can Influence Argentine Jet Fuel Prices?

Having analyzed the Argentina jet fuel market and some of the more common methods use to predict oil prices and its derivatives, it is time to analyze what variables can be used as predictors to forecast jet fuel prices. As can be observed, the election of the methodology to approach the problem influences the amount of data required to obtain a comprehensive model capable of describing reality accurately.

Before mentioning the variables that have often been used to predict jet fuel prices, two elements have to be stated: first, oil reserve estimates are problematic and confusing (Cavallo, 2003); second, the market price of oil is decoupled from the production cost (De Dicco, 2004; Cavallo, 2003). Based on these statements, we can conclude that the market price for oil derivatives does not reflect neither how rapidly reserves are being consumed by society nor the influence of production costs in the supply chain.

With this in mind, from the analyzed methods some interesting conclusions with respect to predictor variables can be drawn:

1. As a crude oil derivate, jet fuel prices have shown strong correlation with crude oil prices and all the factors that affect oil prices (government policies, economic growth, energy demand and supply) (Kasprzak, 1995: 3-4; U.S. Department of Energy, 2002; De Dicco, 2004; Mercuri, 2001; Coloma, 1998).
2. Supply is influenced by the total capacity to produce jet fuel and the relation of this product with other oil products that are obtained with the same process from the same basic product (crude oil). In that sense, heating oil has been highlighted as a good predictor for jet fuel (Kasprzak, 1995:3-4; BMO Commodity Derivatives Group, 2005).
3. Supply and demand for the product are also influenced by causes related to seasonality and natural disasters (Kasprzak, 1995: 3-4).

Summary

Jet fuel is not only an important asset that allows the AAF to accomplish its mission, but it is also responsible for the largest amount of its material budget. The influence that this element exerts on the AAF budget demands accurate prediction of its price to improve financial and logistics planning. A complex environment characterized by high volatility in prices, high consumption rates, the lack of understanding of the variables that influence jet fuel prices, and the ways in which these variables are interrelated have made it difficult to predict jet fuel prices in the AAF. Several models have been developed to forecast jet fuel prices in the world; they have been developed considering the particular conditions of the market where the models will be applied; conditions that differ from the particular characteristics of the Argentine market turning the application of those models inappropriate. Finding a comprehensive model to predict jet fuel prices in Argentina is a real challenge.

From all the analyzed methodologies, multiple regression analysis is widely accepted in several, very different disciplines such as business, economics, engineering, and the social and biological sciences (Kutner, 2005:2), but successful application of this method requires not only a deep understanding of the underlying theory, but also its practical uses. For that reason before introducing the readers to the analysis of the data to address our research question, Chapter 3 reviews the theory behind multiple regression methodology, its assumptions and limitations, and outlines of the model-building and model-validation process.

3. Methodology

Introduction

Good forecasts enable management to achieve effective and efficient planning. As defined in Chapter 1, predicting Argentine jet fuel prices is essential to improve the financial and logistic planning in the Argentine Air Force. Having defined the problem and its setting in Chapter 1, Chapter 2 helped us to understand the problem inside the Argentine environment, and also to investigate models and potential predictors that have been used to forecast jet fuel prices.

Chapter 2 shows us that an ample array of forecasting methods is available to forecast jet fuel prices. These methods range from the easiest ones to highly complex approaches such us econometric forecasting or neural networks. Complex environments such as predicting oil prices cannot easily be simplified to the application of the simplest forecasting method, and normally requires the analysis of several variables under specific conditions and assumptions. During this chapter we will analyze in depth the multiple regression technique that has been chosen to investigate our research question. We will take a look at the required assumptions that form part of the methodology and how we plan to meet them, some approaches to build useful models with this technique and the validation process to be implemented.

What is a multiple regression model?

In general, regression analysis models the relationship between one or more response variables (also called predictive or dependent variables) and a number of predictors (also called explanatory or independent variables) (McClave et al. 2005:694). The association between only one dependent variable and a unique independent variable is called simple regression, while the use of a set of explanatory variables to predict the behavior of a response is known as multiple regression. It can be established that multiple regression models are probabilistic models in which the behavior of a dependent variable (predictive) is influenced by more than one independent variable (predictors), and that simple regression models could be understood as a simplification of multiple regression (McClave et al. 2005:768-769; Makridakis et al. 1998:248-249).

One of the most important advantages of multiple regression models is that they allow analysts to include both quantitative and qualitative variables in the model (McClave et al. 2005:825). This is not a minor point in some fields like economics and human science where regression models are normally applied. Qualitative variables, also called categorical, indicator and most commonly dummy variables, cannot be measured on a numerical scale as quantitative variables. These variables are used to introduce in the model discrete events as seasonality effects, and holidays like Christmas and Thanksgiving for example (McClave et al. 2005:825), and through them estimate the effect these events have on the response variable. Dummy variables are normally coded as 0 or 1 depending on the studied event has influence or not.

Independently of the inclusion of qualitative and quantitative variables into the model, the general, mathematical form of a multiple regression model can be written as:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \ldots + \beta_k x_{k,i} + \varepsilon_i \,, \qquad (3.1)$$

where y represents the dependent variable (in our case Argentine jet fuel prices), i=1,…,n represent subjects, β_0, …, β_k are the regression coefficients, x_1, …, x_k symbolize the independent variables or predictors and ε is a error term that captures the effects of all omitted variables. Equation 3.1 can also be expressed in vectorial form as:

$$\boldsymbol{y}=\boldsymbol{x\beta}+\boldsymbol{\varepsilon}, \qquad (3.2)$$

where $\boldsymbol{y}$, $\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}$ are the $nx1$, $px1$, and $nx1$ vectors that represent the dependent variable, the regression coefficients and the errors respectively and $\boldsymbol{x}$ is the nxp design matrix that symbolizes what we want to introduce in the model to explain the behavior of our dependent variable. Equation 3.2 can be divided in two parts: a deterministic portion (the product of the β coefficients and the independent variables x), and a probabilistic portion represented by the error term (ε), which represents a random error (McClave et al. 2005).

The set of β coefficients indicates the contribution of each independent variable and has to be estimated from the data. Several methods can be used to estimate the β parameters; one of the most common approaches is known as the method of ordinary least squares (OLS). This method is based on finding the set of β coefficients that minimizes the sum of squared errors (SSE), which is defined as the difference between the observed value (y_i) and the estimated value using the regression model ($\hat{y}_i$).
The least squares approach can be expressed mathematically as minimizing

$SSE = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$, where $\varepsilon_i = y_i - \hat{y}_i$ and $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \hat{\beta}_2 x_{2,i} + \ldots + \hat{\beta}_k x_{k,i}$.

In vectorial form, the regression coefficients, $\hat{\beta}_0$, $\hat{\beta}_1$, …, $\hat{\beta}_k$ can be calculated through the following expression:

$$\underline{\hat{\beta}} = (x^T x)^{-1} x^T \underline{y} \qquad (3.3)$$

The application of OLS is subject to the accomplishment of the following assumptions that involves not only the data but also the probability distribution of the random error:

1. **Continuity**: This assumption implies that the distribution of the dependent variable is relatively continuous. Histograms and steam-and-leaf plots of historical data collected for the response variable can be used to test this assumption.
2. **Linearity**: Normally, it is assumed that the relationships between response and each predictor are linear. Confirming this assumption is not an easy matter, but fortunately multiple regression procedures are not greatly affected by minor deviations from this assumption. However, scatter plots help analysts not only to draw conclusions about the nature and the strength of the bivariate relationships between each of the considered predictor and the response variable, but also to identify the type of relation that exists between them (Kutner et al. 2005:232). If curvature in the relationships is evident, mathematical transformations can be applied to the variables to simulate the behavior of the relationship, which means introducing non-linear terms in the regression model.
3. **Normality**: It is assumed that the model residuals (random errors) are normally distributed with mean zero and constant variance. Departures from normality are not serious except when major departures are present (Kutner et al. 2005:110). Several methods have been applied to test this assumption; graphical representations of the residuals and goodness of fit tests are common. For the latter ones Shapiro-Wilks W test, Kolmogorov-Smirnov, and the chi-square test can be used to test normality of the error

terms (Kutner, et al. 2005:215). All of them are tests to determine whether or not a sample comes from a normal distribution; in the case of the Shapiro-Wilks test, it is conducted by comparing the quartiles of the observed data against that of the best-fitting normal distribution (Kutner et al. 2005:216). P-values higher than the chosen level of significance (normally 0.05) allow concluding that there is not enough evidence to reject the hypothesis that the distribution of the residuals is normally distributed. This test is recommended for sample size smaller than 200 data points; for larger samples Kolmogorov-Smirnov test is generally used (Garson, undated).

4. **Independence**: OLS also assumes that the random errors are independent in the probabilistic point of view what means that no correlation or association of the residuals exists. Although this assumption can be difficult to test, if data is gathered at equal intervals of time, the Durbin-Watson test or runs test are useful tools to consider (Kutner, et al. 2005:114, and 487-490). On the other hand, if data is not equally spaced in time, a detail analysis of the scatter plots of the residuals can help to detect any type of patterns or anomalies. In cases where patterns are present in the residuals, it can be an indication of the necessity to introduce new predictors into the analysis; predictors that explain the lack of randomness in the error terms.
5. **Constant variance**: Another OLS assumption requires the residuals to display constant variance; a descriptive plot (response versus residual)

and the Breusch-Pagan test can be used to test this assumption (Kutner, et al. 2005:234-235). Mathematically, it can be expressed as:

$\chi^2_{BP} = \dfrac{\dfrac{SSR}{df}}{\left(\dfrac{SSE}{n}\right)^2}$, where *SSR* is the regression sum of squares when regressing the e^2 against the explanatory variables of the model, *df* is the degree of freedom of the model, *SEE* is the error sum of squares when regressing *y* against the predictors, and n is the number of data points considered to build the model. χ^2_{BP} follows a Chi-square distribution, so p-values higher than the chosen level of significance (0.05) are preferred because indicates that there is no statistical evidence to reject the hypothesis that the residuals display constant variance (Kutner, et al.2005:118-119).

6. **Outliers**: These are data points that lay more than three standard deviations ($\pm 3\sigma$) away from the mean of the distribution of the residuals; this assumption can be met through an analysis of the residual distribution plot. The presence of outliers should require a detailed analysis of the respective data points to look for the causes and their possible implications in the future model building. If the probability that in *n* observations an outlier will be obtained by chance is small, the data point considered an outlier can be eliminated, but otherwise it has to be retained (Kutner, et al. 2005:115, 390-400).

7. **Multicollinearity**: This is a common problem in many correlation analyses and plays a key roll in the regression model. Multicollinearity is present when explanatory variables are correlated among themselves and with other variables related to the response variable not included in the model. When multicollinearity exists the normal interpretation given to the β coefficients is no longer valid. The notion that only one predictor changes by one unit while the others remain constant is not fully applicable when high correlation exists between predictors. As a result a unique solution for the regression coefficients (β*'s*) according to equation 3.3 cannot be found (multicollinearity does not allow us to find a unique solution for the inverse of the matrix in that equation), and so the regression line cannot be calculated (Kutner, et al. 2005:278-284). Multicollinearity is checked through VIF scores (Variance Inflation Factor). These measures compute how much the variances of the estimated β coefficients are magnified compared to the β coefficients when the explanatory variables are not linearly related (Kutner, et al. 2005: 406-410). High VIF scores (higher than 10) implies the presence of linear redundancy in the explanatory variables which has to be removed to avoid this issue (Kutner, et al. 2005:409).

8. **Influential data points**: Finally, the last element to consider is the existence of influential data points in the data. The presence of influential data points can seriously bias the result by "pulling" or "pushing" the regression line in a particular direction. The elimination of these data

points should be taken carefully; we should balance the accuracy of the chosen model against the manipulation of the data to obtain the model. The Cook's distance approach is used to test for the existence of influential data points; Cook's distance values smaller than 0.25 are preferable, values between 0.25 and 0.50 are consider "moderate" influential data points and values greater than 0.50 are considered "major" influential data points (Kutner, et al. 2005:402-403).

Meeting these assumptions is an important step not only during the model validation process, but also important to determinate the precise limits of the chosen model. Once the assumptions are met and the regression coefficients calculated, it is natural to ask if the observed relation between response variable and predictors is significant. The F-test for overall significance has been developed to test that; this statistic measures the relation between the explained mean square (MS) and the unexplained mean square. Mathematically, it can be expressed as:

$$F = \frac{explainedMS}{unexplainedMS} = \frac{\frac{\sum(\hat{Y}_i - \overline{Y})^2}{m-1}}{\frac{\sum(Y_i - \hat{Y}_i)^2}{n-m}}$$, where m is the number of parameters (coefficients) in the model (Makridakis, et al. 1998:211:215). Software packages normally provide P-values of the F statistic. These P-values represents "*the probability of obtaining an F statistic as large as the one calculated for our data, if in fact the true slope is zero*" (Makridakis, et al. 1998:213). As a result small p-values correspond to significant regression and vice versa.

If the overall F-test indicates significant of the regression model, the next step is to analyze whether the term $\beta_K X_K$ can be dropped from the model. In other words we want to know whether the variable X_K is significant for the regression or not. The goal in the end is to produce a significant but parsimonious model. The t-statistic is used to test that. P-values for the t-statistics lower than the chosen level of significance refers to correlation between the dependent and the analyzed independent variable; it means that this particular variable should remain in the model.

As it can be observed, the analyzed methodology depends to a great extent of the chosen variables to simulate the behavior of the dependent variable. The major conceptual limitation of all regression techniques is that one can only ascertain relationships, but never be sure about underlying causal mechanisms. Due to this fact difficulties arise to determine the correct independent variables that could assure a useful regression model. Several approaches have been developed to face this problem, some of which will be discussed in the following section.

Choosing a Useful Model

George Box's adage “all models are wrong, but some models are useful” is appropriate for those who are too incredulous of models, and for those who are not skeptical enough (Box, 1976). Since models are by their nature approximations to a complicated reality, they are of course literally false. But, on the other hand, models are in practice the only instruments we have for understanding complex phenomena.

Building a regression model for real data is not a simple process. The use of regression methodology assumes that we have specified the appropriate model. I.e., we have been able to find an appropriate set of significant and useful independent variables to explain the behavior of our dependent variable (Freund et al. 2003:125-126). The use of expert opinions and other knowledgeable people are very useful in the process; it was one of the goals of our literature review developed in Chapter 2. But, the development of a useful model is also dependent upon the existence of appropriate historical data. It would be pointless to find a "perfect" variable if the data for this variable is not available or difficult to understand.

Having taken into consideration these two elements (expert opinions and data availability) a set of independent variables can be listed and data can be collected. A subset of explanatory variables could be obtained through the examination of all possible combinations of the original set of variables. This could probably give us the best answer, but this procedure could be hard and tedious depending on the number of variables selected. Fortunately, highly efficient algorithms have been developed and are available in several software packages. One of the most recognized methods is known as stepwise regression. A stepwise regression can be used to help sort out the relevant explanatory variables to introduce in the model (Makridakis et al. 1998: 274-279). Three approaches, forward, backward and forward with a backward look regression, have been used to conduct this analysis. The last one of these approaches is more complex, but gives the better results because it involves an iterative process that combines the forward and backward methods (Makridakis et al, 1998:285-286).

The use of stepwise regression normally produces an array of subsets of variables that can be used to model the behavior of the dependent variable. Some statistics have been developed to help in the final selection of the independent variables. The two more useful statistics are the coefficient of determination (R^2) and the C_p statistic proposed by

Mallows (Freund et al, 2003:129). The R^2 coefficient is a measure of how well the predicted values from a forecast model "fit" with the real-life data, and varies from one to zero; models with larger values of R^2 are preferred to models with lower R^2. Mathematically, the R^2 coefficient can be calculated as (McClave et al, 2005:732):

$$R^2 = 1 - \frac{SSE}{SST}, \tag{3.4}$$

where SSE represents the unexplained variance of the dependent variable (the sum of squared errors as defined earlier) and SST is the total variance of the dependent variable. R^2 measures the proportion of the total sample variability that is explained by the model.

Although relevant, the R^2 calculation has a weakness. While the denominator is fixed for a determinate data set for the dependent variable, the numerator can only increase when we incorporate explanatory variables into the regression model; this could result therefore in a higher R^2 even when the new variable causes the equation to become less efficient (worse). In theory, using an infinite number of independent variables to explain the change in a dependent variable would result in an R^2 of one. In other words, the R^2 value can be manipulated and should be suspect (McClave et al, 2005: 792-793).

The statistic called Adjusted R^2 is used to correct this issue; it is done by adjusting both the numerator and the denominator by their respective degrees of freedom. Unlike R^2, adjusted R^2 can decline in value if the contribution to the explained deviation by the additional variable is less than the impact on the degrees of freedom (Makridakis et al. 1998:279-280). Mathematically, adjusted R^2 can be expressed as:

$$r^2 = 1 - \frac{(1-R^2)(n-1)}{(n-k-1)}, \tag{3.5}$$

where r^2 represent the adjusted R^2, n the number of observations, k the number of independent variables and R^2 the initial correlation coefficient. The values of R^2 and the adjusted R^2 are also used to compute the overall fitting of the regression model. By definition these values take into consideration the total deviation explained by the model and the total deviation; so higher values of these statistics are coincident with the least squares methods applied to calculate the regression coefficients.

Cp values defined by Mallows have also been used as a tool to help analysts to look for a decent model. Mallows defined this coefficient as (Freund et al. 2003:129-131):

$$C_p = \frac{SSE}{SST} - (n - 2k) + 1 . \tag{3.6}$$

As all the variables of equation 3.6 are known, the calculation for a given subset of independent variables can be easily computed. According to Mallows, when C_p is higher than (k+1), evidence exists of bias due to an incompletely specified model; on the other hand, when C_p reaches values lower than (k+1), the model is considered overspecified, containing too many variables (Freund, 2003:129-131).

The application of these discussed techniques will help us to find an appropriate subset of explanatory variables for our problem. From them and after testing the required assumptions of the model described earlier, we can calculate the regression coefficients and determine the regression equation to predict values of our dependent variable. Several software packages have been developed to be used as a platform to compute the statistics required to follow the regression process; JMP® is one of them and has been chosen to perform our analysis.

At this point, it is important to highlight that there is not an exclusive way of searching for a good subset of independent variables to introduce in the regression model; subjective elements like analyst judgment can play an important role into the exploratory process. This means that no automatic procedure will always come across with the "best" model and judgment should play a key role in model building especially for explanatory studies (Kutner et al. 2005:368).

Finally, the amount of available data is also an important element to consider (Mc Clave et al. 2005:789). The number of independent variables to introduce in the model is strongly influenced by data availability, and it has to make sense; it is difficult to imagine a model constructed from 20 or 30 data points that contains 10 or 15 independent variables; these should be a balance between the amount of data and the number of independent variables introduced in the regression model. It is generally accepted that a ratio greater than 6:1 (6 data points for each independent variable present in the model), but if possible greater than 10:1, is preferred for any model building method (Kutner, et al. 2005:372).

The final step in all model-building process is the validation of the model. To be useful, the selected regression model should be validated against reality. Several methods have been developed to perform the validation process of a constructed model; the following section will help us to understand the validation process that will be implemented to validate our regression model.

The Validation Process of the Model

Validation can be defined as a process in which the model and its behavior are compared to the real system and its behavior (Banks et al. 2004:361-365). The objective of the process is a judgment regarding how well suited a particular model is for a specific application. (Hughes and Rolek, 2003:977). Actually, models are merely limited representations of complex reality and for that reason they cannot be totally validated, but the quality of a model depends on how well those that develop the model understand the reality it supposes to represent.

Naylor and Finger (1967) have formulated a three-step validation approach (Banks et al, 2004: 362):

1. Build a model that has high face validity.
2. Validate model assumptions.
3. Compare the model input-output transformations to corresponding input-output transformations for the real system.

Face validity is defined as the extent to which an instrument looks like it is measuring a particular characteristic. Through this measure we look for constructing a reasonable model for users and other people who know how the real system works and understand how it is being simulated. The use of expert opinions and the experiences of users and modelers are very useful to construct face validity (Banks et al. 2004: 362).

The validation of the model assumptions can be classified as structural and data assumptions (Banks et al. 2004: 362). The first ones are related with the simplifications and abstraction inside the methodology used to build the model. In our case it includes the model assumptions presented previously such as continuity, normality and independence between observations. On the other hand, data assumptions involves testing for data reliability, and also testing that the particular environmental conditions used to perform the analysis of the data will be present to allow model's users to extrapolate future values from the original data.

The final validation test, and perhaps the most objective one, is related to the model's ability to predict future values. Kutner describes three basic ways of validating the regression model (Kutner, et al. 2005:369-375):

1. Checking the model's ability to predict values against new data.
2. Compare the result of the model with theoretical expectations, empirical results or simulation.
3. Reserve part of the original data set to be used only in the validation process.

As it will be described in the next chapter, actual data have been chosen to perform the regression analysis; this limits our ability to gather new data to be used to test our model. In addition to that, the lack of a pre-existing methodology to predict jet fuel prices in Argentina makes it difficult to introduce theoretical or empirical evidence to determine whether the chosen model is reasonable. As a result the third way described by Kutner has been selected for our case. Implementing this implies that the modeler normally reserves part of the acquired, historical data for the validation process only. For these data points, values are predicted and confidence intervals calculated with the regression model and then these values are compared to determine how well the model simulated the behavior of the real data.

Also two forecasting error measures and the Theil's U-statistic will be used to evaluate the performance of the model. The two forecasting errors to be used are the

Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE). These measures can be mathematically expressed as follow (Makridakis et al. 1998:42-45):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | e_i |$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} | PE_i |,$$

where $PE_i = \left(\frac{Y_i - \hat{Y}_i}{Y_i} \right) \times 100$, and n represents the number of data points used in the error calculations. The MAE error has the advantage of being more interpretable and easy to explain to non-specialists because it represents the average of the absolute error of the forecast. On the other hand, the MAPE measure is the average percentage of the absolute error of the forecast and it is considered an important measure especially when we want to compare different forecasting models (Makridakis et al. 1998:43-45).

The Theil's U statistic allows a relative comparison of our model with the naïve approach. Mathematically this statistic is defined as $U = \sqrt{\frac{\sum_{i=1}^{n-1}\left(\frac{\hat{Y}_{i+1} - Y_{i+1}}{Y_i} \right)^2}{\sum_{i=1}^{n-1}\left(\frac{Y_{i+1} - Y_i}{Y_i} \right)^2}}$; it can be observed that the numerator represents the sum of the squares of the relation between the error of our forecast model and the previous data point, while the denominator represents the sum of the squared of the relation between the errors of the naïve forecast most commonly used, which is considering our current data point as the forecast of the next period and the previous data point (Makridakis et al. 1998:48-49). U-values greater than 1 indicates that naïve forecast error are lower than our forecast model, so naïve forecast is preferred, while U-values lower than 1 denotes that our forecast model is better than the naïve forecast (Makridakis et al. 1998:50).

Model-building and model-validation processes are important aspects to be considered in multiple regression analysis. To be implemented successfully, the described processes require a large amount of accurate data. The next chapter describes the data used to implement the described methodology to build the regression model to predict Argentine jet fuel prices, as well as, the results obtained by the regression analysis and their implications on the AAF environment.

4. Data Analysis

Introduction

Earlier chapters have helped us to define the problem to forecast Argentine jet fuel prices, to identify potential predictors that affect this asset inside the Argentine environment, and to describe the selected methodology to build the model to make inferences of the price of jet fuel in Argentina. In this chapter, we first describe the data required and how they can be obtained, then based on those data we illustrate how the described model-building process applies to obtain a model that would allow us to predict Argentine jet fuel prices, and finally we relay the validation process to determine the usefulness of our model.

The Data

Statistics is the science of data; no-statistical analysis is possible without the existence of data over which to perform the analysis. For its own characteristics our study can be identified as an exploratory observational study. In this, analysts look for explanatory variables that could be related to the response variable (Kutner et al. 2005:345-346); the main characteristic of this study is that the investigator examines the experimental unit in their natural setting and records the variables of interest (McClave et al. 2005:19).

Looking for the appropriate set of data that can be used to build any statistical model is not an easy matter. Investigators are often forced to search explanatory variables that might plausibly be associated in any form with the response variable under study. In our case, beside Argentine jet fuel prices (measured in Argentine Peso per liter), based on the literature review (Chapter 2), the following list of domestic and international factors have been selected as potential predictors of Argentine jet fuel prices to perform the statistical analysis:

1. International factors:
    a. West Texas Intermediate (WTI), a type of crude oil used as a benchmark in oil pricing, measured in US Dollars per barrel.
    b. JetKero 54 index (JK 54), the price of jet fuel in the Gulf of Mexico, measured in cents of dollar per gallon.
2. Domestic factors:
    a. The value of the Argentine Peso in relation to the US Dollar (VPD), measured in peso per dollar.
    b. Argentine Industrial Growth (IG) as percentage of the previous month.
    c. Consumption Inflation Rate (IR) as percentage of previous month.
    d. Price Index of Argentine-Produced Wholesale Goods (IPP).
    e. Internal Wholesale Price Index (IPIM).
    f. Price Index of Argentine-Produced Wholesale Goods (natural gas and oil) (IPP O&G).
    g. Argentine Total Jet Fuel Production (TJFP) measured in cubic meters.
    h. Argentine Jet Fuel Demand (TJFD), measured in cubic meters.
    i. Relation between the Argentine jet fuel demand and Argentine jet fuel production (RDP).

These factors tend to consider the international influence of oil market over the Argentine oil market, the own characteristic of the market of jet fuel in Argentine, and how the market is influenced by economic indicators.

Monthly data from March, 2002 to September, 2006 involving Argentine jet fuel prices, as well as, data from the same period concerning the described domestic and international factors have been collected from different sources. The Argentine Secretary of Energy and the Argentine Institute of Statistic and Census (official statements of the Argentine government) have been used as source to collect the data for the domestic factors; while the Platts, Co. has been chosen for the data involving the considered international factors.

The data was selected from March, 2002 to avoid possible distortions in prices produced during the financial crisis that affected Argentina in 2001-2002. Although the analysis of this crisis is beyond the scope of this thesis, it is important to highlight that this crisis was one of the most difficult situations that affected the country. This crisis had politic, economic and social implications. Five presidents governed in a two month period. The default of the public debt (which reached values close to 150 billon dollars) had international implications (inability to access to international credit and lost of international credibility) as well as internal implications (instability and fiscal insolvency). Other ramifications included: devaluation of the Argentine currency with respect to U.S. Dollar, unbalance pesification of deposit which affected the whole financial system, and the consequent lost of people purchasing power (Cortés, 2003).

Model to Predict Argentine Jet Fuel Prices

From the whole set of data (55 observations), 43 data points, selected by random, have been used to build the model and test the model assumptions during the validation process; the entire set of data have been used to calculate the forecast error measures and confidence intervals. Although the selection by random can be easily questioned when working in forecasting, it seem to be more appropriated to simulate the behavior of a response variable when the conditions of the market of reference is subject to little instability which is the case of Argentine after the 2001-2002 crisis.

The analysis was performed lagging one month all the considered explanatory variables, including the Argentine jet fuel price also considered as a possible predictor; a fact that has practical and logical implications. The first is that a month is the typical delay to obtain the information; normally all the domestic factors can be easily obtained during the first days of the next month in relation to the monthly information required. Also, as it was described in Chapter 2, Argentine is a price taker with respect to the oil market, so selecting lagging international reference prices of oil and its derivatives seem to be more adequate.

The building-model process can be divided in two steps: reducing the number of predictors and building the regression model. To implement the first step to reduce the number of explanatory variables, a multivariate scatterplot matrix can be obtained using JMP® (Figure 4-1). Table A.1 in Appendix A shows us the corresponding correlation coefficients. As it can be observed, jet fuel prices are highly correlated with previous value of jet fuel prices in Argentina (JFP(L1)), the selected international factors (WTI and JetKero 54) and the Argentine indexes of inflation for wholesales: IPIM, IPP and IPP

(O&G). Small or no correlation can be detected between jet fuel price and jet fuel production, jet fuel demand, demand / production relationship, value of Argentine peso in relation to U.S. dollar, consumption inflation rate, and industrial growth. Also strong correlations can also be observed between WTI and JK 54 index, and between the three selected wholesale inflation indexes. These facts suggest that only one of the international and domestic factors should be introduced in the model to reduce possible multicollinearity issues. A closer look of a new multivariate scatterplot matrix reduced to the explanatory variables that show correlation with our response variable can help us to extract other conclusions (Figure 4.2).

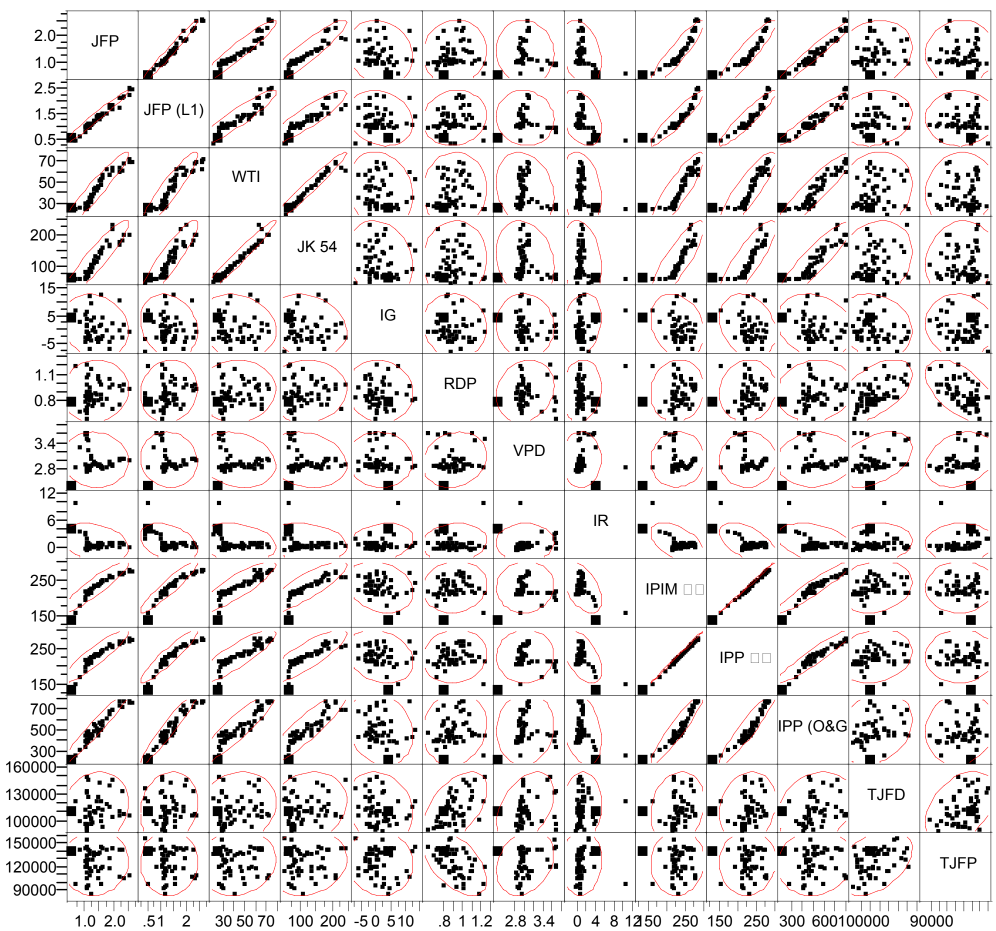

Figure 4.1: Multivariate Scatterplot Matrix

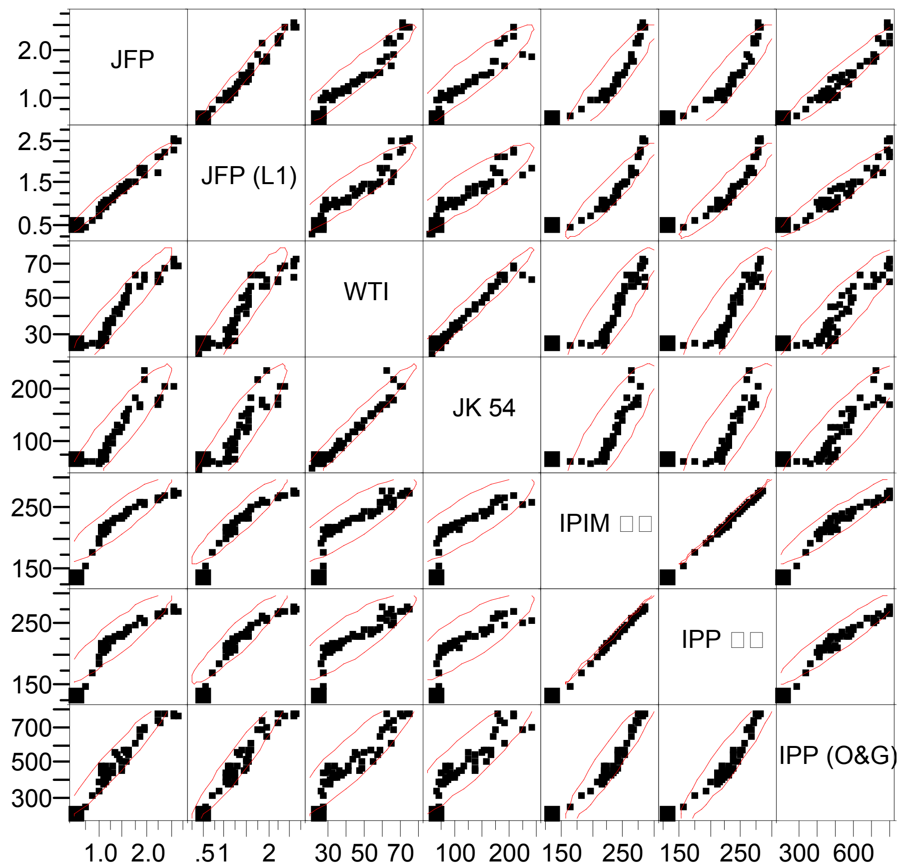

Figure 4.2: Reduced Multivariate Scatterplot Matrix

Figure 4.2 shows us two additional, important facts that should be considered before selecting the explanatory variables to introduce in the model to predict Argentine jet fuel prices. The first fact involves a discrete event that affects the values of the variables from February 2006 to September 2006. Although this discrete event cannot be easily attributable to a specific fact, it can be simulated by the use of a dummy variable to be introduced in the model. This dummy variable takes values of one for data points from February 2006 and zero otherwise. The second fact is related to the presence of non-linear relation between the Argentine jet fuel prices and IPIM and IPP indexes. If the use of these explanatory variables cannot be avoided then the necessity of introducing transformations should be considered. But to avoid multicollinearity issues we have to choose only one of the wholesale inflation indexes. Because of comparable association and that the relation between jet fuel price and IPP (O&G) seem to be more linear, we can avoid transformations selecting this explanatory variable.

As a result of the preceding analysis the selected variables to introduce in the model are: Argentine jet fuel prices (Argentine $/liter) lagged one month, WTI (US$ per barrel) lagged one month, IPP (O&G) index lagged one month, and the described dummy variable. A final Multivariate Scatterplot Matrix for the selected variables is shown in Figure 4.3 to show how the created dummy variable works. The model parameters calculated using JMP® are shown in Tables 4.1: Summary of Fit, Table 4.2: Analysis of Variance, Table 4.3: Parameter Estimates.

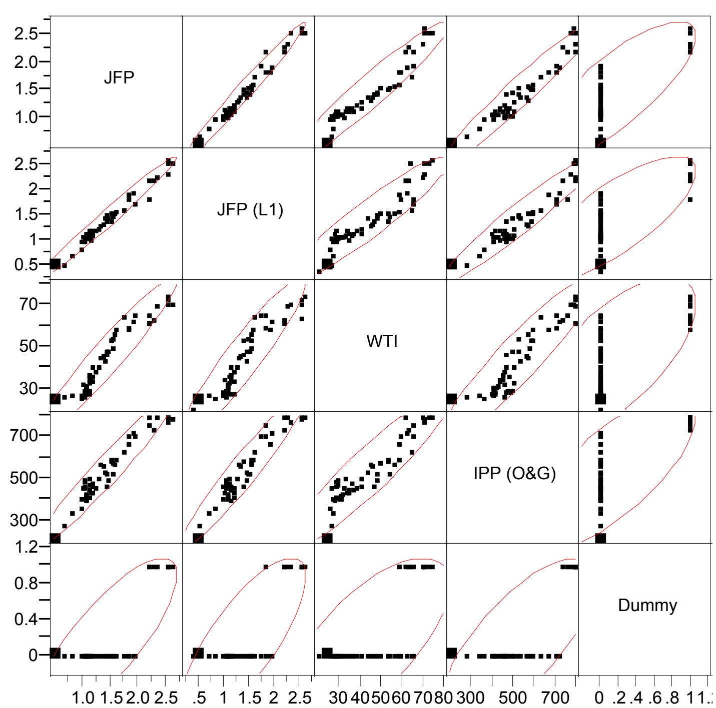

Figure 4.3: Multivariate Scatterplot Matrix of selected explanatory variables

Table 4.1: Model Summary of Fit

RSquare	0.993514
RSquare Adj	0.992832
Root Mean Square Error	0.04413
Mean of Response	1.445581
Observations	43

Table 4.2: Model Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	11.336257	2.83406	1455.256
Error	38	0.074004	0.00195	Prob > F
C. Total	42	11.410260		<.0001

Table 4.3: Model Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.0337725	0.034891	0.97	0.3392
JFP (L1)	0.424805	0.057151	7.43	<.0001
WTI	0.0101046	0.001336	7.56	<.0001
IPP (O&G)	0.0006242	0.000167	3.75	0.0006
Dummy	0.1995058	0.029592	6.74	<.0001

Analyzing Table 4.1, we conclude that the model presents a high Adjusted-R^2 (0.9928), which implies a good overall fitting of our regression model. Table 4.2 shows us that the F-test for overall significance indicates significance of our regression model (p-value is lower than our state level of significance assumed to be 0.05). Finally Table 4.3 also shows us that the selected explanatory variables are also significant (the individual p-values of our explanatory variables are all lower than our level of significance). Summarizing, the multiple regression model to predict Argentine jet fuel prices (after some rounding for display purposes only) can be mathematically expressed as:

$$\hat{y} = 0.034 + 0.425 \times JFP(L1) + 0.01 \times WTI + 0.00062 \times IPP(O\,\&\,G) + 0.1995 \times Dummy$$

where $\hat{y}$ represents our prediction of Argentine jet fuel price in Argentine pesos per liter, JFP(L1) is the Argentine jet fuel price of the previous month in Argentine pesos per liter, WTI is the West Texas Intermediate in US Dollars per Barrel lagged one month, IPP (O&G) is the Price Index of Argentine-Produced Wholesale Goods for natural gas and oil also lagged one month, and the dummy variable takes the value of 1 for calculations from February 2006 and zero otherwise.

The Model Validation Process

As it was described in previous chapters, the validation process is the process by which the model and its behavior are compared to the real system and its behavior. This process implies to demonstrate that the model has high face validity, meets the model assumptions, and is capable of providing similar outputs compared to the real system when they are subject to similar inputs.

Demonstrating that the model has high face validity is perhaps the most difficult part of the analysis because it is commonly based on a subjective point of view of the builder of the model. The use of adequate techniques to select the appropriate explanatory variables to predict the response variable and expert opinions are common elements used to help an analyst to achieve confidence that the model is an instrument that measures what it is supposed to measure. In our case, we have used expert opinions to select our initial list of explanatory variables; also we have obtained data of these variables from recognized (domestic and international sources) sources; and finally we have applied a rational, statistical process to reduce the number of explanatory variables and to obtain our multiple regression model. All these facts allow us to conclude that the reached model presents face validity.

Assuming that our model has high face validity, our next step is to test our model assumptions: normality, independence and constant variance of residuals, linearity of the β's coefficients, outliers and influential data points, and multicollinearity issues; assumptions that have been described in detail in Chapter 3.

1. **Testing Normality on Residuals**: As we know the model residuals are suppose to be normally distributed with mean equal to zero and variance equal to 1. We have used the Shapiro-Wilks test provided by JMP® and a histogram of the random errors to test that the distribution of our residuals is normal. Figure 4.4 shows us the histogram of the residuals compared to a normal distribution; we can observe that the distribution of our residuals looks normal. This is corroborated through the Shapiro-Wilks test (Table 4.4). As it can be seen the p-value of this test (0.8716) is higher than our level of significance (0.05); this fact allows us to conclude that there is no statistical evidence to reject the hypothesis that our residuals are normally distributed.

Table 4.4: Shapiro-Wilks Test (Goodness of Fit test)

W	Prob<W
0.983495	0.8716

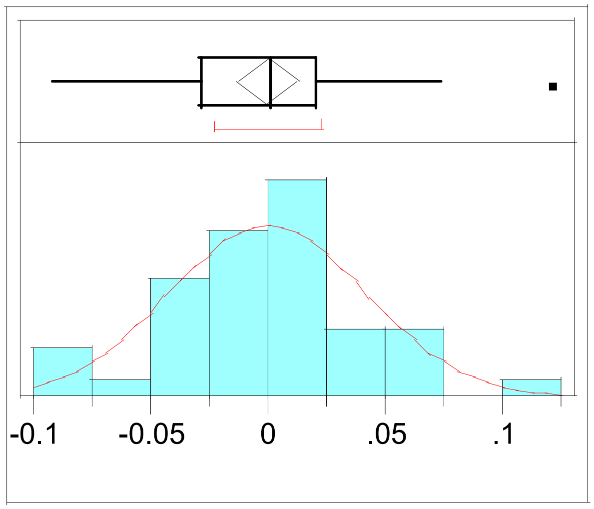

Figure 4.4: Distribution of Residual Jet Fuel Prices

2. **Testing Independence on Residuals**: OLS also assumes that random errors are independent in the probabilistic point of view; neither correlation nor association of the residual exists. The random selection of data points for the validation process deprives us of the capability of using Durbin-Watson or runs test to test for this assumption on the data points used to build the model. The salomonic solution to that is the visual analysis of the scatterplot of the residuals to test for the presence of any trend, pattern, or abnormality, and performing Durbin-Watson test over the entire set of data points which are definitely equally spaces in time. Figure 4.5 shows us the scatterplot of the residuals; no pattern, trend or abnormality can be easily observed through this plot. On the other hand the results of the Durbin-Watson test over the entire set of data can be observed in Table 4.5. The p-value (0.3632) for this test is higher than our

level of significance (0.05), so we can conclude that our residuals are independent over time.

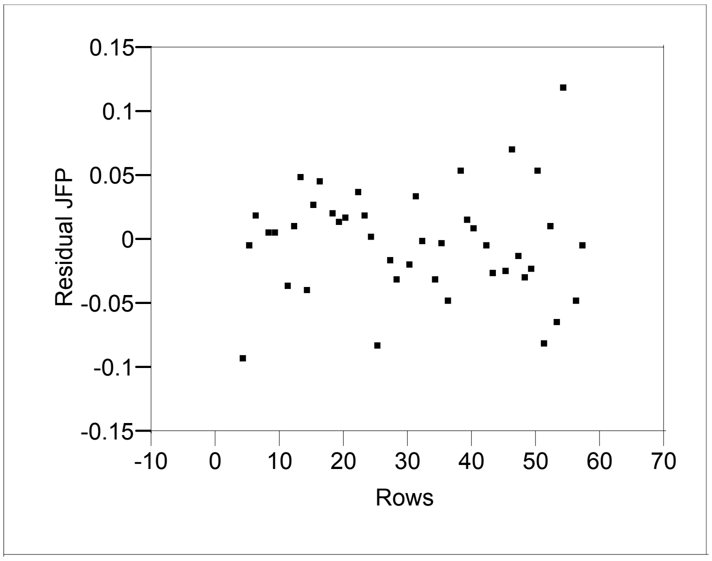

Figure 4.5: Run Plot of Residuals

Table 4.5 Durbin-Watson Test

Durbin-Watson	Number of Obs.	Autocorrelation	P-value
2.0315657	55	-0.0674	0.3632

3. **Testing Constant Variance on Residuals**: Also we know that another OLS assumption requires the residuals to display constant variance; a descriptive plot (Figure 4.6) and the Breusch-Pagan test (Table 4.6), described in Chapter 3, can be used to test this assumption. As it can be observed in Table 4.6, our p-value is equal to 0.02706, lower than our level of significance (0.05), fact that allows us to conclude that constant variance assumption could be an issue. A close analysis of Figure 4.6 shows us that the data points that correspond to April-02 and June-06

could be the problem; if we perform the analysis again excluding these data points we find that the new p-value for the Breush-Pagan test becomes 0.1688, higher that our chosen level of significance (Table 4.7). As a conclusion we can assume that if these data points are not influential, analysis that we are going to perform later on in this chapter, we can keep them in the model, to make a robust model, and assume that constant variance is met. On the other hand if these data points are influential we should remove them of the model and also consider that our assumption that the residuals display constant variance is met.

Table 4.6: Breusch-Pagan Test

n	dfmodel	SSE	SSR	t-statistic	p-value
42	4	0.074004	0.00006803	10.9561676	0.027061072

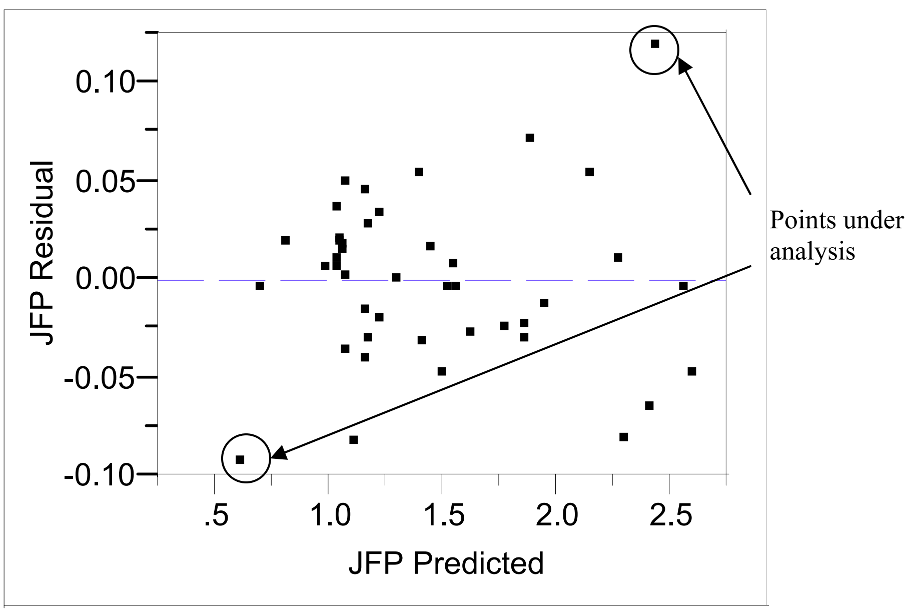

Figure 4.6: Residuals' Scatterplot

Table 4.7: Breush-Pagan Test 2

n	dfmodel	SSE	SSR	t-statistic	p-value
40	4	0.0452725	0.00001649	6.43638017	0.168843347

4. **Testing Linearity on β's coefficients**: This assumption implies that the β's coefficients, which are the slope of the line that model the behavior of each predictor with the response variable, are constant over time. As there is no a real way to test that a non-linear model would be better, we can only take a look to the multivariate scatterplot matrix (Figure 4.3). As we can observe the relation between response and each selected explanatory variable looks linear, which allow us to assume that this assumption is also met.

5. **Testing for the Existence of Outliers and Influential Data Points**: As we have said in previous chapters the existence of outliers and influential data points can bias our regression model. To test for these we use the Cook's Distance overlay plot (Figure 4.7). According to the figure, the point corresponding to April 2002 seems to be an influential data point; the figure does not allow us to conclude in the same form when we look at the data point corresponding to June 2006. Although eliminating the April 2002 data point could be considered appropriate because it is close to the period of crisis that affected Argentine's economy, a deeper analysis of the

model parameters and p-values for the overall model and for the independent explanatory variable without this data point is required.

As it can be observed in Table 4.8: Summary of Fit, Table 4.9:Analysis of Variance, and Table 4.10: Parameters estimates, the adjusted R^2 of the new model, the F-test for overall significance and the p-values for each independent explanatory variables do not show us changes when we exclude this point from the analysis. These facts suggest that keeping these data points is appropriate and would allow us to increase the robustness of our model.

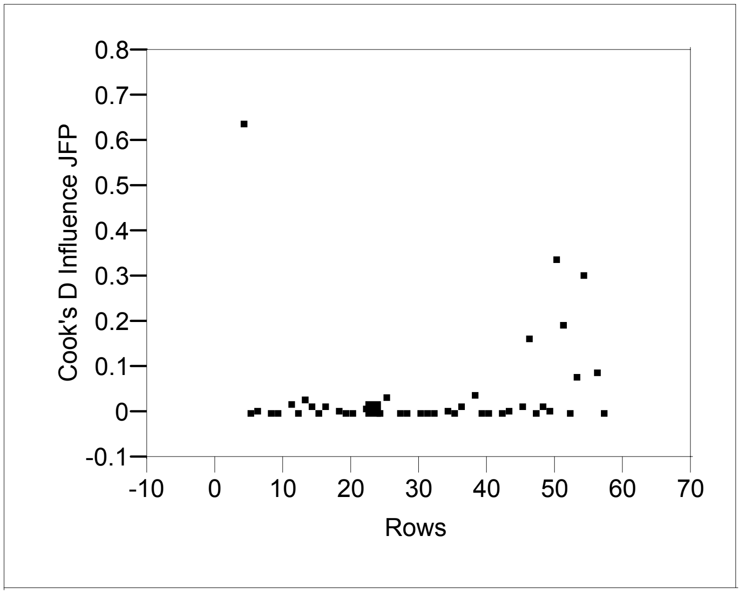

Figure 4.7: Cook's Distance Overlay Plot

Table 4.8: Model 2 Summary of Fit

RSquare	0.99416
RSquare Adj	0.993529
Root Mean Square Error	0.040773
Mean of Response	1.467619
Observations (or Sum Wgts)	42

Table 4.9: Model 2 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	10.471653	2.61791	1574.765
Error	37	0.061509	0.00166	Prob > F
C. Total	41	10.533162		<.0001

Table 4.10: Model 2 Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.0893476	0.03808	2.35	0.0244
JFP lagged 1 month	0.405903	0.053252	7.62	<.0001
WTI	0.0115271	0.001339	8.61	<.0001
IPP (O&G)	0.0004489	0.000167	2.69	0.0106
Dummy	0.2267128	0.029086	7.79	<.0001

6. **Testing for Multicollinearity issues**: Possible multicollinearity is present when explanatory variables are correlated among themselves and with other variables related to the response variable not included in the model. This test is very important because the presence of multicollinearity directly affects the calculation of the β's coefficients. It was highlighted in Chapter 3 that the Variance Inflation Factor is normally used to test this issue. JMP® provides the VIF scores which are shown in Table 4.11: Parameter Estimates and VIF scores.

   We can observe that some of the VIF scores are higher than 10, which can alert us about some multicollinearity issues. The presence of this problem should be expected due to the fact that we are trying to predict prices of jet fuel which is a derivative of a commodity (oil). Additionally it is important to consider that high VIF scores are frequently tied to high p-values which is not our case. Although the VIF scores show that predictors overlap each other, the individual contributions of each predictor to the predictive variable are high, as we can observe in the p-values of Table 4.11. This fact allows us to conclude that keeping the selected explanatory variables into the model is appropriate.

Table 4.11: Model Parameter Estimates and VIF scores

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	0.0337725	0.034891	0.97	0.3392	
JFP lagged 1 month	0.424805	0.057151	7.43	<.0001	18.413055
WTI	0.0101046	0.001336	7.56	<.0001	9.621175
IPP (O&G)	0.0006242	0.000167	3.75	0.0006	13.435453
Dummy	0.1995058	0.029592	6.74	<.0001	2.63525

Having analyzed how our model respond to the theoretical model assumptions, we now need to know if the model behaves as the real system behaves. Figure 4.8 illustrates the real price of jet fuel during the analyzed period of time and the result obtained applying the constructed model to predict Argentine jet fuel prices. The apparently good response showed by the model in Figure 4.8 has to be corroborated with the use of statistical measures that allow us to quantify the level of response of our model compared to the real system behavior. As we have defined in Chapter 3, three different measures have been used to determine the accuracy of our model: Forecasting error measures (Mean Absolute Error and Mean Absolute Percentage Error), Theil's U statistic, and percentage of prediction that fall inside the confidence interval of the real output. We look for low forecasting errors (lower than 5%), values of Theil's U lower than one, and more than 95 % of predictions to fall inside the respective confidence intervals.

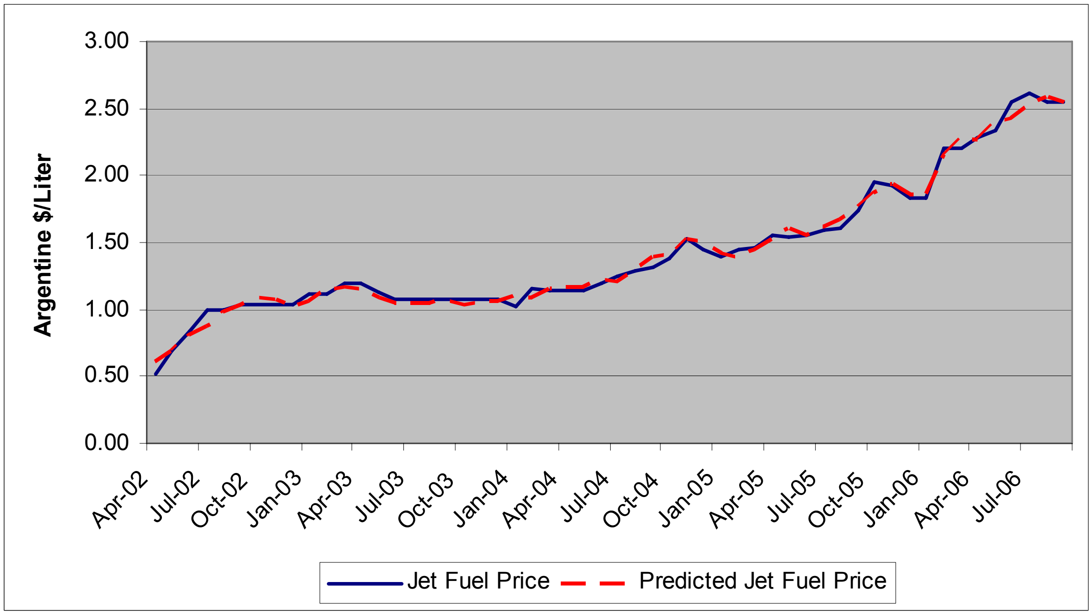

Figure 4.8: Real and Predicted Jet Fuel Prices Comparison

Table B.1 of Appendix B shows us the entire calculation of these measures which are summarized in Table 4.12. It can be observe that the average absolute error is lower than 4 cents of Argentine pesos, while the average absolute percentage error is lower than 3 %. These measures indicate that our model shows a good behavior compared to the real system. The same table shows us that the Theil's U statistic (0.5564) is lower than 1; this fact implies that the selected regression model provides better outcomes than the naïve approach of considering the last jet fuel price as the price of the following period. Finally, the table also illustrates that 100% of the model predictions fall inside the confidence intervals of the real system.

Table 4.12: Summary of Model Behavior Results.

MAE	MAPE	Theil's U	% Inside CI
0.0372	2.86%	0.5564	100%

Having built and validated the model, the following chapter presents the conclusion of this thesis work, the limitation of the developed model, and suggestions of future development that would improve this work.

5. Conclusion

This chapter presents the answer to the investigative and research questions proposed in Chapter 1, the limitation of the developed model to predict Argentine jet fuel prices, an exploration of possible areas of further research, and the research summary.

Addressing the Research Questions

During this section, we first review the investigative questions traced in Chapter 1 to address the research question in conjunction with the result of our thesis. Finally, as a summary of the work we address the research question itself.

Can jet fuel prices be adequately predicted using multiple regression models?

Yes, multiple regression models have shown to be effective to predict the prices of oil and its derivatives in the United States market. Although other methods such as econometric forecasting and neural networks have normally shown better results, their complexity have been an impediment to select one of these models. Introducing a new methodology in a complex environment such as the AAF requires a balance between complexity and accuracy. Multiple regression analysis provides a good trade off between these two aspects permitting us to obtain a model easy to understand, practical and useful.

What are the necessary variables to introduce in the model to predict jet fuel price in Argentina?

The successful application of any methodology strongly depends on the particular conditions of the market where it is applied. It makes no sense to believe that a model that has proved to be useful predicting jet fuel prices in the U.S. market or any other market in the world can be directly applied to the Argentine market. Accordingly twelve variables including international and domestic factors that may affect Argentine jet fuel prices have been analyzed to select the best predictors. Using a stepwise process the original number of potential predictors were reduced to six, and further analysis allowed us to choose four significant predictors of Argentine jet fuel prices: the price of Argentine jet fuel lagged 1 month, the West Texas Intermediate Index (WTI), the Price Index of Argentine-Produced Wholesale Goods (natural gas and oil) (IPP O&G), and a dummy variable which takes values of one from February 2006 and zero otherwise. The adjusted R^2 of the resulting model is high (approximately 0.99) showing an excellent goodness of fit to the real data in the analyzed period of time.

What are the necessary data to solve the problem? Are they available?

Any statistical approach requires the analysis of a considerable amount of data. In our case, monthly data of the selected variables (international and domestic factors) from March 2002 to September 2006 have been used to build the multiple regression model. To minimize the possible negative effect of the considered assumption of complete and accurate data, we have used data from worldwide providers of oil market information (Platts, Co.), and two different Argentine governmental organizations: the Argentine Secretary of Energy, and the Argentine Institute of Statistics and Census.

Another important factor considered during this thesis has been the repeatable characteristics of any thesis work; this factor strongly depends on data availability. All the data used in this paper are available on the internet; any person should be capable of obtaining the same results using the analyzed methodology, which assures the thesis repeatability.

Would a multiple regression model provide a useful planning and decision aid for the Argentine Air Force?

To be useful, a model should be validated against model assumptions and real system behavior. A model validation process including the validation of the theoretical model assumption and the comparison between model results and real system behavior have been included in this thesis work. In relation to the comparison of the model with the real system behavior, the obtained high adjusted R^2 (0.99) shows us an excellent goodness of fit of the model. A reduced average absolute error (2.98%) of the model has also corroborated this fact. Finally, the resulting Theil's U statistic (0.55) lower than 1 allows us to conclude that the model presented in this thesis is better than using the classical naïve approach to forecast Argentine jet fuel prices. All these calculations have proved that the model could provide a useful planning and decision aid for the Argentine Air Force.

How can the Argentine Air Force better predict jet fuel prices to improve financial and logistic planning?

This thesis has proved that accurate predictions of Argentine jet fuel prices are essential to improve financial and logistic planning. It has also demonstrated that predicting Argentine jet fuel price is neither easy nor impossible. The application of a logical methodology to the correct data is the key to achieve success. A systematic application of statistical principles has allowed us to build a multiple regression model to predict the price of jet fuel considering the particular conditions of the Argentine market.

The usefulness of any model is always based on a trade off between model accuracy and model complexity. The presented model has proved to be accurate

(Average Absolute Error lower than 3%) and better than the naïve approach (Theil's U statistic lower than 1), methodology normally used when no model exists to predict jet fuel prices in the AAF. Added to that, model complexity has been reduced trough the use of only four variables which are easily available in normally consulted URL addresses such us the Argentine Secretary of Energy and the Institute of Statistics and Census; this fact increases the usefulness of the model, and at the same time facilitates its introduction in the AAF environment. As a result, the application of the presented model would help the AAF to increase forecast accuracy of jet fuel prices facilitating budget process and logistic planning.

The model developed to predict Argentine jet fuel prices can mathematically be written (after some rounding for display purposes only) as:

$$\hat{y} = 0.034 + 0.425 \times JFP(L1) + 0.01 \times WTI + 0.00062 \times IPP(O\,\&\,G) + 0.1995 \times Dummy \ ,$$

where $\hat{y}$ represent our prediction of Argentine jet fuel price, JFP (L1) is the price of jet fuel in Argentina in the previous month in Argentine Pesos per Liter, WTI is the West Texas Intermediate in US Dollars per Barrel lagged one month, IPP (O&G) is the Price Index of Argentine-Produced Wholesale Goods for natural gas and oil also lagged one month, and the dummy variable takes the value of 1 for calculations from February 2006 and zero otherwise.

As any model, our model is not perfect; it is only a simplification of the real world and presents some limitations. For our case, the most important limitation is related with the assumption that current Argentine economic conditions will continue in the future. Having considered data from 2002 to 2006, a period in which economic indicators of the country has grown; the model is limited to forecast Argentine jet fuel prices where the present conditions continue.

Areas of Further Research

This thesis has presented a systematic, statistical approach to forecast Argentine jet fuel prices. Although this approach has proved to be effective, the develop process can be considered static because past data have been used to predict future jet fuel prices. A new interesting point of view could present a more dynamic approach. This further area of research should address the problem month by month, regression coefficients could be recalculated each month when new data is available, new variables could be analyzed and introduced in the model if they show to be significance. This more dynamic approach should allow the AAF to obtain a more responsive forecast of jet fuel prices generating a process of continue improvement for budgeting and logistic planning.

It is important to remember that Argentine oil companies have taken advantage of favorable conditions in the international market to increase oil exportations. Chapter 2 of this thesis has shown us that Argentine exportations of crude oil and its by-products have grown in the last decade; if this trend continues, it would be interesting to analyze the relationship between the level of exportation and the number of wells drilled in Argentina, how exportations and level of production could impact domestic prices of crude oil and its derivative products, and how government intervention could influence these prices. Finally, it would be also attractive to study the relation that exists between OPEC production and non-OPEC production and how this relation affects oil prices not only in the international market, but also in the particular characteristics of the Argentine market.

Research Summary

Jet fuel is considered an important asset to accomplish the Argentine Air Force (AAF) missions; it is also the element with highest demand and the most expensive item supported by the AAF. Crude oil price instability, the main component on the production of jet fuel, added to high consumption rates and other unique factors of the Argentine market have caused problems that have directly affected budget process and logistics planning. The situation has created a real challenge for military personnel working in the acquisition of jet fuel for the Argentine Air Force. For years, they have tried to predict the price of this asset to improve financial and logistic planning, but the great number of variables that affect the problem and the lack of an adequate methodology have been the biggest impediments to achieve an acceptable solution.

This thesis shows us that no magical way exists to find solution of complex problems; using a rational line of attack, it shows that Argentine jet fuel prices can be accurately predicted through the use of multiple regression analysis. This logic process has included the problem definition (Chapter 1), a literature review (Chapter 2), the description of the methodology to be applied (Chapter 3), and the application of this methodology to real data to build the model to predict Argentine jet fuel prices (Chapter 4). The model provided in this thesis can help the AAF to considerably improve the forecast accuracy of Argentine jet fuel prices; forecast that could become in an important tool to introduce improvements in its budget and logistic process.

Correlation Coefficients of Potential Predictors

Table A.1: Multivariate Correlations of Potential Predictors

	JFP	JFP (L1)	WTI	JK 54	IG	RDP	VPD	IR	IPIM	IPP	IPP (O&G)	TJFD	TJFP	Dummy
JFP	1.0000	0.9819	0.9514	0.9260	-0.1502	0.0509	-0.0692	-0.2974	0.9237	0.9183	0.9676	0.1878	0.1129	0.6899
JFP (L1)	0.9819	1.0000	0.9208	0.9004	-0.1782	0.0856	-0.0766	-0.3747	0.9373	0.9345	0.9566	0.1873	0.0639	0.6422
WTI	0.9514	0.9208	1.0000	0.9781	-0.1450	0.0235	-0.2199	-0.2029	0.8621	0.8555	0.9076	0.1847	0.1484	0.5347
JK 54	0.9260	0.9004	0.9781	1.0000	-0.1724	0.0336	-0.2033	-0.1933	0.8376	0.8313	0.8970	0.2288	0.1787	0.4622
IG	-0.1502	-0.1782	-0.1450	-0.1724	1.0000	-0.0704	-0.0459	0.2562	-0.2248	-0.2307	-0.1923	-0.1881	-0.0734	0.0911
RDP	0.0509	0.0856	0.0235	0.0336	-0.0704	1.0000	0.1614	0.1117	0.0816	0.0840	0.0944	0.5420	-0.5459	0.0232
VPD	-0.0692	-0.0766	-0.2199	-0.2033	-0.0459	0.1614	1.0000	0.0092	0.0399	0.0277	0.0734	0.4328	0.2717	0.0449
IR	-0.2974	-0.3747	-0.2029	-0.1933	0.2562	0.1117	0.0092	1.0000	-0.5241	-0.5402	-0.3511	0.0909	0.0448	-0.0321
IPIM	0.9237	0.9373	0.8621	0.8376	-0.2248	0.0816	0.0399	-0.5241	1.0000	0.9990	0.9396	0.1952	0.0743	0.4925
IPP	0.9183	0.9345	0.8555	0.8313	-0.2307	0.0840	0.0277	-0.5402	0.9990	1.0000	0.9364	0.1940	0.0692	0.4877
IPP (O&G)	0.9676	0.9566	0.9076	0.8970	-0.1923	0.0944	0.0734	-0.3511	0.9396	0.9364	1.0000	0.2703	0.1435	0.6337
TJFD	0.1878	0.1873	0.1847	0.2288	-0.1881	0.5420	0.4328	0.0909	0.1952	0.1940	0.2703	1.0000	0.4002	0.0266
TJFP	0.1129	0.0639	0.1484	0.1787	-0.0734	-0.5459	0.2717	0.0448	0.0743	0.0692	0.1435	0.4002	1.0000	0.0033
Dummy	0.6899	0.6422	0.5347	0.4622	0.0911	0.0232	0.0449	-0.0321	0.4925	0.4877	0.6337	0.0266	0.0033	1.0000

Appendix B

Forecasting Error Measures

Table B.1: Error Measure Calculation

		Multiple regression model											Confidence Interval		
	Jet Fuel Price	Jet Fuel Price lagged 1 month	WTI (U$S/Barrel) lagged 1 month	IPP (Oil and gas) lagged 1 month	Dummy	Predicted Jet Fuel Price	e	|e|	PE	APE	Theil's U		Upper bound	Lower-bound	Prediction included?
Mar-02	0.48														
Apr-02	0.52	0.4800	24.42	203.20	0	0.61	-0.0913	0.0913	-17.55%	17.55%	0.0362	0.0069	0.75	0.48	y
May-02	0.69	0.5200	26.27	275.75	0	0.69	-0.0022	0.0022	-0.32%	0.32%	0.0000	0.1069	0.83	0.56	y
Jun-02	0.83	0.6900	27.02	335.29	0	0.81	0.0208	0.0208	2.51%	2.51%	0.0009	0.0412	0.95	0.67	y
Jul-02	0.99	0.8300	25.52	361.11	0	0.87	0.1204	0.1204	12.16%	12.16%	0.0210	0.0372	1.01	0.73	y
Aug-02	0.99	0.9900	26.94	408.79	0	0.98	0.0083	0.0083	0.84%	0.84%	0.0001	0.0000	1.12	0.85	y
Sep-02	1.04	0.9900	28.38	465.95	0	1.03	0.0081	0.0081	0.77%	0.77%	0.0001	0.0026	1.17	0.90	y
Oct-02	1.04	1.0400	29.67	497.14	0	1.09	-0.0457	0.0457	-4.39%	4.39%	0.0019	0.0000	1.22	0.95	y
Nov-02	1.04	1.0400	28.85	491.24	0	1.07	-0.0337	0.0337	-3.24%	3.24%	0.0011	0.0000	1.21	0.94	y
Dec-02	1.04	1.0400	26.27	459.59	0	1.03	0.0121	0.0121	1.16%	1.16%	0.0001	0.0000	1.16	0.89	y
Jan-03	1.12	1.0400	29.42	473.50	0	1.07	0.0516	0.0516	4.61%	4.61%	0.0025	0.0059	1.20	0.93	y
Feb-03	1.12	1.1200	32.94	505.15	0	1.16	-0.0377	0.0377	-3.37%	3.37%	0.0011	0.0000	1.29	1.02	y
Mar-03	1.20	1.1200	35.87	477.85	0	1.17	0.0297	0.0297	2.48%	2.48%	0.0007	0.0051	1.31	1.03	y
Apr-03	1.20	1.2000	33.55	433.21	0	1.15	0.0470	0.0470	3.92%	3.92%	0.0015	0.0000	1.29	1.02	y
May-03	1.13	1.2000	28.25	413.42	0	1.09	0.0429	0.0429	3.80%	3.80%	0.0013	0.0034	1.22	0.95	y
Jun-03	1.07	1.1300	28.14	399.76	0	1.05	0.0223	0.0223	2.09%	2.09%	0.0004	0.0028	1.18	0.91	y
Jul-03	1.07	1.0700	30.72	408.32	0	1.05	0.0164	0.0164	1.53%	1.53%	0.0002	0.0000	1.19	0.92	y
Aug-03	1.07	1.0700	30.76	402.87	0	1.05	0.0194	0.0194	1.81%	1.81%	0.0003	0.0000	1.19	0.91	y
Sep-03	1.07	1.0700	31.59	430.90	0	1.08	-0.0065	0.0065	-0.61%	0.61%	0.0000	0.0000	1.21	0.94	y
Oct-03	1.07	1.0700	28.29	411.98	0	1.03	0.0387	0.0387	3.61%	3.61%	0.0013	0.0000	1.17	0.90	y

	Jet Fuel Price	Multiple regression model											Confidence Interval		
		Jet Fuel Price lagged 1 month	WTI (U$S/Barrel) lagged 1 month	IPP (Oil and gas) lagged 1 month	Dummy	Predicted Jet Fuel Price	e	|e|	PE	APE	Theil's U		Upper bound	Lower-bound	Prediction included?
Nov-03	1.07	1.0700	30.33	407.32	0	1.05	0.0210	0.0210	1.96%	1.96%	0.0004	0.0000	1.18	0.91	y
Dec-03	1.07	1.0700	31.09	423.09	0	1.07	0.0034	0.0034	0.32%	0.32%	0.0000	0.0000	1.20	0.93	y
Jan-04	1.02	1.0700	32.15	460.81	0	1.10	-0.0808	0.0808	-7.92%	7.92%	0.0057	0.0022	1.24	0.96	y
Feb-04	1.15	1.0200	34.27	438.42	0	1.09	0.0630	0.0630	5.48%	5.48%	0.0038	0.0162	1.22	0.95	y
Mar-04	1.14	1.1500	34.74	450.11	0	1.15	-0.0143	0.0143	-1.25%	1.25%	0.0002	0.0001	1.29	1.02	y
Apr-04	1.14	1.1400	36.76	447.51	0	1.17	-0.0288	0.0288	-2.53%	2.53%	0.0006	0.0000	1.30	1.03	y
May-04	1.14	1.1400	36.69	452.81	0	1.17	-0.0314	0.0314	-2.76%	2.76%	0.0008	0.0000	1.31	1.04	y
Jun-04	1.20	1.1400	40.28	468.82	0	1.22	-0.0177	0.0177	-1.48%	1.48%	0.0002	0.0028	1.35	1.08	y
Jul-04	1.25	1.2000	38.02	459.46	0	1.21	0.0355	0.0355	2.84%	2.84%	0.0009	0.0017	1.35	1.08	y
Aug-04	1.29	1.2500	40.69	500.55	0	1.29	0.0016	0.0016	0.13%	0.13%	0.0000	0.0010	1.42	1.15	y
Sep-04	1.32	1.2900	44.94	572.17	0	1.39	-0.0730	0.0730	-5.53%	5.53%	0.0032	0.0005	1.53	1.26	y
Oct-04	1.38	1.3200	45.95	562.39	0	1.41	-0.0299	0.0299	-2.16%	2.16%	0.0005	0.0021	1.55	1.27	y
Nov-04	1.52	1.3800	53.13	584.36	0	1.52	-0.0016	0.0016	-0.11%	0.11%	0.0000	0.0103	1.66	1.39	y
Dec-04	1.45	1.5200	48.46	523.62	0	1.50	-0.0460	0.0460	-3.17%	3.17%	0.0009	0.0021	1.63	1.36	y
Jan-05	1.40	1.4500	43.33	535.96	0	1.42	-0.0221	0.0221	-1.58%	1.58%	0.0002	0.0012	1.56	1.29	y
Feb-05	1.45	1.4000	46.84	468.54	0	1.39	0.0557	0.0557	3.84%	3.84%	0.0016	0.0013	1.53	1.26	y
Mar-05	1.46	1.4500	47.97	492.53	0	1.44	0.0181	0.0181	1.24%	1.24%	0.0002	0.0000	1.58	1.31	y
Apr-05	1.55	1.4600	54.31	526.16	0	1.53	0.0188	0.0188	1.21%	1.21%	0.0002	0.0038	1.67	1.40	y
May-05	1.54	1.5500	53.04	595.20	0	1.60	-0.0597	0.0597	-3.88%	3.88%	0.0015	0.0000	1.74	1.46	y
Jun-05	1.55	1.5400	49.83	578.43	0	1.55	-0.0025	0.0025	-0.16%	0.16%	0.0000	0.0000	1.69	1.42	y
Jul-05	1.59	1.5500	56.26	567.57	0	1.61	-0.0250	0.0250	-1.57%	1.57%	0.0003	0.0007	1.75	1.48	y
Aug-05	1.60	1.5900	58.70	595.66	0	1.67	-0.0742	0.0742	-4.64%	4.64%	0.0022	0.0000	1.81	1.54	y
Sep-05	1.74	1.6000	64.97	629.08	0	1.76	-0.0226	0.0226	-1.30%	1.30%	0.0002	0.0077	1.90	1.63	y
Oct-05	1.95	1.7400	65.57	706.89	0	1.88	0.0733	0.0733	3.76%	3.76%	0.0018	0.0146	2.01	1.74	y
Nov-05	1.93	1.9500	62.37	717.63	0	1.94	-0.0103	0.0103	-0.53%	0.53%	0.0000	0.0001	2.08	1.80	y
Dec-05	1.83	1.9300	58.30	665.12	0	1.86	-0.0279	0.0279	-1.53%	1.53%	0.0002	0.0027	1.99	1.72	y
Jan-06	1.83	1.8300	59.43	703.26	0	1.85	-0.0207	0.0207	-1.13%	1.13%	0.0001	0.0000	1.99	1.71	y

	Multiple regression model													Confidence Interval		
	Jet Fuel Price	Jet Fuel Price lagged 1 month	WTI (U$S/Barrel) lagged 1 month	IPP (Oil and gas) lagged 1 month	Dummy	Predicted Jet Fuel Price	e	|e|	PE	APE	Theil's U		Upper bound	Lower-bound	Prediction included?
Feb-06	2.20	1.8300	65.51	755.28	1	2.14	0.0559	0.0559	2.54%	2.54%	0.0009	0.0409	2.28	2.01	y
Mar-06	2.21	2.2000	61.63	798.77	1	2.29	-0.0792	0.0792	-3.58%	3.58%	0.0013	0.0000	2.43	2.15	y
Apr-06	2.28	2.2100	62.90	736.64	1	2.27	0.0125	0.0125	0.55%	0.55%	0.0000	0.0010	2.40	2.13	y
May-06	2.34	2.2800	69.69	795.12	1	2.40	-0.0623	0.0623	-2.66%	2.66%	0.0007	0.0007	2.54	2.27	y
Jun-06	2.55	2.3400	70.94	777.31	1	2.43	0.1207	0.1207	4.73%	4.73%	0.0027	0.0081	2.57	2.29	y
Jul-06	2.62	2.5500	70.96	785.73	1	2.52	0.0960	0.0960	3.66%	3.66%	0.0014	0.0008	2.66	2.39	y
Aug-06	2.55	2.6200	74.41	796.00	1	2.60	-0.0450	0.0450	-1.77%	1.77%	0.0003	0.0007	2.73	2.46	y
Sep-06	2.55	2.5500	73.05	797.53	1	2.55	-0.0025	0.0025	-0.10%	0.10%	0.0000	0.0000	2.69	2.42	y

ME	**MAE**	**MPE**	**MAPE**	**Theil's U**	**Total Yes**	54
0.0003	0.0372	-0.13%	2.86%	0.5564	**Total No**	0
					Total Points	54
					% Yes	100.00%

Bibliography

Argentine Air Force Financial Report 2001-2005. Argentine Air Force, Condor Building, Buenos Aires, Argentina, February 2006.

Argentine Secretary of Energy. Market Information, Oil Market. 14 August 2006 http://energia.mecon.gov.ar/home_pet/home_pet.asp.

Banks Jerry, John S Carson II, Barry L Nelson, and David M. Nicol. *Discrete-Event System Simulation (4th Edition)*. New Jersey. Pearson, 2004.

BMO Commodity Derivatives Group. *Hedging Jet Fuel Purchases.* September 2005. 14 August 2006 http//www.bmocm.com/products/marketrisk/commodity/images/ hedging_jetkero_prices.pdf.

Box George. "Science and Education". *Journal of American Statistical Association.* Number 71:791-799, 1976.

Burke Kenneth. *Building a Consensus Forecast for Crude Oil Prices.* MS thesis, AFIT/GLM/ENV/06-04. Air Force Institute of Technology, Wright Patterson Air Force Base OH, March 2005.

Cavallo Alfred. *Predicting the Peak in World Oil Production*. 2003. 14 August 2006 http//www.springerlink.com.

Coloma German. *Analysis of the behavior of the Argentine fuel market*. Political Economy Asociation, XXXIII Annual Meeting, 1998. 14 August 2006 http//www.aaep.org.ar/espa/anales/resumen_98/coloma.htm.

Cortés Conde, Roberto. *The Argentine Crisis 2001-2002*. Book of Economy, Volume 40, Number 121, Pages 762-767, December. 2003. ISSN 0717-6821

De Dicco Ricardo Andrés. *El costo del barril de petróleo crudo en Argentina*. El Salvador University, IDICSO. Área Recursos Energéticos y Planificación para el Desarrollo. August 2004. 14 August 2006.http//www.salvador.edu.ar/csoc/idicso/energia/ papelago5.htm.

Dees Stephane, Karadeloglou Pavlos, Robert Kaufmann and Marcelo Sanchez, *Modelling the World Oil Market Assessment of a Quarterly Econometric Model.* 14 August 2006. http//:www.bu.edu/cees/people/faculty/kaufmann/documents /oil_market_oct05.pdf.

Freund Rudolf, Ramon Littell, and Lee Creighton. *Regression Using JMP.*Cary, Noth Carolina. Wiley Interscience, 2003.

Gadano Nicolas and Federico Sturzenegger. *La Privatización de Reservas en el Sector Hidrocarburifero. El caso de Argentina.* Torcuato Di Tella University. Buenos Aires, Argentina, September, 1998. 14 August 2006. http//www.utdt.edu /~fsturzen/sem240998.pdf.

Garson David. *Testing for Assumptions.* Undated. 17 November 2006. http://www.2.chass.ncsu.edu/garson/pa765/assumpt.htm

Gately Dermot "Strategies for OPEC's Pricing and Output Decisions. *The Energy Journal*, Volume 16, Number 3, 1995.

Hughes Tom and Evan Rolek. "Fidelity and Validity: Issues of Human Behavioral Representation Requirements Development". *Proceedings of the 2003 Winter Simulation Conference, 2003*. 19 October 2006. http//www.informs-cs.org/wsc03papers/119.pdf.

Kasprzak Mary. *Forecasting Jet Fuel Prices Using Artificial Neural Networks.* MS Thesis. Naval Postgraduate School, Monterrey, California., March, 1995.

Kutner Michael, Christopher Nachtsheim, John Neter, and Willian Li. "*Applied Linear Statistical Models (5th Edition)"*. New Jork, NY. McGraw-Hill, 2005.

Leedy Paul and Jeanne Ormrod. *Practical Research. Planning and Design (8th Edition).*Columbus OH. Pearson, 2005.

Makridakis Spyros, Steven C. Wheelwright, and Rob J. Hyndman. *Forecasting: Methods and Applications (3rd Edition)*. New Jersey. Willey, 1998.

McCalve James, P. George Benson, and Terry Sincich. *Statistics for Business and Economics (9th Edition)*. New Jersey. Pearson, 2005.

Mercuri Pablo Antonio. "*Asimetrías en la respuesta de los precios de los combustibles líquidos a cambios en el precio del crudo: El caso argentino*." Political Economy Asociation., XXXIV Annual Meeting, 2001. 14 August 2006 http//www.aaep.org.ar/espa/anales/resumen_01/mercuri.htm.

Michalewicz Zbigniew and David Fogel. *How to solve it: Modern Heuristics (2nd Edition).*Germany. Springer, 2004.

New York Mercantile Exchange Glossary, Edition 2003, page 24. 14 August 2006 http://www.nymex.com/media/glossary.pdf.

Platts, Co. 2006. Oil Market information, 1 August 2006 http//www.platts.com/oil/resources.

Powell Stephen G. "The Target Capacity-Utilization of OPEC and the Dynamic of the World Oil Market," *The Energy Journal,* Volume 11, Number 3, 1990:27-37.

Ramcharran Harri. "Oil production responses to price changes: an empirical application of the competitive model to OPEC and non-OPEC countries". *Energy Economics*, Volume 24, Number 2, March 2002: 97-106.

Scheimberg Sebastian. "Comentario al trabajo de German Coloma: Analisys of the behavior of the Argentine fuel market". Political Economy Asociation, XXXIII Annual Meeting, 1998. 14 August 2006. http//www.aaep.org.ar/espa/anales/comentario-replicas/Coloma%2Bcomentario_scheimberg.pdf.

Senge Peter. *The Fifth Discipline.(1st Edition).*New York. Doubleday, 1990.

Senge Peter, Kleiner Art, Roberts Charlotte, Ross Richard, Roth George and Smith Bryan. *The Dance of Change: The Challenges to Sustaining Momentum in Learning Organizations (1st Edition)* .New York, Doubleday, 1999.

Stergiuo Christos and Dimitrios Siganos. *Neural Networks*. 14 August 2006. http//www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html.

United States Department of Energy, Energy Information Administration. *The Energy Modeling System: An Overview 2003*. 14 August 2006 http//www.eia.doe.gov/aiaf/aeo/overview/download.html.

- - - - Energy Information Administration. *Short term Energy Outlook. Energy Prices Model Description.* Washington, 2002 . 14 August 2006. http//www.eia.doe.gov/emeu/steo/pub/document/textpr.html.

United States General Accounting Office. *Defense Logistics: Better Fuel Pricing Practices Will Improve Budget Accuracy*. Report to Congressional Committees, June 2002. 14 August 2006. http//www.gao.gov/newitems/d02582.pdf.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
09-03-2007	Master's Thesis	June 2006 - March 2007

4. TITLE AND SUBTITLE
Predicting Argentine Jet Fuel Prices

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Juan A. Salaverry, Lieutenant Colonel, Argentine Air Force

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/GLM/ENC/O7M-01

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE, DISTRIBUTION UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Oil distillates are considered important elements to accomplish the missions of the Argentine Air Force (AAF). Of all oil products consumed by the AAF, jet fuel is the resource with highest demand and at the end of the day the most expensive support item procured by the Argentine Air Force. Accurate predictions of Argentine jet fuel prices are necessary to improve AAF financial and logistics planning. This thesis presents a systematic, statistical regression approach to forecast Argentine jet fuel prices. The application of this methodology has allowed us to obtain a very useful model that utilizes information available on the internet to produce forecasting of jet fuel prices with average percentage absolute errors lower than 3%

15. SUBJECT TERMS
Forecasting - Regression Analysis - Budget Improvement

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	UU	84	DR.EDWARD D. WHITE III, PhD
					19b. TELEPHONE NUMBER *(Include area code)* (937) 255-3636 X4540

Reset

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18