Southwest Region University Transportation Center

Managing Commodity Risks in Highway Contracts: Quantifying Premiums, Accounting for Correlations Among Risk Factors, and Designing Optimal Price-adjustment Contracts

SWUTC/11/161104-1

Texas Transportation Institute
Texas A&M University System
College Station, Texas 77843-3135

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
SWUTC/11/161104-1		

4. Title and Subtitle	5. Report Date
Managing Commodity Risks in Highway Contracts: Quantifying Premiums, Accounting for Correlations Among Risk Factors, and Designing Optimal Price-adjustment Contracts	September 2011
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Xue Zhou and Ivan D. Damnjanovic	Report 161105-1
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Texas Transportation Institute Texas A&M University System College Station, Texas 77843-3135	11. Contract or Grant No. 10727
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Southwest Region University Transportation Center Texas Transportation Institute Texas A&M University System College Station, Texas 77843-3135	14. Sponsoring Agency Code

15. Supplementary Notes

Supported by general revenues from the State of Texas.

16. Abstract

It is a well-known fact that macro-economic conditions, such as prices of commodities (e.g. oil, cement and steel) affect the cost of construction projects. In a volatile market environment, highway agencies often pass such risk to contractors using fixed-price contracts. In turn, the contractors respond by adding premiums in bid prices. If the contractors overprice the risk, the price of fixed-price contract could exceed the price of the contract with adjustment clauses. Consequently, highway agencies have opportunity to design a contract that not only reduces the future risk of exposure, but also reduces the initial contract price.

The main goal of this report is to investigate the impact of commodity price risk on construction cost and the optimal risk hedging of such risks using price adjustment clauses. More specifically, the objective of the report is to develop models that can help highway agencies manage commodity price risks. In this report, weighted least square regression model is used to estimate the risk premium; both univariate and vector time series models are estimated and applied to simulate changes in commodity prices over time, including the effect of correlation; while genetic algorithm is used as a solution approach to a multi-objective optimization formulation. The data set used in this report consists of TxDOT bidding data, market-based data including New York Mercantile Exchange (NYMEX) future options data, and Engineering News-Record (ENR) material cost index data. The results of this report suggest that the optimal risk mitigation actions are conditional on owners' risk preferences, correlation among the prices of commodities, and volatility of the market.

17. Key Words	18. Distribution Statement
Unit Cost, Contracting, Escalation Clauses, Pricing Risk	No restrictions. This document is available to the public through NTIS: National Technical Information Service 5285 Port Royal Road Springfield, Virginia 22161

19. Security Classif.(of this report)	20. Security Classif.(of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	174	

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

MANAGING COMMODITY RISKS IN HIGHWAY CONTRACTS: QUANTIFYING PREMIUMS, ACCOUNTING FOR CORRELATIONS AMONG RISK FACTORS, AND DESIGNING OPTIMAL PRICE-ADJUSTMENT CONTRACTS

Xue Zhou (Ph.D)
Graduate Research Assistant
Zachry Department of Civil Engineering
Texas A&M University System

Ivan D. Damnjanovic (Ph.D)
Assistant Professor
Zachry Department of Civil Engineering
Texas A&M University System

SWUTC/11/161104-1

Southwest Region University Transportation Center
Texas Transportation Institute
Texas A&M University System
College Station, Texas, 77843-3135

September 2011

ABSTRACT

It is a well-known fact that macro-economic conditions, such as prices of commodities (e.g. oil, cement and steel) affect the cost of construction projects. In a volatile market environment, highway agencies often pass such risk to contractors using fixed-price contracts. In turn, the contractors respond by adding premiums in bid prices. If the contractors overprice the risk, the price of fixed-price contract could exceed the price of the contract with adjustment clauses. Consequently, highway agencies have opportunity to design a contract that not only reduces the future risk of exposure, but also reduces the initial contract price.

The main goal of this report is to investigate the impact of commodity price risk on construction cost and the optimal risk hedging of such risks using price adjustment clauses. More specifically, the objective of the report is to develop models that can help highway agencies manage commodity price risks. In this report, weighted least square regression model is used to estimate the risk premium; both univariate and vector time series models are estimated and applied to simulate changes in commodity prices over time, including the effect of correlation; while genetic algorithm is used as a solution approach to a multi-objective optimization formulation. The data set used in this report consists of TxDOT bidding data, market-based data including New York Mercantile Exchange (NYMEX) future options data, and Engineering News-Record (ENR) material cost index data. The results of this report suggest that the optimal risk mitigation actions are conditional on owners' risk preferences, correlation among the prices of commodities, and volatility of the market.

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ACKNOWLEDGMENT

The authors recognize that support for this research was provided by a grant from the U.S. Department of Transportation, University Transportation Centers Program to the Southwest Region University Transportation Center which is funded, in part, with general revenue funds from the State of Texas.

EXECUTIVE SUMMARY

How to quantify the effects of commodity price risk factors and develop optimal strategy to manage them is the topic of this report. Commodity prices, such as, asphalt, crude oil, cement and steel, have been identified as one of the most important risk factors affecting the cost of construction projects. In fact, the escalation of commodity prices often results in significant financial hardships for unprepared contractors and owners. Contractors rely on contingency plans to deal with volatile commodity prices. This holds true particularly for contracts where price adjustments are not permitted. Consequently, it is of great importance for highway agencies to negotiate contracts that would optimally manage risk exposure. Price adjustment clauses with trigger values have been used by many public agencies to manage risks with volatile commodity prices. The overall goal of this research is to develop a comprehensive methodological framework to assess the effect of trigger values and correlations of commodity prices, and to determine the optimal risk hedging strategies using escalation clauses with trigger barriers. The developed framework accounts for correlations of commodity price risks when balancing the objectives from the perspective of the agency.

To determine optimal hedging of commodity risk using escalation clauses, it is essential to first determine the price of risk in these contracts. The report presents a link between the average risk premium and the expected change and volatility of commodity prices. This part explains the relationship between the unit bid prices of selected control items and the risk factors (e.g., increase in commodity prices); in other words, it estimates the impact that volatile commodity prices have on the unit bid prices (i.e., risk premiums).

Then the report provides an optimal way to manage commodity-based correlated risks in contracting using an escalation clause with trigger barriers. Such an escalation clause can be used to specify the amount of risk the agency would like to be exposed to during construction via the barrier levels. It allows balancing between (1) initial payment in the form of risk premium before a contract begins, and (2) future risk exposure during construction. The developed framework also accounts for correlations of commodity risks when balancing the objectives.

The optimal risk mitigation actions are conditional on the owner's risk preferences specified using CVaR-based measures. The solution approach to the problem is based on a multi-objective optimization formulation (or a single-objective degenerate case) and

genetic algorithms as a solution approach. The key insights of this study are as follows: (1) TxDOT could have used the escalation clauses on the specific control items to reduce the premium and take on the risk in the historical extreme volatile commodity market environment. A case study investigates what would have happened if they did, and what should have been the optimal strategy using the information that was available at that time; (2) It should be aware that an owner's risk preference affects the choice of contracting. The resulting multi-objective optimization problem considers a number of trade-offs in optimal solutions. If the decision-making process is governed by a "risk-seeking" policy, the decision-maker could consider optimal solutions in the bottom-right corner of the Pareto front; for a "risk averse" posture, the owner should consider optimal solutions in the upper-left corner of the Pareto front; and for a "risk neutral" attitude, the middle part should be considered; (3) The effect of correlations (between prices of commodities) should not be ignored. The potential risk exposure of the project is considerably under-estimated if the commodity prices are forecasted independently. The effect of correlation between the prices of commodities was captured using the vector Time Series model; and (4) It is essential to consider multi-objective optimization with multiple barrier levels using the vector time series model when making optimal decisions to hedge against the risks from volatile prices of commodities. This is because the solutions from multi-variate optimization can considerably reduce the total cost of a project.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

TABLE Page

1. INTRODUCTION

Project risk management plays an important role in development and management of projects (Olsson, 2008). Construction projects, much like any other engineering projects, exhibit many uncertain factors, both internal and external to the project environment (Jaafari, 2001; Rahman and Kumaraswamy, 2002). In order to assure that the project meets the target values, such as, initial capital expenditure, project managers need to carefully identify, assess, and analyze the factors and events that can cause deviations from the plans.

Risks associated with the cost of delivery of capital projects are often correlated. This correlation can be observed from a project environment perspective, where risks associated with different work packages and unit bid items show correlations. In fact, the correlations often come from the factors or events that affect all work packages. For example, uncertain material prices may impact the costs of all work packages that include this type of material.

Project risks are typically managed using comprehensive strategies that include selection of delivery and contracting methods. Here, contracting method allows for allocation of the risks between the owner and the contractor. For example, projects can be delivered using lump-sum, unit-based, or reimbursable contracts which imply different risk allocation schemes. Further, contracts can be adjusted to reflect who is willing to take on specific risk factors and at what level. Here, these contracts are referred to as contracts with adjustment clauses that are triggered based on realization of certain pre-defined events.

Price adjustment clauses have been used by many public agencies to manage risks with volatile commodity prices. For example, Washington State Department of Transportation (WSDOT) applies the same trigger value of 10 percent for fuel cost adjustments as long as the projects meet certain requirements (e.g. projects longer than 200 working days) (AASHTO, 2009).

How to quantify the effects of commodity price risk factor and develop optimal strategy to manage them is the topic of this report. This section introduces the motivation for this report, presents the goals and objectives, summarizes the contributions, and outlines the remainder of the report.

1.1 Background and Identified Problem

Commodity prices, such as, asphalt, crude oil, cement and steel, have been identified as one of the most important risk factors affecting the cost of construction projects (Kangari, 1995; Zhi, 1995; Hastak and Shaked, 2000; Jaafari, 2001; Baloi and Price, 2003; Mendell, 2006). In fact, the escalation of commodity prices often results in significant financial hardships for unprepared contractors and owners. For example, in the mid-2000s, a number of highway contractors were affected by escalating material prices (the cost of liquid asphalt, cement, oil, and steel) (Gallagher and Riggs, 2006). Even though contractors observed materials cost increase in 2001-2003, they still failed to adequately prepare for unexpected price increases in 2004-2005 (Mendell, 2006). Construction cost is particularly sensitive to the cost of energy. Wilmot and Cheng (2003) identified the increase in the cost of petroleum products and construction machinery as the main cause of the rise in construction costs. For example, energy costs propagate through a number of construction activities from petroleum-derived products, machinery costs, to transportation costs. Thus, volatile prices of commodities including energy typically result in volatile costs of construction projects.

Contractors rely on contingency plans to deal with volatile commodity prices. This holds true particularly for contracts where price adjustments are not permitted. When contractors cannot obtain firm price quotes from material suppliers for the duration of the project, they typically inflate the bid prices to protect their marginal profit against possible price increases (FHWA, 1980). In highway contracting, both locked-in unit prices and unit prices with adjustment clauses are currently being used. In fixed-price unit-based contracts, owners transfer the price increase risk to contractors, while in the unit-based contracts with price adjustments, owners accept such risk. If contractors overprice the risk, the prices of fixed-price contracts could exceed those of contracts with adjustment clauses. In fact, "large" contingencies are often included in the initial estimates of bid items to hedge against the risk exposures (Gallagher and Riggs, 2006). Consequently, it is of great importance for highway agencies to negotiate contracts that would optimally manage risk exposure.

Starting from 1980s, the Federal Highway Administration (FHWA) of the U.S. Department of Transportation began to encourage state highway agencies to consider price adjustment provisions to respond to price volatility of construction materials (FHWA, 1980). Adjustment clauses include both upward and downward adjustments for volatile prices of construction materials. For example, escalation clauses are triggered if the crude oil spot price exceeds a trigger barrier. These contracts are often referred to as knock-in options under trigger clauses in which a holder receives a payment conditional on the underlying prices reaching a certain trigger barrier. In theory, if such escalation clauses are added to the contracts, it is expected that the contractors will lower their bids. This is because a portion of the risk is transferred from the contractors to the highway agencies. For example, WSDOT reported that the Hot Mix Asphalt (HMA) and fuel escalation clause had improved competitive bidding climate (GNB, 2007). Price escalation lessened contractors' fear of increased prices, while in return WSDOT took on the risk of future material cost. WSDOT reported that it had obtained the bids that reflected the current market condition and excluded speculative pricing (GNB, 2007).

Even though price adjustment clauses are being used by many highway agencies for a while (e.g., since 1980s), to the author's best knowledge, there are no studies that investigate the procedure to set the trigger values and account for the effect of correlations among commodity prices on unit bid prices. McGoldrick (2006) recommended that it was better for the owner to pay the actual increase in costs rather than pay a significant contingency included within the contract price, which might ultimately be higher than the cost of material increases. However, this result should be taken with caution since risk preference may contribute to the decision to go one way or the other. Thus, when risk and uncertainty from volatile markets result in overpriced bids, the potential payoff of including adjustment clauses for highway agencies is high. The investigation of this phenomena and the development of optimal risk hedging strategies based on defining adjustment clauses is the main topic of this study.

1.2 Research Goal and Objectives

The overall goal of this research is to develop a comprehensive methodological framework to assess the effect of trigger values and correlations of commodity prices, and to determine the optimal risk hedging strategies using escalation clauses with trigger barriers. The developed framework accounts for correlations of commodity price risks when balancing the objectives from the perspective of the agency. The objectives pertaining to this larger goal are as follows:

Objective 1: Develop models that can be used to price bid items and estimate the average risk premiums due to commodity price risks. The developed models should consider the available data and factors influencing the unit bid prices;

Objective 2: Identify the correlations between risk factors (commodity prices) that affect unit bid prices in highway contracts. This relationship should reflect the changes of economic environment as it occurs in reality;

Objective 3: Develop forecasting models for commodity prices to account for historical changes in commodity prices. Univariate time series should fit the historical series independently, while vector time series should account for co-movement; and

Objective 4: Formulate optimization models to determine the optimal hedging strategies. The multi-objective optimization formulation should take into account the effect of correlations among commodity risk factors, and the agencies' risk preferences.

1.3 Research Contributions

This research contributes to the field of construction engineering and management in two major ways. The first way of contributing to the field is by developing the risk premium pricing model for highway construction projects when unit bid prices are significantly influenced by uncertain economic conditions, such as, volatile commodity prices. The second way of contributing to the field is by developing the optimal risk hedging model which is based on the agency's risk preferences and takes into account the effect of correlations between the risk factors. There are a number of benefits to the agencies from this study, for example, 1) assessing the risk premiums in their bids; 2) designing optimal risk hedging contracts with escalation clauses; and 3) evaluating the contracting consequences of different pavement designs such as material requirements.

1.4 Study Limitations

The report focuses on developing optimal risk hedging strategies for volatile prices of commodities (such as, crude oil, cement and steel). The limitations of this study are as follows:

(1) It is based on Texas Department of Transportation bidding data;

(2) It is focused on highway construction projects;

(3) It considers unit cost types of contracts;

(4) It is focused only on a limited number of unit bid items;

(5) It only considers risks related to the prices of commodities;

1.5 Report Outline

This report is organized in eight sections. Following this section, in which the motivation, objectives, and contributions of this research are introduced, the next section presents an overview of the background literature, covering four related topics: construction cost forecasting, construction risk management, risk preference measures, and optimization and its solution methods. After that, Section 3 presents the overall methodology and the data sets used for developing the models.

In Section 4, the methodological framework for pricing bid item and risk is presented. The discussion includes identifying risk factors for unit bid items, estimating the risk premiums due to the impact of changes in the commodity prices, and explaining how the risk premiums vary according to different barrier levels.

Univariate and vector time series models for simulating commodity prices are presented in Section 5. The process for developing autoregressive integrated moving average (ARIMA) model and vector autoregressive moving average (VARMA) model is shown. The VARMA model is based on relaxing the independence in the assumption of ARIMA model.

The formulation and the solution approach to optimal risk hedging problem are presented in Section 6. This section presents the formulation for both single-objective and multi-objective optimization, and discusses the advantage of a multi-objective approach.

Section 7 presents a case study for the models developed in the report. The case study illustrates the overall process using data from real TxDOT projects, and discusses the implications of the results.

Finally, Section 8 summarizes major findings and limitations, and presents the directions for future research work.

2. LITERATURE REVIEW

This section presents an overview of the background literature in four major areas pertaining to this report: construction cost forecasting, construction risk management, measuring risk preferences, multi-objective optimization and its solution approach. In the first subsection, a general background on construction cost forecasting is introduced. In the second subsection, a brief review of construction risk management is presented. In the third subsection, the approaches for risk measures are reviewed, while in the fourth subsection, the most commonly-used methods for multi-objective optimization are identified.

2.1 Construction Cost Forecasting

Regression methods have been extensively applied for forecasting future highway construction costs. For example, the models for predicting the unit cost of highway construction contracts in terms of dollars per mile have been developed by Hartgen and Talvitie (1995) and Stevens (1995), while Koppula (1981) and Hartgen *et al.* (1997) have developed models based on extrapolation of past trends in cost index movements. However, the validity of these models is mostly limited to the short term prediction.

Since 1970s, regression analysis has been used to use historical data to relate construction costs to the explanatory factors. Trost and Oberlender (2003) applied regression into cost estimating, while Lowe *et al.* (2006) described the development of regression models to predict the construction cost of buildings by performing both forward and backward stepwise analysis. In 1990s, neural networks have appeared as an alternative for estimating construction cost (Kim *et al.*, 2004). Emsley *et al.* (2002) used a neural networks model to predict total construction cost. Wilmot and Mei (2005) developed a neural network based procedure that can be used to estimate the escalation in the highway construction costs over time. Even though neural networks models have greater freedom to fit data than regression models (Wilmot and Mei, 2005), Emsley *et al.* (2002) reported that there was no significant difference in prediction accuracy between neural networks and regression models.

A number of risk factors affect the cost of construction projects. Kangari (1995) discussed risk factors as site access, availability of labor, equipment and material, productivity of labor (fatigue and safety) and equipment, defective designs, changes in work, safety, inflation, quality of work and others. Zhi (1995) identified risk factors for overseas construction projects as political situation, economical and financial situation, market fluctuations, law and regulations, labor, materials, equipments, and others. According to RS Means (2008), the cost of building construction varies based on a number of variables. In addition to quality, time, and productivity, the main factors affecting the costs include a) the size of project: the scope of work, and the type of construction project can have a significant impact on the cost. Economies of scale can reduce costs for large projects. Further, the risk of project complexity is usually attributed to the project size and long project durations; b) location: in dense urban areas, traffic and site storage limitations may increase costs. The projects located in a central business district (CBD) often exhibit the cost that is higher than the projects in remote rural location; c) time of year for bidding; d) weather conditions, and others. Due to the availability of data, this report only considers the risks from volatile prices of commodities.

To quantify the effect of the risk factors associated with oil market, Damnjanovic and Zhou (2009) developed a model that links the risk premium to both expected change and volatility in crude oil prices. The behavioral analysis provided evidence of the impact of fluctuated crude oil prices on unit bid prices without price adjustment clauses. This report expands this line of research to include the effect of price movement of other commodities as well.

2.2 Construction Risk Management

Probabilistic analysis, Monte Carlo and discrete-event simulation have been used to quantify risk and uncertainty in construction risk management. In probabilistic analysis, the mean and standard deviation of input variables are used as statistical measures of risk (Paek *et al.*, 1993), while the Monte Carlo simulation, a form of stochastic simulation, is used to obtain the probability of project outcome by carrying out a number of samples depending on the degree of confidence required (Akintoye and MacLeod, 1997). In addition to these methods, there are other risk analysis methods used in the literature, such as, decision analysis (Ng and Bjornsson, 2004), subjective probability analysis (Akintoye and MacLeod, 1997), fuzzy logic methods (Kangari and Riggs, 1989; Dikmen *et al.*, 2007), and formal risk management process (Tah and Carr, 2000; Carr and Tah, 2001). The objective of all these analyses is to quantify uncertainty in numerical terms.

For specific measure of uncertainty, contractors assign specific risk premium (i.e., reward for taking on the risk). This risk premium strategy is often used in construction projects to determine contingency allowance and cover unforeseen eventualities (Akintoye and MacLeod, 1997; Kartam and Kartam, 2001). The premium placed on sources of risk depends on the risk exposure from the sources, the likelihood of occurrence, the experience of the contractor in dealing with the particular type of risk, and decision makers' risk attitude (Akintove and MacLeod, 1997; Kartam and Kartam, 2001).

To the author's best knowledge, there are no studies that investigate and quantify risks of contracting with escalation clauses, even though escalation clauses have been widely used as an approach for risk mitigation. Further, there are no published studies on how to set an optimal level of barriers given agencies' risk preferences and the budget levels.

2.3 Risk Preference Measures

Risk measures are used to comprehend and compare the risk (e.g., the deviation from the expected value) and make decisions about the risk level willing to accept. Risk measures including variance (or standard deviation), Value-at-Risk (VaR), Conditional Value-at-Risk (CVaR) have been broadly used in the practice of portfolio optimization. Variance was the first proposed risk measure by Markowitz (1952) and is still the most widely used measure of risk (Roman *et al.*, 2007). VaR and CVaR are more commonly used and accepted, because they are concerned only with tails of the distribution (extremely unfavorable outcomes). They represent the mean shortfall at a specified confidence level (Mansini *et al.*, 2007; Roman *et al.*, 2007).

Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) can consider an arbitrarily defined percentile of potential losses. Both of these measures have been broadly applied in finance and corporate management. Value-at-Risk focuses on the tail of the distribution of unfavorable outcomes (losses), whereas Conditional Value-at-Risk focuses on the conditional expected losses given VaR (see Figure 2.1 [Sarykalin *et al.*, 2008]). For example, a decision-maker can set an event (amount of loss) that should not be exceeded, or exceeded only with a very small probability. For a specified probability level, β, the $\beta-VaR$ of a distribution is the lowest amount, α, such that with probability β, losses will not exceed α, whereas the $\beta-CVaR$ is the conditional expectation of losses greater than α (that is, CVaR measures the expected losses given a set of worst-case scenarios). VaR has a number of undesirable properties such as lack of sub-additivity, non-convexity, and non-smoothness (Palmquist *et al.*, 1999). In contrast, CVaR is sub-additive and convex; moreover, it is a coherent measure of risk (i.e., risk measures that satisfy the four desirable properties that presented by Artzner *et al.* [1999], are called "coherent"). Furthermore, mean-risk bi-criteria optimization models, such as mean-CVaR models, where the expected return is maximized and some (scalar) risk measure is minimized, have been widely used for optimizing portfolios (Mansini *et al.*, 2007; Roman *et al.*, 2007).

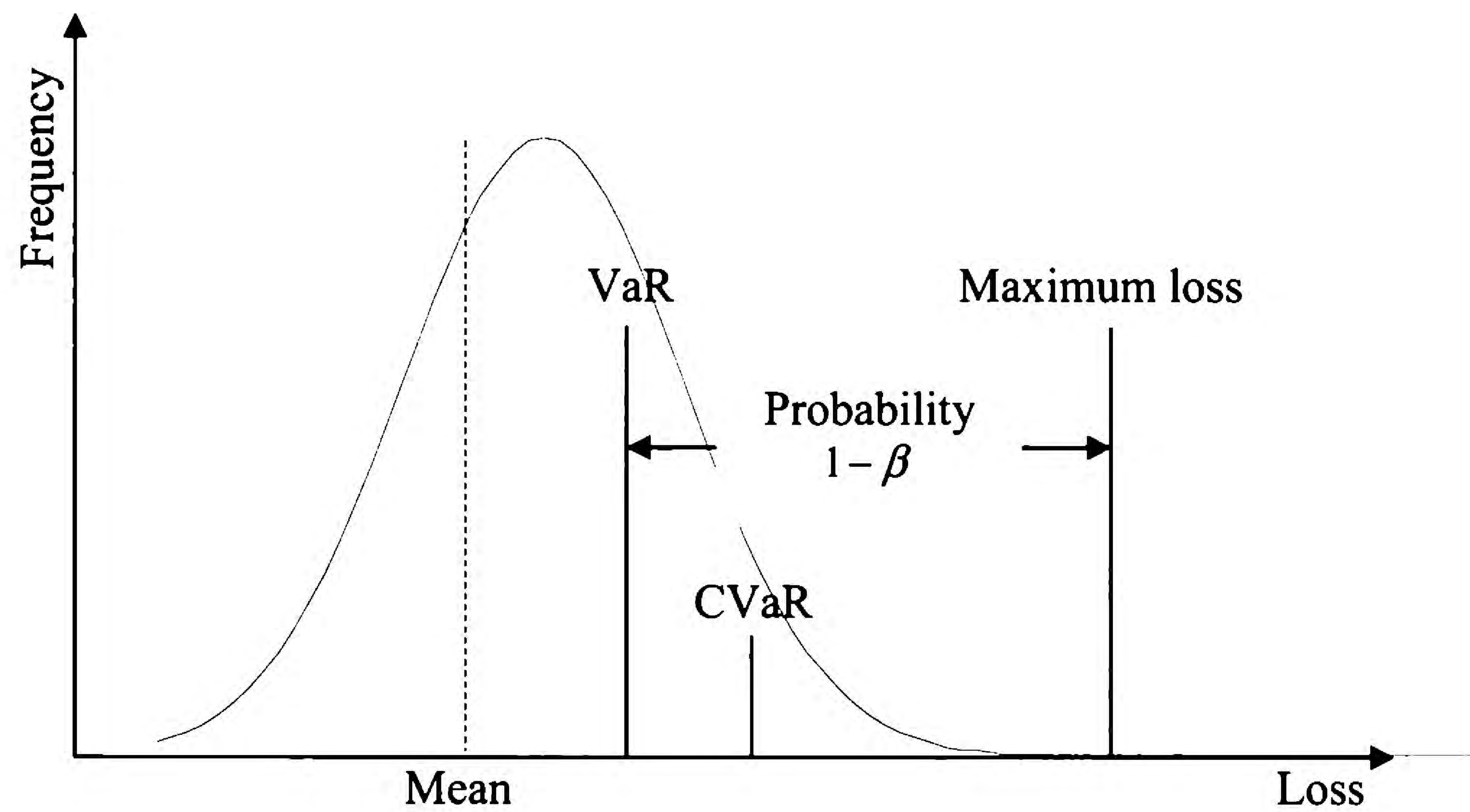


Figure 2.1 CVaR, VaR and deviation

Based on the loss distribution shown in Figure 2.1, this report defines two ways of quantifying losses: 1) "expected loss", which measures the average loss of the whole distribution of simulated losses, and 2) "unexpected loss", which represents the average of the losses greater than VaR. More details and discussions about these two defined losses will be shown in Section 6 and Section 7.

The previous research in construction risk management have been limited only on Value-at-Risk (VaR) as a risk measure for decision makings on bid/no bid on portfolio management level (Caron *et al.*, 2007); there are no construction risk management methods that considers Conditional Value at Risk as a measure for flexible decision makings to deal with economic risk factors, such as volatile commodity prices.

The agencies aim to minimize the initial cost of project, while accepting a tolerance level of risk. Hence, the optimization problems can be formulated as: 1) 'budget-unexpected loss' model, that is, the objective is to minimize the total budget cost at a given level of acceptable risk (which is represented as an extended application of mean-CVaR model); or 2) multi-objective, that is, minimizing both the budget (initial) cost and future the exposure. Due to their complexity, both of these optimization formulations require specialized solution algorithms.

2.4 Multi-Objective Optimization and Solution Approach

Multi-objective optimization is often used to solve the problems when project-related decisions include multiple conflicting criteria simultaneously. This modeling approach is used when the utility functions are not well known prior to the optimization process. This implies that the objectives could not be combined into a single one. In other words, the problem must be treated as a multi-objective optimization problem (MOP) with non-commensurable objectives (Tamaki *et al.*, 1996).

In multi-objective optimization models, the Pareto optimal solutions are generated first, and then, decision makers make choices and specify their preference information. In other words, a set of Pareto-optimal solutions exist in the absence of preference information. The choice for the "best" solution is then made based on higher-level information which is non-technical, qualitative, and experience-driven.

There are two goals in a multi-objective optimization: (1) convergence to the true Pareto-optimal front, and (2) maintenance of diversity among Pareto-optimal solutions (Deb, 1999b). The multi-objective optimization aims to find the global Pareto-optimal frontier, representing the "best" possible objective values (Deb and Gupta, 2005). As it is important for a multi-objective optimization to find solutions near, or on the true Pareto optimal front, it is necessary to obtain solutions that are as diverse as possible. If majority of solutions are found in a small region near, or on the true Pareto-optimal front, the purpose of multi-objective optimization is not served, because in such cases, many interesting solutions with large trade-offs among the objectives may not be discovered. (Deb, 1999b)

Genetic algorithms, a popular solution approach to multi-objective optimization problems, can be used to achieve the "two goals" of multi-objective optimization. For example, the Nondominated Sorting Genetic Algorithm II (NSGA-II) has a superior mechanism in terms of finding a diverse set of solutions and in converging to the true Pareto-optimal set (Deb *et al.*, 2002; Coello Coello, 2006). In the last decade, GA has been applied to construction industry problems as a robust approach for finding a near-optimal solution (Al-Tabtabai and Alex, 1999; Zheng *et al.*, 2004). The key advantage of GA when compared to conventional optimization methods is the ability to quickly solve difficult problems that are non-convex, integer, and non-continuous. In fact, optimization problems that involve risk measures are often non-convex.

GA is a heuristic random search technique based on the concept of natural selection and natural genetics of a population. Evolutionary algorithms (such as, genetic algorithm) deal simultaneously with a set of possible solutions (the so-called population) which allows to find several members of the Pareto optimal set in a single run of the algorithm (Coello Coello, 2006). Hence, it is a "population-based" method of searching large combinatorial spaces to find the near-optimal combination (Tam *et al.* 2001). Three operators—selection, crossover, and mutation—are implemented to generate the offspring that are ready for the next cycle (Zheng *et al.*, 2004). The application of GA in construction management include time-cost trade-off (Feng *et al.*, 1997; Li and Love, 1997), construction scheduling (Chan *et al.*, 1996; Leu and Yang, 1999; Dawood and Sriprasert, 2006), resource leveling (Leu et al., 2000; Senouci and Eldin, 2004), labor use optimization (Tam *et al.*, 2001), building portfolio management (Tong *et al.*, 2001), reliability-based optimization (Deshpande *et al.*, 2010), and others.

2.5 Summary

This section presents the literature review relevant to the overall objectives of the report and introduces the necessary background to analyze risks associated with adjustment clauses in highway construction contracts. The literature review identifies a lack of methodology for investigating the optimal risk hedging strategies using contracting methods with escalation clauses. In the following section, the overall methodology for studying such structure is presented along with the data sets required for its implementation.

3. THE OVERALL METHODOLOGY AND DATA SETS

The overall methodology and the available data set for this study are presented in this section. The framework contains three main parts as shown in Figure 3.1: (1) a model to price the unit bid items and the risk premiums, including the contract design using an escalation clause where barrier level is considered as the decision variable; (2) the time series models used for simulating commodity prices; and (3) the multi-objective optimization model where agencies' risk preferences (e.g., willingness to take on the risk of price escalation) are specified using CVaR-based risk measures. The following sections present the details of these components along with the corresponding data sets.

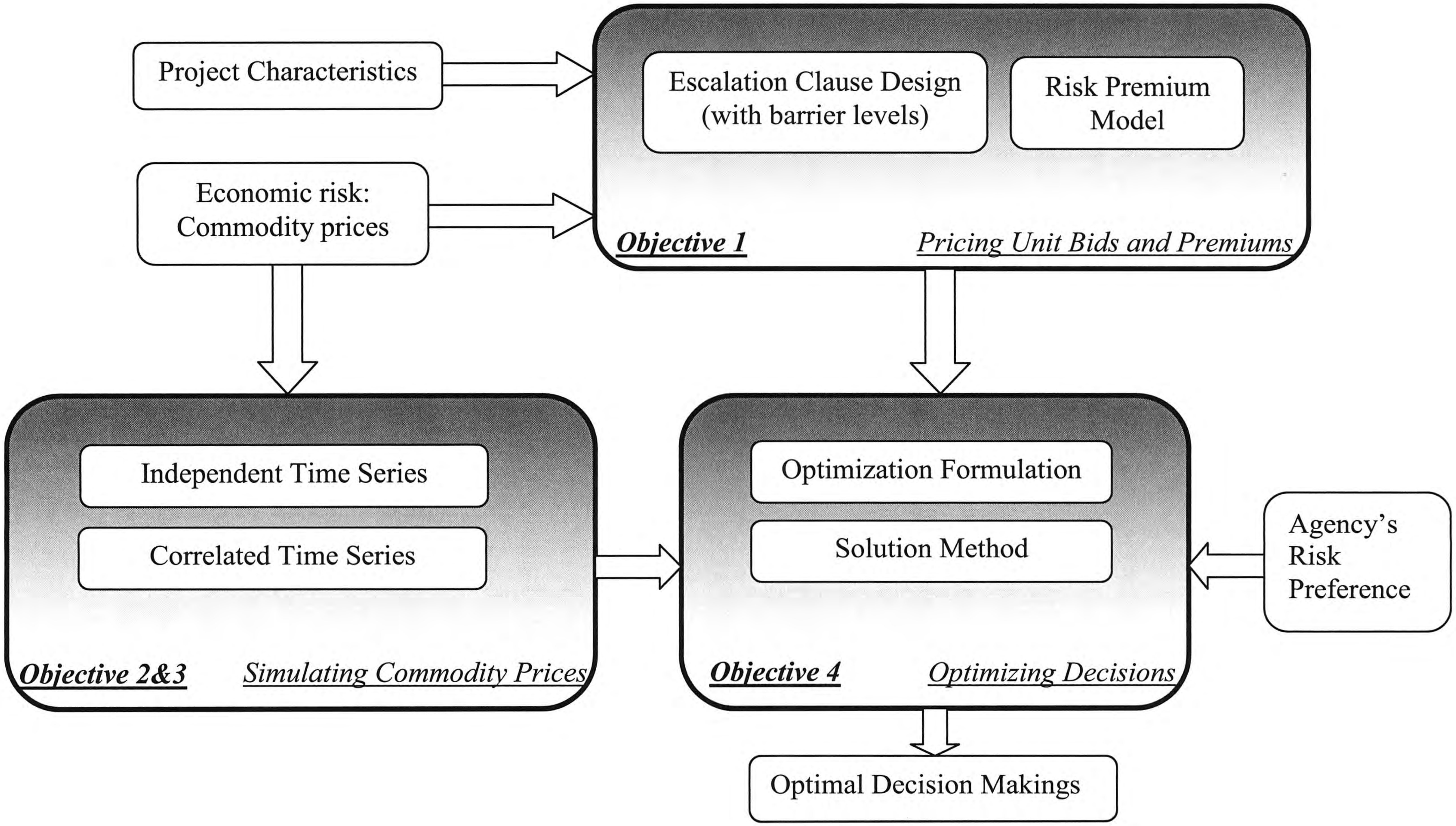


Figure 3.1 Research framework

3.1 Objective 1 – Pricing Risk Premiums

As previously discussed, size, scope of work, and type of construction project have a significant impact on the cost. For example, economies of scale can reduce the cost for large projects, which results in lower unit bid prices for projects with large quantities. However, the longer the duration of the project, the higher risk premium may be included in the initial bid; since the contractors are concerned with the volatile prices of commodities over an extended period of time.

Changes in commodity prices can significantly affect the risk premiums added to unit bid prices. For example, in mid 2000s, an increase in commodity prices drove the increase in construction bid prices (Gallagher and Riggs, 2006; Barraza *et al.*, 2000). Damnjanovic and Zhou (2009) showed that anticipated changes in oil prices (the difference between future and spot price) and the implied volatilities of oil prices significantly affected the prices of bid items for contracts without price adjustment clauses. The analysis linked the risk premium to both expected change and volatility in commodity prices (risk factors). Here the expected change represents the difference between future price and current price, while volatility measures variability or dispersion about a central tendency – a measure of the degree of price movement (Kotze, 2005). In fact, these two parameters (long-term drift rate and volatility) could be interpreted by a more formal representation of the market behavior of commodity prices from a stochastic modeling perspective (Trigeorgis, 2002). Hence, risk premium pricing depends on market behavior and the type of project and contract (i.e., escalation clause with trigger barrier level).

3.2 Objective 1 – Escalation Clauses with Trigger Barriers

FHWA and state departments of transportation (DOTs) have historically used various price adjustment mechanisms in highway contracts. For example, many state DOTs have implemented price adjustments that apply for both upward and downward movement of prices. In fact, there has not been a consistent way that price adjustments are implemented. The result of a survey performed by an AASHTO subcommittee in the fall of 2009 (AASHTO, 2009) shows that the specifications of price adjustments vary for different materials (fuel, asphalt cement, steel, Portland cement, and others), trigger values (0 percent, 5 percent, 10 percent, 15 percent, 20 percent and others), and other essential features, such as, the maximum escalation limit and the definition of controlling price index for adjustment. Hence, potential losses differ and must be computed according to specific contract provisions.

This study considers only upward price adjustments with trigger barriers for commodities. In fact, the trigger level is considered as a decision variable. The potential losses to a highway agency are then calculated if the simulated spot prices of commodities reach the trigger barrier at the beginning of any given payment period. The baseline level is set to be the initial spot price of the commodity when the bid is awarded. For example, options for the escalation clause are triggered if crude oil spot price exceeds a trigger barrier. These contracts are often referred to as knock-in options under trigger clauses in which holders receive a payment conditional on the underlying prices reaching a certain trigger barrier. In figure 3.2(a), when a trigger barrier, B_1, is exceeded at time t_1, the agency needs to reimburse the contractor the difference between the simulated spot prices and the initial spot prices of commodities for the work in the corresponding period. As the commodity price keeps increasing and exceeds the higher barrier levels, such as B_2 at time t_2, then barrier B_2 can be triggered according to the policy of the contract. This study only considers one barrier level in the life of project. Suppose that the loss is computed on a monthly basis. The aggregated loss to the owner when the project is finished is then shown in Equation 3.1:

$$Loss = \gamma \cdot UBP \cdot Qn \sum_{i=1}^{m=T/\Delta t} (S_i - S_0)(Qn_i - Qn_{i-1}) \cdot \exp(-rt_i) \qquad (3.1)$$

where γ= the coefficient that represents the relationship between the difference in the simulated spot price and the initial spot price of a commodity $(S_i - S_0)$ and the unit bid price, UBP, for unit bid items; Qn= total quantity of the bid item; S_i= simulated spot price of a commodity at the beginning of i^{th} month; S_0= initial spot price of a commodity when the bid is awarded; Qn_i= percentage of (total) completed quantity at the end of i^{th} month; r= risk-free interest rate; t_i= time periods for i months; T= duration of the project; and m= number of months of duration. It should be noted that the value of γ could be transformed from the coefficient, which describes the relationship between the expected change of commodity prices and the unit bid prices from the regression model.

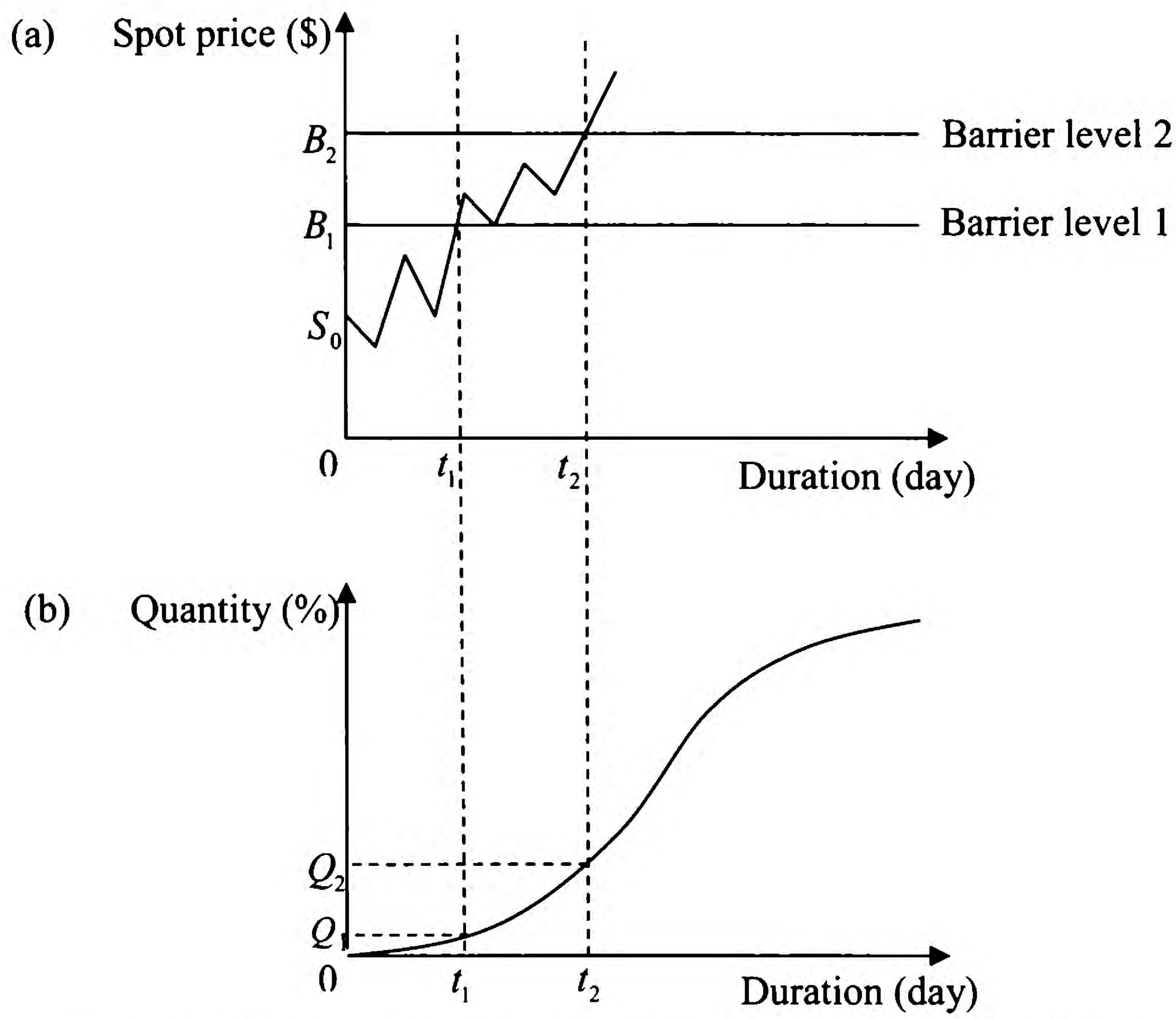


Figure 3.2 Commodity spot price and project completed quantity percentage

The corresponding work-completed quantity during each month could be obtained from actual data. For modeling purposes, it is assumed to follow an S-shaped work progress curve (Figure 3.2[b]). A convenient way to represent a project is to measure the percentage of work completed based on the cost for each activity in the project (Barraza *et al.*, 2000). In this study, the quantity of completed work for each month can be represented by the product of the percentage of project progress and the total quantity of work. More details about the escalation clauses and the relationship between the risk premiums and the barrier levels are provided and discussed in Section 4.

3.3 Objective 2 & 3 – Commodity Prices Simulations

The study models the changes in commodity prices using time series. It is known that volatile commodity prices may be correlated, thus vector time series model might be also considered to allow for correlation and co-modeling of the series. The effect of correlations between commodity prices may yield different losses when pricing risk with adjustment clauses.

The autoregressive integrated moving average (ARIMA) model and vector autoregressive moving average (VARMA) model are used for univariate and vector time series, respectively.

The ARIMA model, which is a commonly used statistical model in linear time series analysis (Box *et al.*, 1994; Montgomery *et al.*, 1998), has been historically applied to forecasting commodity prices (see Weiss, 2000; Morana, 2001; Buchananan *et al.*, 2001; Chinn *et al.*, 2001; Contreras *et al.*, 2003). This general stochastic model can be used to model commodity prices that have a long-term equilibrium level while short-term variations are due to demand or supply shocks (Guzel, 2003). Many economic time series seemed to be well described by a combination of autoregressive and moving average parameters fitted to the series itself (Kinney, 1978). An ARIMA model analyzes and predicts a value in a univariate time series as a linear combination of its own past values, past errors, and current and past values of other time series (SAS Institute Inc., 2010). When an ARIMA model includes other time series as input variables, the model is referred to as an ARIMAX model. This study does not consider other input variables, but only the past values of time series itself.

In many applications where variables are related to each other, vector time series models are applied for making use of all relevant information in forecasting. The vector generation of the ARIMA model is called the vector autoregressive moving average (VARMA) model (Montgomery *et al.*, 1998). This important extension from the univariate models to multivariate or vector AR and ARMA models (VAR; VARMA) were introduced by Quenouille (1957) and Tian and Box (1981). Analyzing and modeling the series jointly enables us to understand the dynamic relationships over time among the series and to improve the accuracy of forecasts for individual series (Clements and Hendry, 2008).

3.4 Objective 4 – Formulation of Optimization Models

Highway agencies seek optimal solutions that will not only minimize the initial project costs, but also minimize potential risk exposure during construction. Thus, the optimization problem involves two conflicting objectives: *Objective 1*—minimize the "unexpected losses" due to fluctuations in commodity prices if an escalation clause is added; and *Objective 2*—minimize the initial project cost (bidding cost), that is, minimize the risk premium.

As the optimal risk hedging problem considers two conflicting objectives (pay now in initial bid price versus pay later in the risk exposure during construction), a multi-objective optimization problem is formulated. In fact, multi-objective optimization is often applied to many engineering problems where the objectives under consideration conflict with each other. Optimizing a particular solution with respect to a single objective can result in unacceptable results with respect to the other objective. By comparison, a solution to a multi-objective problem is a set of solutions, each of which satisfies the objectives at the optimal level without being dominated by any other solution. This set of solutions, commonly used for comparing solutions in a multi-objective optimization, is known as a Pareto optimum set. Pareto sets are often preferred to single solutions, as the final decision is left to a user who can conduct trade-offs (Konak *et al.*, 2006). The corresponding set of objective values is termed as the Pareto front (Zheng *et al.*, 2004).

3.5 Data Processing and Final Data Sets

The overall data set used in this report study is shown in Figure 3.3. There are two major data sets requirements for the report - "data set 1" for simulating commodity prices, and "data set 2" for pricing unit bids and risks. Details about the two main data sets are presented in the next two subsections.

3.5.1 Data set 1 – Historical material price series

Historical material prices are obtained from Engineering News-Record (ENR) e-journal and Infor's EAM Datastream crude oil spot price data. Specifically, ENR provides "Material Cost Index" (2002-2006) data which consists of historical prices of cement, steel, and lumber. In order to keep them uniformly recorded, all the collected material prices are based on monthly data. An example time series of historical commodity prices is shown in Table 3.1.

Table 3.1 Example for prices of commodities

	ENR material prices		Datastream spot price futures price	
Date	Cement ($/ton)	Steel ($/cwt)	Brent crude oil spot price ($/barrel)	Brent crude oil 3 month future price ($/barrel)
May, 2003	83.17	25.89	25.86	26.86

3.5.2 Data set 2 – Bidding data and futures options data

There are a number of risk factors contributing to increase in bid prices. Some of them may be hard to quantify as they are not recorded or available; however, some can be collected and do relate to project risk factors. Having more variables increases the level of confidence in accepting or rejecting the hypotheses, and increases the level of significance of the estimates. The key variables in this sense are volatility and expected change in prices of construction materials.

However, data sets must be processed before they can be effectively used for modeling. The objective of the data processing is to synthesize the final data set specification for developing a model that can estimate bid prices based on two impact important variables: market-implied expected change (what market expects to occur to the prices) and volatility in commodity prices (how the prices are fluctuated).

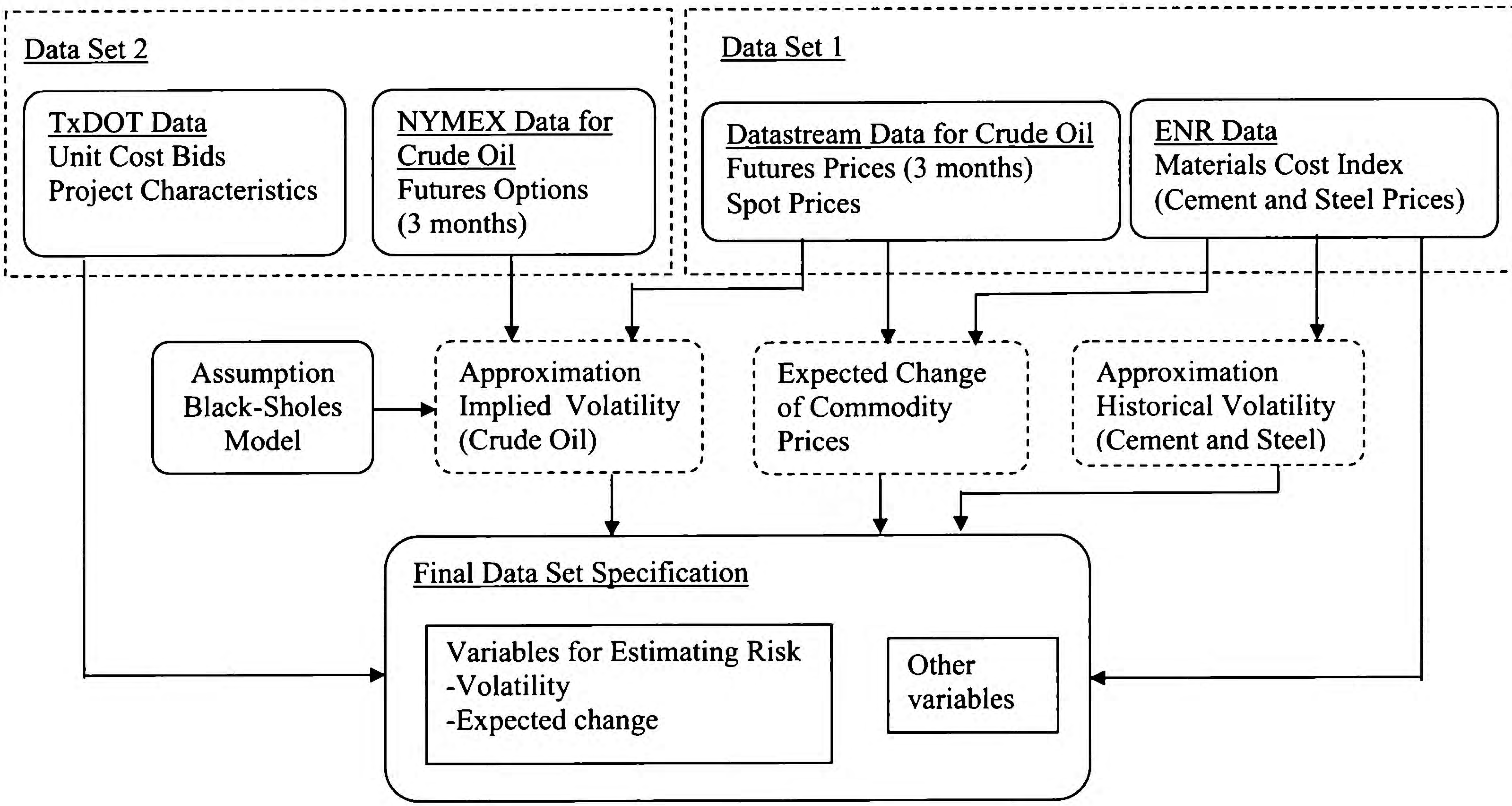


Figure 3.3 Data processing for pricing unit bids and risks

To generate needed data for pricing unit bids and risks, four distinct data sources with inflation-adjusted observations from 1998 to 2006 were pulled together: (1) Texas Department of Transportation (TxDOT) bidding data, (2) Datastream future and spot prices data for crude oil from Thomson Financial, (3) New York Mercantile Exchange (NYMEX) future options data for crude oil, and (4) Engineering News-Record (ENR) data for cement prices and steel prices.

The first data component (bid item data) contains the unit bid prices (lowest prices) for typical construction line items and other contracting parameters including quantity, project duration, total bid price, estimated total price by owner, bidding date, number of bidders, county in which the project is located, bidder, etc. The bid items data cover highway construction contracts awarded in Texas during 1998-2006. It includes four bid item categories and sixteen bid elements as defined in Highway Construction Index (HCI) developed by TxDOT (see Table A-1 in Appendix A).

The second data component containing spot prices and three month future prices of crude oil was obtained from Datastream produced by Thomson Financial and used to calculate market-implied expected change of the oil prices over three month period. The motivation for using three-month option was purely based on data set limitation. As future options are trading instruments, their price is based on supply and demand. In this context, options with shorter maturity date are more actively traded and their price is available for each trading date. This was not the case for futures option with long maturity. If consistent data for longer maturity options become available, the analysis could be replicated. This change is represented using a difference between future and spot prices.

The third data component was used to estimate market-implied volatility of oil prices at the bidding time with future options data from NYMEX covering trade date, contract year and month, settle price, and strike price. This implied volatility represents a measure of how volatile prices of crude oil will be over a time period (Christensen and Prabhala, 1998). For example, if there is absolute certainty that the prices will increase or decrease for a specific amount over a considered time period, the volatility would equal zero, no uncertainty – no volatility. Further, it is important to note that implied volatility is a measure that is implied by the traders at NYMEX: the value of the option implies the expected volatility of the price of oil over a time period, or for the case used in this study, over three month period. The methodology to calculate implied volatility from the available data will be presented in subsection 3.5.2.1. Finally, the fourth data component – ENR journal data – includes historical prices of cement and steel prices from 1997 to 2006. Historical volatilities and expected changes are accordingly estimated: the expected change of cement and steel prices represents the difference between the price in the current period and the price in the previous period; the historical volatility will be presented in subsection 3.5.2.2.

3.5.2.1 Implied volatility in crude oil prices

Market implied volatility which is embedded in the option's market price can be calculated using observed data of call options, or in the case of crude oil derivatives, future call options. Implied volatility is widely interpreted as "the market's" volatility forecast and is used in pricing options (Figlewski, 1997). A call option is a financial contract between two parties. The buyer of the option has the right, but not the obligation to buy an agreed quantity of a particular commodity from the seller of the option at a certain time (the expiration date) for a certain price (the strike price) (Stoll, 1969). The classic option pricing framework for valuing European call options on a non-dividend paying stock is developed by Black and Scholes (1973) and Merton (1973). A European option may be exercised only at the expiry date of the option, whereas an American option may be exercised at any time before the expiry date.

A useful property of the Black-Scholes option pricing model, shown in Equation 3.2, is that all the parameters except the volatility are directly observable from the market. Given the current price of a specific option contract along with the model's other parameters, the model can be solved backwards for the value of the volatility parameter implied by the current price of the option which is implied volatility of the option.

$$C = SN(d_1) - Xe^{-rT}N(d_2) \tag{3.2}$$

$$d_1 = \frac{\ln(S/X) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \tag{3.2a}$$

$$d_2 = \frac{\ln(S/X) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T} \tag{3.2b}$$

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du \tag{3.2c}$$

The value of a European call option, stock price, strike price, risk-free interest rate, time to expiration and volatility are denoted by C, S, X, r, T, σ, respectively. $N(x)$ is the cumulative normal density function.

The widely used Black-Scholes option pricing model implies that underlying stochastic process for oil prices is geometric Brownian motion, a typical representation of a process in financial economics (Bernabe *et al.*, 2004; Postali and Picchetti, 2006; Al-Harthy, 2007). A number of models were developed to represent uncertainty in oil prices including mean-reversion (Bessembinder *et al.*, 1995; Baker *et al.*, 1998) and mean-reversion with jumps (Pelet, 2003). While there is an active debate to what models can best fit the reality, Damnjanovic and Zhou (2009) adopted a standard geometric Brownian motion to specify the stochastic behavior of oil prices (the principle of parsimony is used as a guiding criterion); here, the drift rate tells the trend or the mean change in price of oil, while volatility reflects the variance of the price in period *T*. However, it is important to note that other models can be adopted to represent uncertainty in prices.

Even though there is no closed-form solution for implied volatility from the Black-Sholes model, a number of approximated solutions are reported in the literature. An approximation developed by Corrado and Miller (1996) - Equation 3.3 - is validated when stock prices deviate from discounted strike prices as discussed by Li (2005). Li also developed approximate solutions for implied volatility and provided a uniform framework for deriving the approximations. The approximation models include classification based on a number of factors such as classification deep in- or out-of-the-money calls.

$$\sigma \approx \sqrt{\frac{2\pi}{T}} \frac{1}{S+K} \left[C - \frac{S-K}{2} + \sqrt{\left(C - \frac{S-K}{2} \right)^2 - \frac{(S-K)^2}{\pi}} \right] \tag{3.3}$$

where K is the discounted strike price and the other variables are the same as used in Equation 3.2.

The three month implied volatility is calculated using Equation 3.3, as the data support the assumptions used in the approximation. The approximation results are shown in Figure B-1 in Appendix B. This approximation yields implied volatility values that are nearly identical to the values from the actual Black-Scholes model when numerically solved. Please see details in Appendix B for formulas used in computing implied volatility.

While implied volatility is a forward-looking measure of likely future volatility conditions, historical volatility is a backward-looking measure of recent volatility conditions (Kawaller *et al.*, 1994). Since the data for options and futures for cement and steel prices are not available, historical volatility is calculated and used instead of implied volatility for cement and steel prices.

3.5.2.2 Historical volatility for cement and steel prices

In theoretical option pricing models, Black and Scholes derived their option valuation equation under the assumption that stock returns, "log price relatives", followed a logarithmic diffusion process in continuous time with a constant drift and volatility parameters, as shown in Equation 3.4:

$$\frac{dS}{S} = \mu dt + \sigma dz \qquad (3.4)$$

where dS/S is the instantaneous proportional change in the price of the underlying asset, μ is the annual mean return, σ is the volatility, dt indicates an infinitesimal unit of time and dz represents Brownian motion. Starting from an initial value S_0, the return over the period from 0 to T is given by Equation 3.5:

$$R = \ln\left(S_T/S_0\right) \qquad (3.5)$$

and R has a Normal distribution, with

$$\begin{aligned} Mean &= \left(\mu - \sigma^2/2\right)T \\ Standard\ deviation &= \sigma\sqrt{T} \end{aligned} \qquad (3.6)$$

The standard historical volatility estimate (the most basic method) follows the following procedure. Historical volatility is typically computed as the standard deviation of the percentage price changes over a recent period (Kawaller *et al.*, 1994). Consider a set of historical prices that follow the process defined in Equation 3.4: $\{S_0, S_1, \ldots, S_T\}$. Then the log price relatives are computed, i.e., the percentage price changes expressed as continuously compounded rates:

$$R_t = \ln\left(S_t/S_{t-1}\right), \text{ for } t \text{ from 1 to } T \qquad (3.7)$$

The estimate of the (constant) mean μ of the R_t is the simple average:

$$\bar{R} = \frac{\sum R_t}{T} \tag{3.8}$$

The variance of the R_t is given by Equation 3.9:

$$v^2 = \frac{\sum \left(R_t - \bar{R}\right)^2}{(T-1)} \tag{3.9}$$

Annualizing the variance by multiplying by N, the number of price observations in a year and taking the square root yields the volatility,

$$\sigma = \sqrt{Nv^2} \tag{3.10}$$

If the constant parameter diffusion model of Equation 3.4 is correct, the above procedure gives the best estimate of the volatility that can be obtained from the available price data. This estimate then becomes the forecast for the volatility going forward, over a time horizon of any length. Simply projecting observed past volatility into the future is a common way to make a forecast, but it is only one of several common methods, and need not be the most accurate (Figlewski, 1997). The calculated historical volatilities for cement and steel prices are shown in Appendix B.

3.5.2.3 Variables for estimating risk premiums

As previously mentioned, two characteristic variables are considered for estimating risk premiums: (1) the market-implied expected change of commodity prices (the difference between three-month futures price and spot price for crude oil, or the difference between current price and the price in three months lags of cement and steel); and (2) duration-based volatility (three-month duration-based implied volatility for crude oil prices, or duration-based historical volatility for cement and steel prices). The *duration-based volatility* represents a transformed testing variable defined as a product of contract duration and market-implied volatility. The motivation for specifying this "synthetic" variable lies in the fact that contract length affects the risk: The longer the contract duration, the more significant volatility becomes. For example, if the contract duration is short, even high volatility should not affect the prices as the contractor has fuel/materials available before the contract begins. On the other hand, the impact of volatility can be significant even for modest volatility measures for long-duration contracts (Damnjanovic and Zhou, 2009).

3.5.2.4 Final specification of data set for pricing unit bids and risk

The unit bid price is considered as a dependent variable while total quantity, number of bidders, bidding date (in quarters of year), project location (city location or not), districts (a vector of districts in Texas), duration of project, volatility and expected change of prices of commodities (cement, steel and oil) are considered to be explanatory variables.

3.6 Summary

This section presents the overall methodology and data sets for developing the models in this report. Three main model parts are explained: (1) a model to price the unit bid items and the risks, including the contract design using escalation clauses with barrier levels; (2) the time series models used for simulating commodity prices; and (3) the multi-objective optimization model. Data processing and two final data sets are also presented. Data set 2 (TxDOT bidding data) is used for unit bid price regression model and estimating "build in" premiums. Data set 1 (monthly historical material prices) is used for identifying expected change and volatility of uncertain commodity prices, thus for simulating time series prices and for contracting optimizations. In the following section, the models for pricing unit bids and risks are presented.

4. PRICING UNIT BIDS AND AVERAGE RISK PREMIUMS

The highway construction cost showed a significant increase from 2003 to 2008. For example, compared to 1997, the purchasing power of the Texas Department of Transportation (TxDOT) has decreased considerably. A construction project in 2006 is valued two times more than a similar project in 1997 (Pandit *et al.*, 2009). As the cost of materials and oil-based fuels has significantly increased over the same time period, it is evident that there is a direct relationship between the cost of construction and commodity prices (Gallagher and Riggs, 2006). In the mid-2000s, a number of highway contractors were influenced by escalating material prices (the cost of liquid asphalt, cement, fuel, and steel). In such settings, where the commodity market is volatile, the contractors are cautious when bidding on the jobs without adjustment clauses. As a result, the bids include even larger contingencies amplifying the volatility effect to the construction market.

To determine optimal hedging of commodity risk using escalation clauses, it is essential to first determine the price of risk in these contracts. This section presents a link between the average risk premium and the expected change and volatility of commodity prices. The average risk premium here is the average amount of money that highway agencies pay to the contractors given the historical bidding information of highway projects. This section aims at explaining the relationship between the unit bid prices of selected control items and the risk factors (e.g., increase in commodity prices); in other words, it aims to estimate the impact that volatile commodity prices have on the unit bid prices (i.e., risk premiums).

The increase in cost of highway construction is largely a result of escalation in the cost of commodities, such as crude oil, steel and cement (Gallagher and Riggs, 2006; Mendell, 2006; GNB, 2007; FHWA, 2007). Highway construction depends on many products that are derived from crude oil. This implies that potential fuel price escalation can cause contractors to submit higher bid prices not only for materials derived from crude oil, but also for the work items that require significant use of fuel to power machinery (McFall, 2005; GNB, 2007). In addition to oil, price of steel significantly impacts the cost of construction. During 2004-2006, steel price rose significantly; sharp hikes to steel prices made concrete structures more competitive thus increasing the demand and the price of cement (McGoldrick, 2006).

In this report, oil, steel and cement prices are viewed as risk factors that impact unit bid prices of bid items. According to the result of a survey conducted by the American Association of State Highway and Transportation Officials (AASHTO) in 2008, forty states out of fifty-two member departments used fuel price adjustment clauses; forty-two states used asphalt cement price adjustment clauses; fifteen states used steel price adjustment clauses; and three states used Portland cement price adjustment clauses (AASHTO, 2008a).

As mentioned in Section 3, based on the available data, the unit bid price is considered as a response variable while total quantity, number of bidders, bidding date (in quarters of year), project location (city location or not), districts (a vector of districts in Texas), duration of project, volatility and expected change of commodity prices are considered to be explanatory variables. Explanatory variables are typically the variables representing the variation in the response variable. The final specification of data set used for pricing unit bids and risks is shown in Table 4.1.

Table 4.1 Variables for pricing unit bids and risks

No.	Explanatory variable	Unit	Symbol
1	Quantity	C.Y.; L.F.	*Qn*
2	Number of bidders	count	*N*
3	Bidding date (quarter of year)	dummy	*T*
4	Location (district)	dummy	*L*
5	City or not	dummy	*C*
6	Duration-based volatility	day×volatility	*VD* (*V* for volatility; *D* for duration)
7	Expected change	$/bbl;$/ton;$/cwt	*oilFS*; *cementFS*; *steelFS*
8	Unit bid price (response variable)	$/quantity	*UBP*

4.1 Investigated Bid Items and Identified Risk Factors

Most projects contain a small number of work items that together comprise a significant portion (e.g. 75 percent) of the total cost. In highway construction, the major items are typically Portland cement concrete pavement, structural concrete, structural steel, asphalt concrete pavement, and embankment (FHWA, 2004). Federal Highway Administrations' composite bid price index (FHWA CBPI), a popular construction cost index in the highway industry, reflects these items. The FHWA CBPI is composed of six indicator items: common excavation, to indicate the price trend for all roadway excavation; Portland cement concrete pavement and bituminous concrete pavement, to indicate the price trend for all surfacing types; and reinforcing steel, structural steel, and structural concrete, to indicate the price trend for structures (FHWA, 2006). These items were identified by first observing which construction sections experience major expenditure, and then identifying the largest expenditure pay item expressed in unit costs in each section (Wilmot and Mei, 2005; Cheng and Wilmot, 2008). Further, a number of states track project costs for a number of selected bid items. For example, WSDOT tracks the cost of construction materials of seven typical construction bid items—crushed surfacing, concrete pavement, structural concrete, hot mix asphalt, roadway excavation, steel reinforcing bar and structural steel (WSDOT, 2011).

Based on the identified significant impact of the items on the cost of projects, seven control items shown in Table 4.2 are selected from a set of thirty-four control items used in the TxDOT's HCI (Appendix A). These control items are (1) roadway excavation; (2) roadway embankment; (3) flexible base; (4) hot mix asphaltic concrete; (5) continuous reinforced concrete pavement; (6) regular beams; and (7) retaining walls. For these seven control items, this report considers the impact of three commodity risk factors (cement, steel, and oil).

Table 4.2 Selected items and risk factors

Model	Selected control item	Identified risk factor
1	Roadway excavation	Oil price
2	Roadway embankment	Oil price
3	Flexible base	Oil price
4	Hot mix asphaltic concrete	Oil price
5	Regular beams	Steel price
6	Continuous reinforced concrete pavement	Cement price
7	Retaining walls	Cement price

The method for matching the most significant risk factor to each control item is based on several steps. First, it considers the description of the each item by the AASHTO (2008b), as it is important to determine the contributing materials to each item. Second, if a control item uses more than one type of material, one can use correlation coefficients between the unit bid price and the commodity prices as a criterion to determine the most significant risk factor.

Risk factors identified for each item are shown in Table 4.2. The justification, for considering the seven control items and the corresponding risk factors, is presented as below:

Control items: Excavation and embankment

It is well known that the cost of highway construction is affected by the cost of crude oil. While this relationship is highly visible for construction items such as asphalt cement (a byproduct in the process of refining oil), the effects of the crude oil prices on the cost of other construction items, such as concrete cement or construction operations, are less direct but equally important (Damnjanovic and Zhou, 2009).

Excavation and embankment are identified among the most fuel intensive types of work, and therefore are susceptible to the changes in fuel prices (FHWA, 1980; FHWA, 2007; Damnjanovic and Zhou, 2009). Excavation and embankment which include excavating, hauling, placing, disposing and compacting materials require considerable usage of fuel. Thus, the costs of oil-based fuels significantly impact the costs of excavation and embankment (FHWA, 2007; Damnjanovic and Zhou, 2009). Increase in the cost of excavation and embankment has been dramatic since 2003. For example, the percentage of change in the cost of excavation in Texas during that period was estimated to be 47.2%.

Control item: Flexible base

The cost of transporting granular materials for flexible bases represents a major factor in pavement construction costs. Pavement construction requires transportation of significant amount of aggregates to the construction site (Nash *et al.*, 1995). In other words, it requires the use of equipment (e.g., trucks, rollers) for delivering base materials to the roadway, for hauling and placing flexible base, for compaction and rolling, and other processes. (Note that the cost of aggregate materials is also a major factor contributing the unit bid price of flexible base; however, due to data availability, crude oil price was a sole risk factor considered for flexible base control item.)

Control item: Hot mix asphaltic concrete (HMAC)

The increasing cost of crude oil has a direct effect on the cost of hot mix asphaltic concrete bid item. Asphalt prices are closely related to crude oil prices. As asphalt is a product derived from crude oil and HMAC work item deals with mixing, and placing asphalt concrete using fuel intensive machines, the crude oil price is considered as the most significant risk factor for HMAC. According to the state of Washington (FHWA, 2007), the costs of hot mix asphalt grew by 64.2 percent from 2003 to 2006, compared to 9.8 percent from 2000 to 2003. This to a large degree mimics the cost of crude oil.

Control item: Continuous reinforced concrete pavement (CRCP)

Concrete pavement control item is one of the items that affect the highway construction cost. Steel and cement prices affect the cost of continuous reinforced concrete pavement as they represent the major material requirements. In this report, the cement price is chosen as the most significant risk factor contributing to the cost of continuous reinforced concrete pavement control item. To identify which commodity price has the highest impact on this control item, a correlation analysis is conducted. The result is shown in Table 4.3.

Table 4.3 Correlations between unit bid prices of CRCP and commodity prices

	D	*Qn*	*N*	*cementFS*	*oilFS*	*steelFS*
UBP	-0.446	-0.682	-0.163	**0.127**	0.055	0.082

The coefficient of correlation between the unit bid price of CRCP and the price of cement is 0.127 which shows that cement price is more correlated to the unit bid price of CRCP than both prices of crude oil and steel. In addition, contract duration, quantity and the number of bidders show negative correlations to the unit bid price of CRCP. Therefore, the price of cement is selected to be the risk factor for this control item.

Control items: Retaining walls

Retaining walls are considered as a part of concrete structures. Cement price is identified as the risk factor for retaining walls. The retaining walls control item deals with furnishing, constructing, and installing retaining walls. The required materials for this item are mainly aggregates and concrete cement.

Control items: Regular beams

Regular beams are considered as a part of steel structures. Table 4.4 shows that both steel price and cement price have considerable influence on the unit bid price (i.e., the corresponding correlation coefficients are 0.24 and 0.24, respectively). As cement price was selected for CRCP item, the price of steel is considered as the risk factor for regular beams item.

Table 4.4 Correlations between unit bid prices of regular beams and commodity prices

	D	*Qn*	*N*	*steelFS*	*oilFS*	*cementFS*
UBP	-0.124	-0.266	-0.198	**0.24**	0.035	0.24

4.2 Average Risk Premiums Estimation

4.2.1 Regression model

This subsection presents development of regression models to determine the impact of risk factors on the unit bid prices. Based on the identified risk factor for each bid item, regression models are developed to price the unit bids and risk premiums. The models include project characteristics as explanatory variables in addition to commodity price variables used to estimate the risk premium.

Multiple linear regression models in this report are based on weighed least square (WLS) estimation. The main advantage of WLS is its ability to deal with non-constant error variance (Sheather, 2008). The basis of WLS implementation is illustrated as follows. Consider the straight linear regression model in Equation 4.1:

$$Y_i = \beta_0 + \beta_1 x_i + e_i \tag{4.1}$$

where the e_i has a mean 0 but the variance of σ^2 / w_i, the weights w_i need to be taken into account when estimating the regression parameters β_0 and β_1. This is achieved by considering the weighted version of the residual sum of squares in Equation 4.2:

$$WRSS = \sum_{i=1}^{n} w_i \left(y_i - \hat{y}_{wi}\right)^2 = \sum_{i=1}^{n} w_i \left(y_i - b_0 - b_i x_i\right)^2 \tag{4.2}$$

The values of b_0 and b_1 that minimize $WRSS$ are required to obtain the WLS estimates.

The explanatory variables used in this regression model include project characteristics (quantity, number of bidders, bidding date, and location), and risk factor characteristics (duration-based volatility and expected change of commodity price). The response variable in the model is the unit bid price. The two risk factor variables — duration-based volatility and expected change of commodity price — are added to estimate the risk premium. Let UBP_R represent the regressed unit bid price given duration-based volatility (VD) and the expected (average) change of commodity price (FS) from market data sets, while UBP_0 represent the regressed unit bid price where the change and the volatility of commodity prices are equal to zero, $V \approx 0$, and $FS = 0$ (this corresponds to the case where the prices are fixed and no premium is needed); then, the risk premium hidden in the unit bid price is estimated as $\Delta UBP = UBP_R - UBP_0$. Thus, the total risk premium is $R = \Delta UBP \times Qn$, where Qn is the total quantity of bid item. It is important to note that the following regression model is not intended to predict future prices of the bid unit items, but to estimate average risk premium hidden in the bid price. Due to lack of data, the model ignores possible price manipulation due to unbalancing bids and other important explanatory variables such as contractor's size, construction market condition, and others.

4.2.2 Steps for building model and validation

It is well known that a key step in any regression analysis is assessing the validity of the given model. The steps for developing and validating models follow the process presented below (Sheather, 2008):

Step 1: Generate scatter plots, box plots and normal QQ plots to assess the distributions of variables

Scatter plots show the general pattern between the two variables; Box plots and normal QQ plots can be used to further see the shape of the distribution of variables. They are useful in the early stages of the analysis when exploring data before actually calculating a correlation coefficient or fitting a regression curve.

Step 2: Transform data using Box-Cox methods if the variables show non-normality and/or non-constant variance

When data needs to be transformed to address the problems such as nonnormality and non-constant variance, Box-Cox method is used to overcome these problems by identifying the most appropriate transformation. Box and Cox (1964) provided a general method for transforming strictly positive variables. The Box-Cox procedure aims to find a transformation that makes the transformed variable close to normally distributed. Admittedly, transformations do not perform well in every situation especially when important predictors which interact with other are not included in the model.

Step 3: Re-check step 1

This step aims to check the effect of transformation on the relationships between the new variables, that is, to see whether the strength of the linear relationship between the new variable is better than the linear relationship between the original un-transformed ones.

Step 4: Select variables based on backward method

The "best" model with a set of predictors should be chosen from a class of multiple regression models using a variable selection method. In general, the more predictor variables included in a model, the lower the bias of the predictions, but the higher the variance. Including too many predictors in a regression model is commonly referred to as over-fitting (Sheather, 2008). Backward elimination is used for the variable selection in this report. The starting point is the model with all of the available variables. The partial-F tests are then used to determine if the "worst" variable at each testing stage can be deleted.

Step 5: Run regression model and estimate the parameters

With all the final selected variables from the procedure of backward elimination included in the model, the output associated with fitting model shows all the coefficients for the variables and their significance. In other words, this step estimates the model parameters.

Step 6: Model validation and regression diagnostics

A comprehensive set of tests is used to validate regression models. These tests include plots of standardized residuals against each predictor, plots of fitted values, diagnostic plots, marginal model plots, added variable plots, and variance inflation factor (VIF) test. 1) A crucial assumption in any regression analysis is that the errors have constant variance. The regression models are considered valid when there is a random pattern from plots of the standardized residuals against each predictor, a straight line fit to fitted values of dependent variable, and normality of the errors from diagnostic plots (Sheather, 2008). 2) Marginal plots have a wider application than residual plots while added variable plots are used to assess the effect of each predictor variable on the response variable having adjusted for the effect of other predictor variables. 3) Validation of regression model also depends on no multicollinearity where VIF for each predictor should be less than five. When highly correlated predictor variables are included in a regression model, they are effectively carrying very similar information about the response variable. Then, it is difficult for least squares to distinguish their separate effects on the response variable.

The process for multiple linear regression models developed in this study is shown, explained, and discussed in Appendix C where the bid item "excavation" is considered as an example. The same process is then repeated for the other six bid items.

4.2.3 Results

4.2.3.1 Item 1 – roadway excavation

The final model specification and parameter estimation results are shown in Equation 4.3 and Table 4.5, respectively. All the variables are statistically significant and show the expected signs (See the regression output 1 from statistical software R – Item 1 roadway excavation in Appendix C.).

$$UBP = \beta_0 \cdot (V \cdot D)^{\beta_1} \cdot Qn^{\beta_2} \cdot N^{\beta_3} \cdot \beta_4^{oilFS} \cdot \beta_5^{T_1} \cdot \beta_6^{T_2} \cdot \prod_{j=1}^{14} \beta_{j+6}^{L_j} \qquad (4.3)$$

Table 4.5 Model estimation value - excavation

Parameter	Estimated parameter value for excavation	Parameter	Estimated parameter value for excavation
β_0	31.28	$\beta_{11}(d7)$	0.773
β_1	0.022	$\beta_{12}(d9)$	0.766
β_2	-0.218	$\beta_{13}(d10)$	1.119
β_3	-0.226	$\beta_{14}(d11)$	1.231
β_4	1.088	$\beta_{15}(d12)$	1.139
β_5	0.945	$\beta_{16}(d16)$	0.79
β_6	0.911	$\beta_{17}(d18)$	0.787
$\beta_7(d1)$	0.871	$\beta_{18}(d19)$	0.577
$\beta_8(d3)$	0.799	$\beta_{19}(d21)$	1.26
$\beta_9(d4)$	1.27	$\beta_{20}(d22)$	0.846
$\beta_{10}(d5)$	1.147		

The result of the regression analysis in Table 4.5 shows that the data support the claim that the market-implied change (β_4) and the volatility of prices (β_1) in crude oil market affect the unit prices of excavation for contracts without adjustment clauses (because the coefficients of these two variables are greater than zero). The model indicates that every 1 percent increase in duration-based implied volatility and the expected change (a difference between futures and spot prices of crude oil), will result in (on average) a 0.022 percent and 8.4 percent increase in unit bid prices of excavation, respectively.

In addition, Table 4.5 indicates the effect of project characteristics variables: every 1 percent increase in quantity and number of bidders of projects leads to a 0.218 percent decrease and a 0.226 percent decrease respectively in the unit bid prices of excavation, which demonstrate that the larger quantity and the number of bidders reduce the unit bid prices; Quarters in which projects are let and the location of projects also affect the unit bid price of excavation. The unit bid prices in the quarter 1 of year are estimated to be approximately 5.7 percent lower than those in quarter 3 and quarter 4; the unit bid prices in quarter 2 is estimated to be approximately 9.3 percent lower than those in quarter 3 and quarter 4; Project location (i.e., Texas districts as discussed in Section 3), has a significant impact on the unit bid price of excavation. The unit bid price in d1, d3, d7, d9, d16, d18, d19 and d22 are estimated by the model to be 13.8 percent, 22.5 percent, 25.7 percent, 26.7 percent, 23.6 percent, 24 percent, 55 percent and 16.7 percent respectively lower than those in the remaining districts in Texas; the unit bid prices in d4, d5, d10, d11, d12 and d21 are estimated by the model to be 23.9 percent, 13.7 percent, 11.2 percent, 20.8 percent, 13 percent and 23.1 percent respectively higher than those in the remaining districts. (The twenty-five districts in Texas are shown in Table 4.6. The project location categorized in districts can also reflect the influence by city or rural.)

Table 4.6 Twenty-five districts in Texas

Variable	District	Variable	District	Variable	District
d1	Abilene	d10	Dallas	d19	Pharr
d2	Amarillo	d11	El Paso	d20	San Angelo
d3	Atlanta	d12	Fort Worth	d21	San Antonio
d4	Austin	d13	Houston	d22	Tyler
d5	Beaumont	d14	Laredo	d23	Waco
d6	Brownwood	d15	Lubbock	d24	Wichita Falls
d7	Bryan	d16	Lufkin	d25	Yoakum
d8	Childress	d17	Odessa		
d9	Corpus Christi	d18	Paris		

Considering the model estimation in Table 4.5 and an example contract structure of excavation item let in 2004 in Lubbock (Texas) shown in Table 4.6, the total risk premium without adjustment clauses can be estimated using Equation 4.3. UBP_R, the regressed unit bid price given input values of both duration-based implied volatility VD and expected change of crude oil prices $oilFS$, is 1.384 (\$/CY); UBP_0, the regressed unit bid price given $V = 0.0005 \approx 0$ and $oilFS = 0$, is 0.995 (\$/CY). V is set to be approximately zero because implied volatility of crude oil prices will never be zero. Thus the risk premium in unit bid price accounting for expected change and volatility of crude oil price is $\Delta UBP = UBP_R - UBP_0 = \$0.389 / CY$ and then total risk premium without any adjustment is $R = \Delta UBP \times Quantity = \$613,248$ (13 percent of total bid price).

Table 4.7 Parameters for an excavation bid

T	L	V	D (day)	Qn (C.Y.)	N	$oilFS$ (\$/bbl)
12/07/04	Lubbock (d15)	0.5018	860	1576453	4	2.11

4.2.3.2 Item 2 – roadway embankment

The final model specification for this control item is shown in Equation 4.4 and Table 4.7. All the variables are statistically significant and of the expected signs (See the regression output 2 from statistical software R – Item 2 roadway embankment in Appendix C). The result of regression analysis shows that the data support the claim that market-implied change of prices in crude oil market affect the unit price of embankment for contracts without adjustment clauses. The model indicates that every 1 percent increase in the expected change (a difference between futures and spot prices of crude oil), will result in (on average) a 9.43 percent increase in unit bid prices of embankment.

$$UBP = \beta_0 \cdot Qn^{\beta_1} \cdot N^{\beta_2} \cdot \beta_3^{oilFS} \cdot \prod_{j=1}^{7} \beta_{j+3}^{L_j} \quad (4.4)$$

Table 4.8 Model estimation value - embankment

Parameter	Estimated parameter value for excavation	Parameter	Estimated parameter value for excavation
β_0	31.218	$\beta_6(d10)$	1.182
β_1	-0.215	$\beta_7(d11)$	1.315
β_2	-0.158	$\beta_8(d19)$	0.533
β_3	1.099	$\beta_9(d21)$	0.87
$\beta_4(d4)$	1.16	$\beta_{10}(d24)$	1.288
$\beta_5(d5)$	1.323		

Considering the model estimation in Table 4.7 and an example contract structure of embankment item let in 2004 in Lubbock (Texas) as shown in Table 4.8, the total risk premium without adjustment clauses can be estimated using Equation 4.4. UBP_R, the regressed unit bid price given input values of the expected change of oil price $oilFS$, is 1.608 (\$/CY); UBP_0, the regressed unit bid price given $oilFS = 0$, is 1.317 (\$/CY). Thus the risk premium in unit bid price accounting for expected change of steel price is $\Delta UBP = UBP_R - UBP_0$= 0.29 (\$/CY) and then total risk premium without any adjustment is $R = \Delta UBP \times Quantity = \$259{,}598$ (10 percent of original bid price).

Table 4.9 Parameters for an embankment bid

T	L	D (day)	Qn (C.Y.)	N	$oilFS$ ($/bbl)
12/07/04	Lubbock(d15)	860	893,984	4	2.11

When compared to excavation item, the effect of oil price (risk factor) on the unit bid prices of embankment is less. However, for project with large quantity, the effect is still substantial as shown in the example above where the risk premium is estimated to be $259,598. Thus, an escalation clause should be considered to assess the benefits of taking on the future risk while getting a lower risk premium for the bids.

4.2.3.3 Item 3 – flexible base

The final model specification for this control item is shown in Equation 4.5 and Table 4.9. All the variables are statistically significant and of the correct expected signs (See the regression output 3 from statistical software R – Item 3 flexible base in Appendix C). The result of regression analysis shows that the market-implied change in crude oil prices affects the unit prices of flexible base for contracts without adjustment clauses at a 0.1 significance level. Although the quantities of projects and numbers of bidders affect the unit bid prices of flexible base more significantly (i.e., p-value is smaller), the impact (coefficients of estimates) of these two parameters is less than the effect of crude oil price on the unit bid price. The model indicates that every 1 percent increase in the expected change (a difference between futures and spot prices of crude oil), will result in (on average) a 7.85 percent increase.

$$UBP = \beta_0 \cdot Qn^{\beta_1} \cdot N^{\beta_2} \cdot \beta_3^{oilFS} \cdot \beta_4^{T_1} \cdot \beta_5^{T_2} \cdot \prod_{j=1}^{11} \beta_{j+5}^{L_j} \qquad (4.5)$$

Table 4.10 Model estimation value - flexible base

Parameter	Estimated parameter value for excavation	Parameter	Estimated parameter value for excavation
β_0	52.353	$\beta_9(d10)$	1.24
β_1	-0.131	$\beta_{10}(d13)$	1.309
β_2	-0.097	$\beta_{11}(d16)$	1.209
β_3	1.081	$\beta_{12}(d18)$	1.165
β_4	0.929	$\beta_{13}(d19)$	0.651
β_5	0.947	$\beta_{14}(d20)$	0.783
$\beta_6(d5)$	1.694	$\beta_{15}(d22)$	1.256
$\beta_7(d7)$	1.164	$\beta_{16}(d24)$	1.184
$\beta_8(d9)$	1.262		

Considering the model estimation in Table 4.9 and an example contract structure of flexible base item let in 2005 as shown in Table 4.10, the total risk premium without adjustment clauses can be estimated using Equation 4.5. UBP_R, the regressed unit bid price given input values of expected change of crude oil price $oilFS$, is 13.52 ($/CY); UBP_0, the regressed unit bid price given $FS = 0$, is 12.2 ($/CY). Thus the risk premium in unit bid price accounting for expected change of price is $\Delta UBP = UBP_R - UBP_0 = 1.32\,(\$/\mathrm{CY})$ and then total risk premium without any adjustment is $R = \Delta UBP \times Quantity = \$26{,}116$ (6.8 percent of original bid price).

Table 4.11 Parameters for a flexible base bid

T	*L*	*D* (day)	*Qn* (C.Y.)	*N*	*oilFS* ($/bbl)
05/11/05	Cass (d3)	196	19,776	3	1.32

Compared to the effect of crude oil price on excavation and embankment, the effect on flexible base is not that significant. This might be due to the factor that was not considered in the model – the cost of aggregate material used for the flexible base item.

4.2.3.4 Item 4 - hot mix asphaltic concrete (HMAC)

The final model specification for this control item is shown in Equation 4.6 and Table 4.11. All the variables are statistically significant and of the correct expected signs (See the regression output 4 from statistical software R – Item 4 HMAC in Appendix C.). The result of regression analysis shows that the market-implied change in crude oil prices affects the unit price of HMAC for contracts without escalation clause significantly. The model indicates that every 1 percent increase in the expected change (a difference between futures and spot prices of crude oil), will result in (on average) a 9.37 percent increase in unit bid prices of HMAC.

$$UBP = \beta_0 \cdot Qn^{\beta_1} \cdot N^{\beta_2} \cdot \beta_3^{oilFS} \cdot \prod_{j=1}^{11} \beta_{j+3}^{L_j} \quad (4.6)$$

Table 4.12 Model estimation value - HMAC

Parameter	Estimated parameter value for excavation	Parameter	Estimated parameter value for excavation
β_0	81.78	$\beta_8(d8)$	1.251
β_1	-0.124	$\beta_9(d11)$	1.11
β_2	-0.109	$\beta_{10}(d14)$	1.112
β_3	1.099	$\beta_{11}(d15)$	1.3
$\beta_4(d2)$	1.146	$\beta_{12}(d16)$	1.148
$\beta_5(d4)$	0.891	$\beta_{13}(d19)$	0.803
$\beta_6(d5)$	1.141	$\beta_{14}(d22)$	1.053
$\beta_7(d6)$	1.122		

Considering the model estimation in Table 4.11 and an example contract structure of HMAC item let in 2004 in Lubbock (Texas) as shown in Table 4.12, the total risk premium without adjustment clauses can be estimated using Equation 4.6. UBP_R, the regressed unit bid price given input values of expected change of oil price $oilFS$, is 25.93 ($/CY); UBP_0, the regressed unit bid price given $oilFS = 0$, is 21.24 ($/CY). Thus the risk premium in unit bid price accounting for expected change of steel price is $\Delta UBP = UBP_R - UBP_0$= 4.68 ($/CY) and then total risk premium without any adjustment is $R = \Delta UBP \times Quantity = \$604,387$ (22 percent of original bid price).

Table 4.13 Parameters for an HMAC bid

T	*L*	*Qn* (C.Y.)	*N*	*oilFS* ($/bbl)
12/07/04	Lubbock (d15)	129,083	4	2.11

The estimated risk premium indicates that the effect of crude oil price on the unit price of HMAC (which is directly related to asphalt price) is more significant than the effect on the unit bid price of excavation or flexible base. Thus, an escalation clause of this control item should be considered to assess the benefits of taking on the future risk while getting a lower risk premium for the bids (it is noted that the adjustment clause should be for asphalt price if the asphalt prices are available and show to be the most correlated risk factor with the unit bid price of HMAC).

4.2.3.5 Item 5 - Regular beams

The model specification for this control item is shown in Equation 4.7 and Table 4.13. All the variables are statistically significant and of the correct expected signs (See the regression output 5 from statistical software R – Item 5 regular beams in Appendix C.). The result of regression analysis shows that the data supports the claim that expected change (the observation of change in current and previous period) and historical volatility of steel prices affect the unit prices of regular beams for contracts without escalation clauses. The model indicates that every 1 percent increase in duration-based historical volatility and the expected change in steel price, will result in (on average) a 0.114 percent and 9.8 percent increase in unit bid prices of regular beams, respectively.

$$UBP = \beta_0 \cdot (V \cdot D)^{\beta_1} \cdot Qn^{\beta_2} \cdot N^{\beta_3} \cdot \beta_4^{steelFS} \cdot \prod_{j=1}^{3} \beta_{j+4}^{L_j} \qquad (4.7)$$

Table 4.14 Model estimation value - regular beams

Parameter	Estimated parameter value for excavation	Parameter	Estimated parameter value for excavation
β_0	83.346	β_4	1.103
β_1	0.114	$\beta_5(d2)$	1.188
β_2	-0.1	$\beta_6(d15)$	1.465
β_3	-0.126	$\beta_6(d24)$	1.168

Considering the model estimation in Table 4.13 and an example contract structure of regular beams item let in 2004 in Lubbock (Texas) as shown in Table 4.14, the total risk premium without adjustment clauses can be estimated using Equation 4.7. UBP_R, the regressed unit bid price given input values of duration-based historical volatility VD and expected change of steel price $steelFS$, is 64.66 (\$/LF); UBP_0, the regressed unit bid price given $V = 0.0005 \approx 0$ and $steelFS = 0$, is 32.98 (\$/LF). Thus the risk premium in unit bid price accounting for the expected change of steel price is $\Delta UBP = UBP_R - UBP_0$=31.68 (\$/LF) and then the total risk premium without adjustment clause is $R = \Delta UBP \times Quantity = \$780611$ (30.2 percent of original bid price).

Table 4.15 Parameters for a regular beams bid

T	*L*	*V*	*D* (day)	*Qn* (L.F.)	*N*	*steelFS* (\$/cwt)
12/07/04	Lubbock (d15)	0.1095	680	24,642	4	0.6

The risk premium due to fluctuating steel price without escalation clause is estimated as 30.2 percent of the unit bid price. Thus, an adjustment clause concerning steel price must be considered to assess the benefits of taking on the future risk while getting a lower risk premium for the bids.

4.2.3.6 Item 6 - Continued reinforced concrete pavement (CRCP)

The model specification for this control item is shown in Equation 4.8 and Table 4.15. All the variables are statistically significant and of the expected signs (See the regression output 6 from statistical software R – Item 6 CRCP in Appendix C.). The result of regression analysis shows that the data supports the claim that expected change (the observation of change in current and previous period) and historical volatility of cement prices affect the unit prices of continuous reinforced concrete pavement for contracts without adjustment clauses. The model indicates that every 1 percent increase in the expected change in cement price, will result in (on average) a 14.8 percent increase in unit bid prices of the control item.

$$UBP = \beta_0 \cdot D^{\beta_1} \cdot Qn^{\beta_2} \cdot N^{\beta_3} \cdot \beta_4^{cementFS} \cdot \prod_{j=1}^{6} \beta_{j+4}^{L_j} \tag{4.8}$$

Table 4.16 Model estimation value - CRCP

Parameter	Estimated parameter value for excavation	Parameter	Estimated parameter value for excavation
β_0	388.775	$\beta_6(d10)$	0.89
β_1	-0.057	$\beta_7(d11)$	0.812
β_2	-0.135	$\beta_8(d12)$	0.856
β_3	-0.098	$\beta_9(d13)$	0.687
β_4	1.16	$\beta_{10}(d19)$	0.769
$\beta_5(d5)$	0.832		

Considering the model estimation in Table 4.15 and an example contract structure of CRCP item let in 2004 in Lubbock (Texas) as shown in Table 4.16, the total risk premium without adjustment clauses can be estimated using Equation 4.8. UBP_R, the regressed unit bid price given input values of expected change of cement price $cementFS$, is 62.53 ($/CY); UBP_0, the regressed unit bid price given $cementFS = 0$, is 51.18 ($/CY). Thus the risk premium in unit bid price accounting for expected change of cement price is ΔUBP = $UBP_R - UBP_0$ = 11.35 ($/CY) and then total risk premium without any adjustment is $R = \Delta UBP \times Quantity = \$881523$ (10.5 percent of original bid price).

Table 4.17 Parameters for a CRCP bid

T	L	D (Day)	Qn (C.Y.)	N	$cementFS$ ($/ton)
12/07/04	Lubbock (d15)	680	77,641	4	1.35

As observed from the correlation coefficient matrix and the regression output, the cement prices can play an important role in determining the risk premium of continuous reinforced concrete pavement. It is also noted that the positive correlation coefficients between the unit bid price and steel price and the oil price indicate that a part of the risk premium can be affected by the fluctuating steel and oil prices. Thus, an escalation clause can be considered to assess the benefits of taking on the future risk while getting a lower risk premium for the bids.

4.2.3.7 Item 7 - Retaining wall

The regression model for item 7 - retaining wall is shown in Appendix C. All the variables are statistically significant and of the expected signs, except the expected change of cement price (*cementFS*) which is not significant.

The regression output shows that the expected change of cement price does not significantly affect the unit bid price of retaining walls. This result may be due to issues including the following: 1) the missing important explanatory variables that have not been included in the model; 2) the interactions among the explanatory variables. For example, the effect of the change in cement price on unit bid price of retaining walls is overwhelmed by the duration-based volatility of cement price, where the expected changes and the volatilities are often related; or 3) bidding strategies, such as, unbalancing bids which consider retaining walls as the small item. According to the TxDOT bidding data, retaining walls item is also not always included in a project. This control item is not considered in the following analysis.

4.3 Escalation Clause and Trigger Barrier

As previously discussed, price adjustment clauses can be viewed as an approach to manage the cost risk where trigger barriers specify the amount of future risk agencies are willing to take. This is in accordance with the result of a survey performed by a subcommittee of AASHTO in the fall of 2009 (AASHTO, 2009), where the specifications of price adjustments vary for different materials (fuel, asphalt cement, steel, Portland cement, and others), and trigger values.

This report considers the contract with an escalation clause triggered by *a priori* defined barrier level. For example, an escalation clause can be triggered by an increase in commodity prices of 5 percent. Even when transportation agencies utilize price adjustments, the trigger levels are only arbitrarily defined and are not considered from the risk analysis perspective. For example, the Washington State Department of Transportation (WSDOT) applies the same trigger value of 10 percent for fuel cost adjustments as long as the projects meet certain requirements (e.g., projects longer than 200 working days) (AASHTO, 2009). Developing a model that can be used to determine the optimal level of escalation triggers given agencies' risk preferences is one of the objectives of this report.

4.4 Risk Premium as Function to Barrier Levels

Trigger barriers affect both the risk premiums included in the initial estimate bids and the possible future losses to the agency. Figure 4.1 shows this phenomenon. By simulating commodity prices for a predetermined level of barrier, one can identify how losses and risk premiums are affected by the barrier levels. Contractors build-in the risk premiums by measuring the exposure (i.e., potential losses). As shown in Figure 4.1, the potential loss to the contractor is the largest if escalation clause is not allowed, which causes the contractor to include the 100% risk premium (point A) estimated in this Section; the potential losses will be zero if the trigger barrier of the escalation clause is 0%, which means that there is no need for the contractor to add any risk premium in the bids (point B). Hence the relationship between the risk premium and the trigger barrier for other than extreme conditions (i.e. $B = 0$ or $B = 100\%$) needs to be investigated.

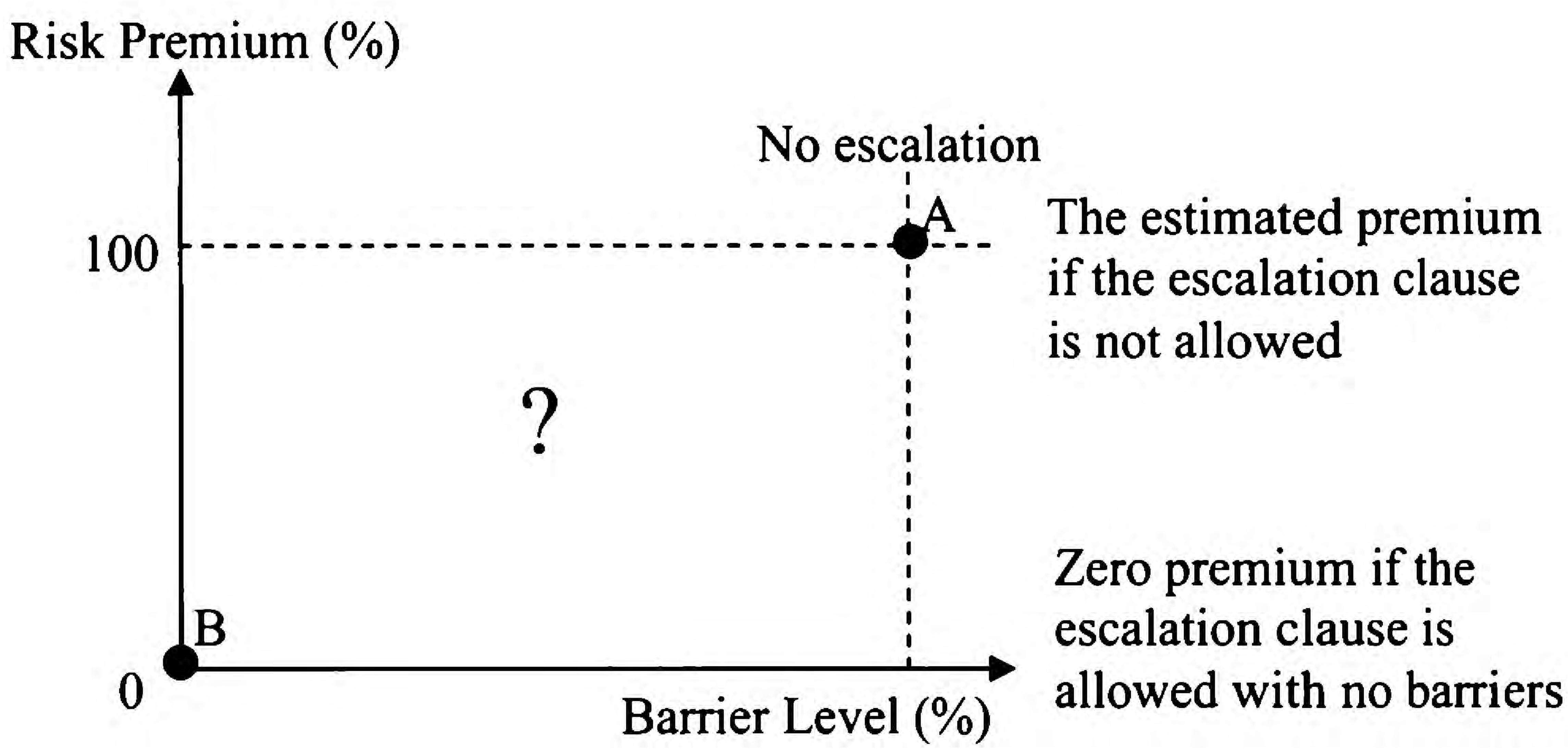


Figure 4.1 Risk premiums and barrier levels

The relationship between the average losses and barrier levels from a simulation study is shown in Figure 4.2. The barrier level for prices of crude oil was varied between 0 and 1. The losses were simulated using sample size of 100,000 paths. The result shows that the average losses decrease smoothly as the barrier level increases. In addition, the losses are not too sensitive at boundaries. This can be concluded from the slope of the losses vs. barriers. It can be seen that the barrier<0.2 or barrier>0.8 is less steep than the slop for 0.2<barrier<0.8.

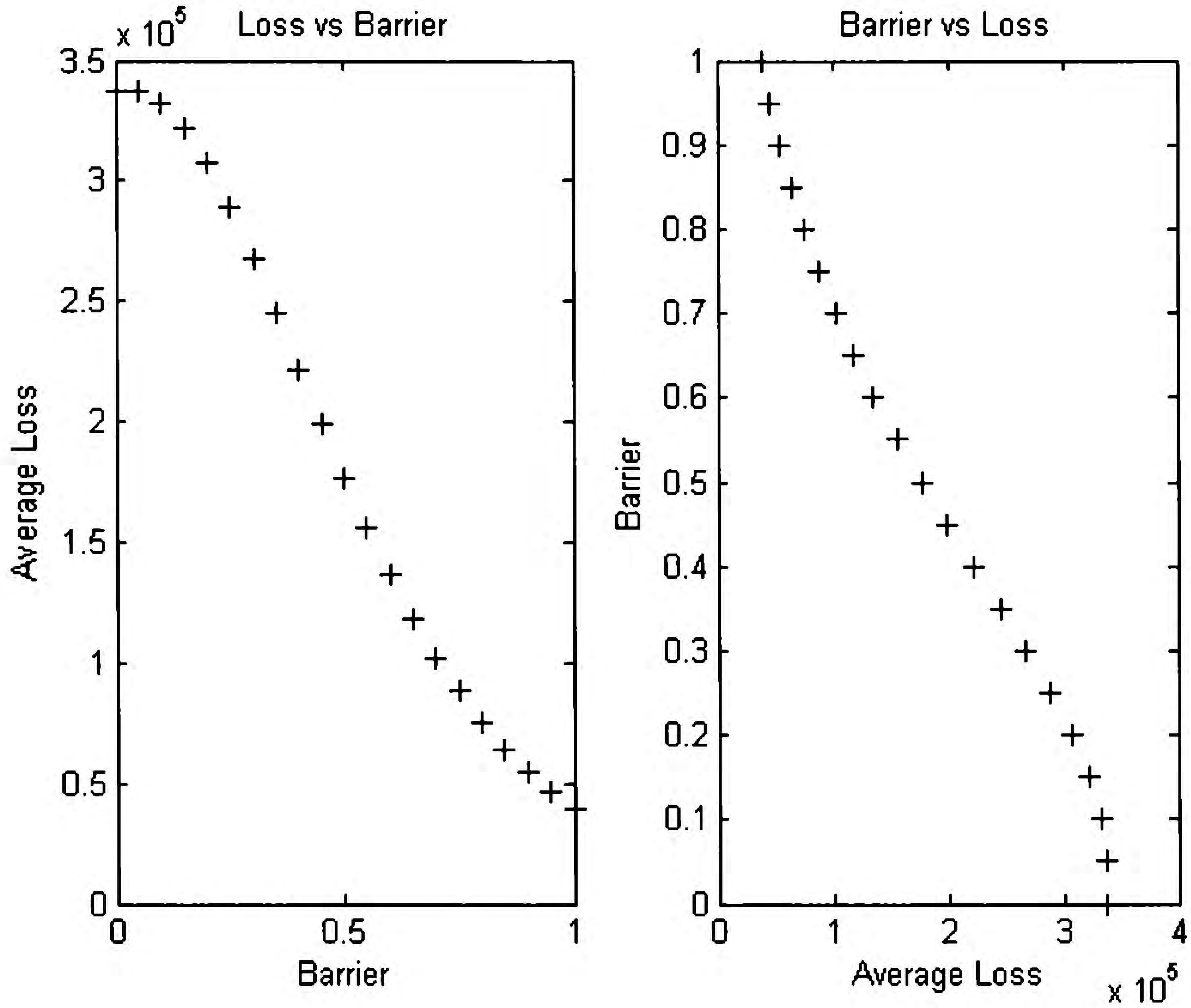


Figure 4.2 The relationship between average loss and barrier level

Based on the observation from Figure 4.2, the risk premium is assumed to be linearly proportional to the total risk premium, where the proportion is set by the barrier level. That is, $\text{Risk Premium} = \text{barrier} \times \text{Total Risk Premium}$, where $0 \le \text{barrier} \le 1$. The higher barrier level, the less average losses the owner would take. From the perspective of the contractor, the higher the barrier level, the higher the risk premium it should include in the bid, and *vice versa.* This relationship between the risk premium and the barrier level is shown in Figure 4.3.

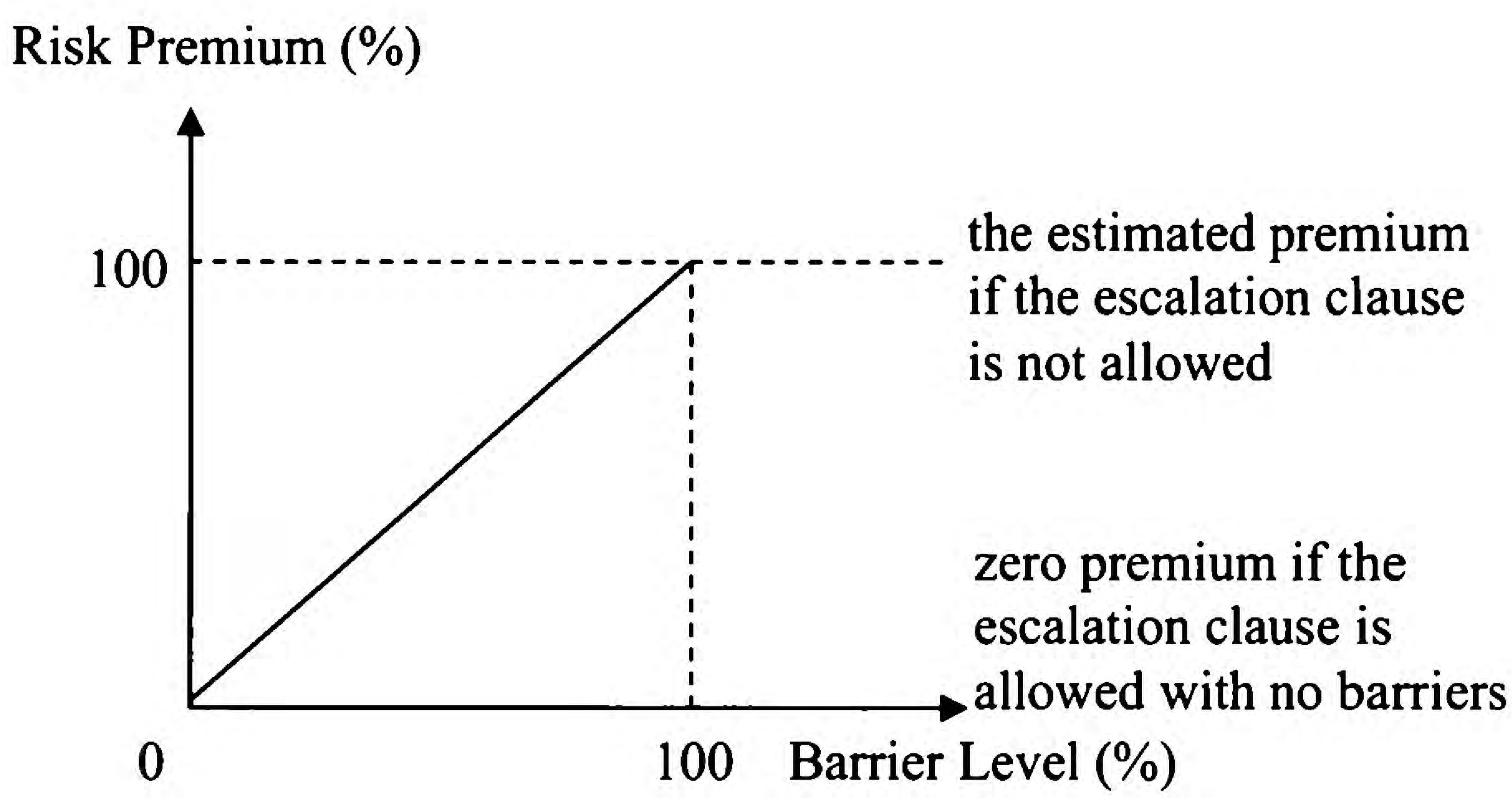


Figure 4.3 The assumed relationship between risk premiums and barrier levels

The risk premium estimated for the models in this Section considers no adjustment clause. This is equivalent to setting the barrier level to positive infinity. However, volatile commodity prices would not increase to positive infinity. It is reasonable to expect that the barrier captures market conditions if it ranges from 0 percent to 100 percent, where the barrier of a 100 percent reflects a 100 increase from the initial price of a commodity.

4.5 Summary

This section describes how commodity prices impact the unit bid prices. Cement, steel and oil prices are selected to be the key risk factors for typical highway construction bid items. The first subsection explains the motivation for selecting the seven bid items and a process for identifying the most significant risk factor for each bid item. The second subsection presents the developed regression models, the steps for building models and validating them, and the discussion about the estimation results. The last two subsections propose how to determine risk premiums for different barrier levels in contracts that allow for escalation. All of these risk premiums estimated for each control item are considered as the input parameters for optimization problems. The optimization problems are also based on the commodity prices, which are modeled in the next section.

5. UNIVARIATE AND VECTOR TIME SERIES MODELS

This section introduces two types of models that represent movements in commodity prices - autoregressive integrated moving average (ARIMA) model and vector autoregressive moving average (VARMA) model. In the first subsection, ARIMA model is presented and estimated using available data sets; in the second subsection, the correlation between the prices of commodities that are considered are presented; then, in the third subsection, VARMA model is developed and discussed.

Time series modeling of commodity prices has been a widely-used modeling approach in the econometric field (Myers, 1994; Contreras, 2003; Tuan, 2010). The series data have a natural temporal ordering. A time series model makes use of this natural ordering of time so that predicted values can be expressed in terms of their past levels. Time series models can capture all the features of commodity price series, such as, high volatilities, stochastic trends, and co-movements in commodity price series (Myers, 1994). Thus, time series models discussed in this section aim at capturing stochastic trends in prices of commodities (i.e., oil, steel and cement).

5.1 ARIMA Model

ARIMA models are typically used to predict values in univariate time series as a linear combination of their past values, past errors, and the current and past values of other time series (SAS Institute Inc., 2010). The model structure with more details is explained next.

The general idea behind ARIMA models is the synthesis of forecasting functions on the basis of discounted past observations. Consider an ARIMA (*p*, *d*, *q*) where (1) *p* indicates the order of the autoregressive (AR) part (AR part of the model indicates that the future values are weighted averages of the current and past series); (2) *d* indicates the amount of differencing; and (3) *q* indicates the order of the moving average (MA) part (MA represents the lagged forecast error part, which shows how current and past random shocks will affect the future values of series) where the difference linear operator is defined as $\Delta Y_t = Y_t - Y_{t-1}$ (Clements and Hendry, 2008). Autoregressive Moving Average Model - ARMA (*p*, *q*) - has the general form:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \qquad (5.1)$$

where Y_t = response variable at time t; $Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p}$ = response variable at time lags $t-1, t-2, \cdots, t-p$, respectively; ε_t = error term at time t; $\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_{t-q}$ = errors in previous time periods that are incorporated in the response Y_t. (Nochai and Nochai, 2006)

The ARIMA model is particularly useful when little knowledge is available on the underlying data generating process, or when there is no satisfactory explanatory model that relates the prediction variable to other explanatory variables (Zhang, 2003). The advantage of the ARIMA model is that it easily allows to add terms to random walk model to correct the model for autocorrelation in the residuals, if it is necessary, such as, adding lags of the differenced series and/or lags of the forecast errors to the prediction equation (random walk model predicts the first difference of the series to be constant).

Although ARIMA models are quite flexible at representing different types of time series (i.e., AR, MA, ARMA), the major limitation lies in the pre-assumed linear form of the model. That is, a linear correlation structure is assumed among the time series values and therefore no nonlinear patterns can be captured by the ARIMA model. The approximation of linear models to complex real-world problem is not always satisfactory (Zhang, 2003).

5.1.1 Steps for building ARIMA model

Building an ARIMA model consists of two iterative steps: 1) model identification, and 2) model estimation and diagnostic check (Montgomery *et al.*, 1998; SAS Institute Inc., 2010)

Step 1: Model identification

The first step of developing ARIMA is to identify the model structure. In this step, one needs to specify the response series and identify the candidate models. To determine whether the series is stationary, the sample autocorrelation function plot (ACF) should be considered. ACF measures how strongly time series lags are correlated with each other. If the series is not stationary, it can often be converted to a stationary series by differencing, that is, the original series is replaced by a series of differences. An ARMA model is then specified for the differenced series.

Step 2: Model estimation and model diagnostic check

The next step is to estimate the parameters for a tentative model. One needs to check for the model adequacy by considering the properties of residuals. The residuals from an ARIMA model should be normally distributed and random. To check the residuals, one can use a number of tests including the following two:

a) Tests for white noise residuals that indicate whether the residual series contains additional information that might be used by a more complex model; and

b) Plots of ACF and partial autocorrelation function (PACF) of the residuals that show if the residuals are truly random.

5.1.2 Model identification, estimation and diagnostic check

5.1.2.1 Cement price model

Step 1: Model identification

Figure 5.1 contains a time series of historical cement prices. The cement price shows gradually increasing trend during 1997-2004, followed by a sharp increase during 2004-2007. There are a number of studies investigating the factors that might have caused such phenomenon (ABARE, 2005; FHWA, 2007).

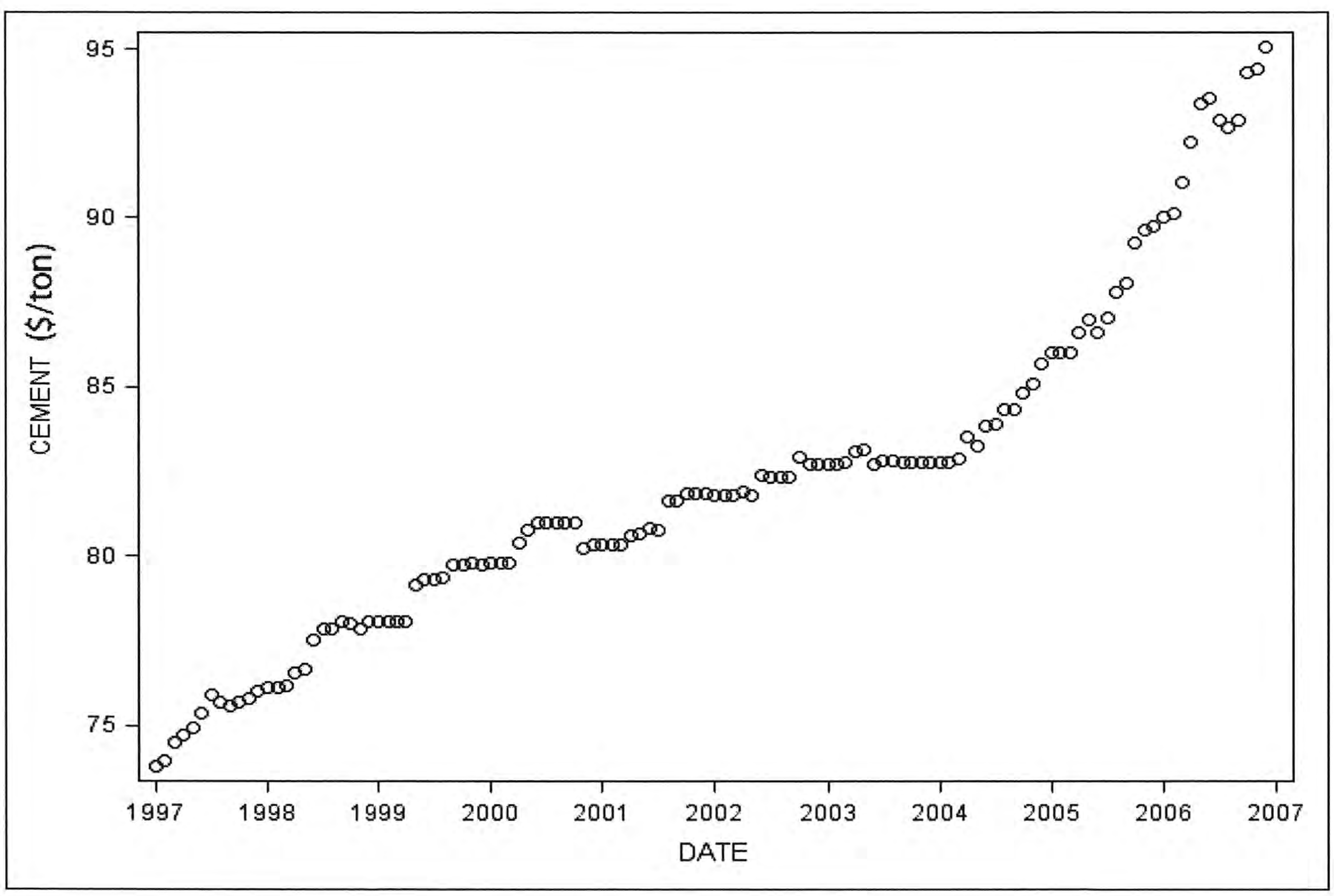


Figure 5.1 Historical cement price

A panel of autocorrelation function plots is used for the series' autocorrelation and trend analysis. The panel in Figure 5.2 contains: (1) the time series plot of the series; (2) the sample autocorrelation function plot (ACF); (3) the sample partial autocorrelation function plot (PACF); and (4) the sample inverse autocorrelation function plot (IACF).

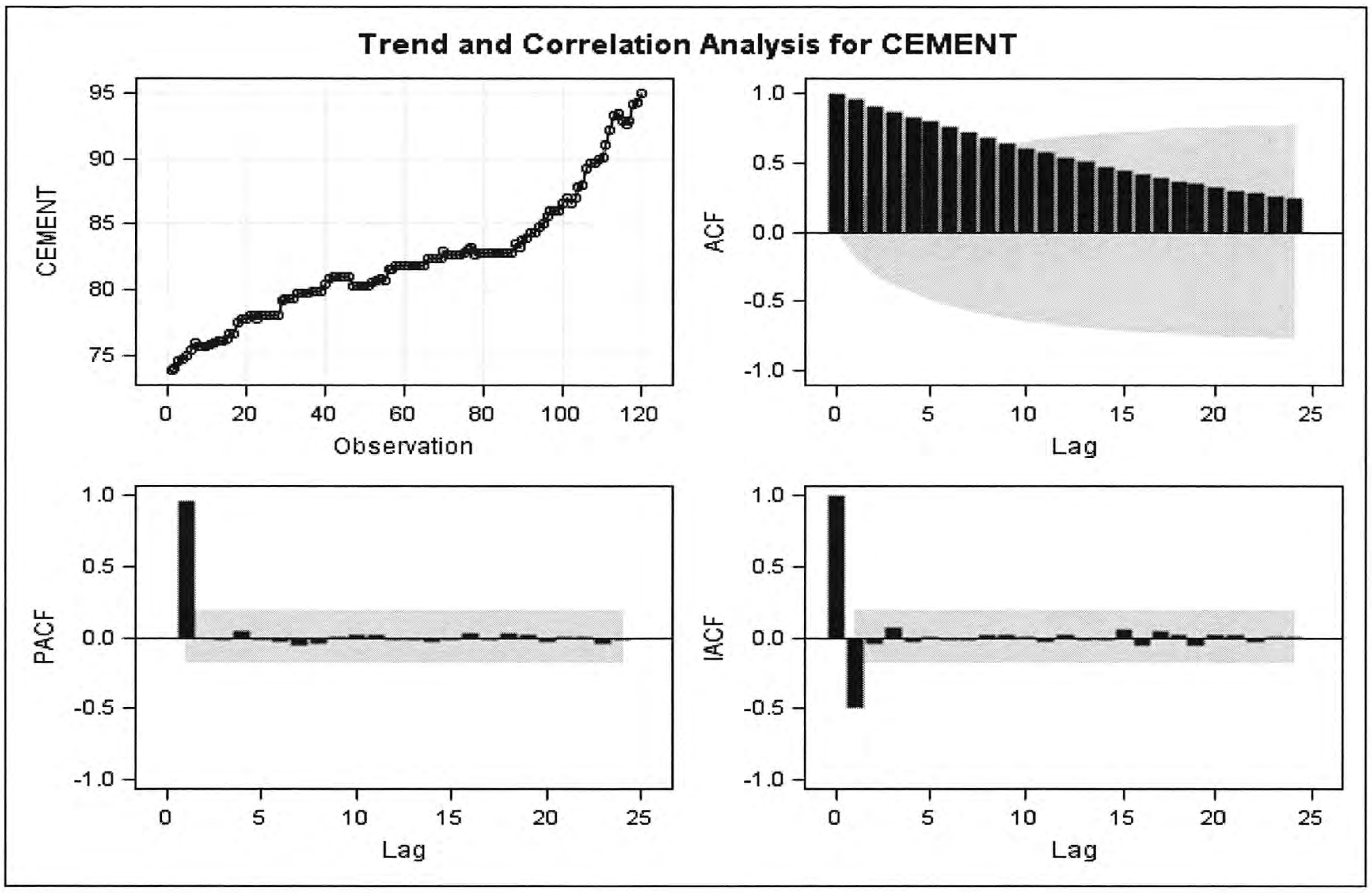


Figure 5.2 Correlation analysis of cement price

These autocorrelation function plots show the degree of correlation with the past values of the series as a function of the number of periods in the past (that is, the lag) at which the correlation is computed. By examining the plots, the series can be checked whether it is stationary or nonstationary. First, the ACF plot, shows the autocorrelation between a time series and lags of itself. Second, the PACF plot, shows the partial autocorrelation between the series and lags of itself. The PACF may intuitively be thought of as the sample autocorrelation of time series with the effects of the intervening observations eliminated. This is because a partial autocorrelation is the amount of correlation between a variable and a lag of itself that is not explained by correlations at all lower-order-lags, that is, the correlation between observations Y_t and Y_{t-p} after removing the linear relationship of all observations between Y_t and Y_{t-p}. Third, the IACF plot, is useful for detecting over-differencing. If the data have been over-differenced, the IACF looks like ACF from a nonstationary process (SAS Institute Inc., 2010). The correct amount of differencing is the lowest order of differencing that yields a time series which fluctuates around a well-defined mean value and whose ACF plot decrease fairly rapidly to zero. However, if the series is "over-differenced" by an unnecessary higher order, then over-differencing may introduce unnecessary correlations into the model and cause the loss of information.

The ACF plot in Figure 5.2 indicates that the cement price series is nonstationary since the ACF decreases very slowly. The autocorrelations are significant for a large number of lags. However, the autocorrelations at lags 2 and for higher lags may be due to the propagation of the autocorrelation at lag 1. This is confirmed by the PACF plot in Figure 5.2. Note that the PACF plot has a significant spike only at lag 1, meaning that all the higher-order autocorrelations are effectively explained by the lag-1 autocorrelation.

Before estimation, the cement price series should be transformed to a stationary series. This can be done by taking the difference of the series from one period to the next and then analyzing the differenced series. The residual series for an ARMA model must be stationary which means both the expected values of the series and the autocovariance functions are independent of time.

The next step in identification stage is the check for the white noise. This is an approximate statistical test of the hypothesis that none of the autocorrelations in the series, up to a given lag, are significantly different from zero. If this is true for all lags, then there is no information in the series to model, and no ARIMA model is needed. In other words, if the series is white noise, then it is a purely random process (SAS Institute Inc., 2010).

The χ^2 test statistics for the residuals series in Table 5.1 indicate that the residuals are not uncorrelated (white noise). The autocorrelations are checked in groups of six. Table 5.1 shows that the white noise (no-autocorrelation) hypothesis is rejected at a high level of significance (the p-values are all less than 0.0001). This means that the series is nonstationary and it needs to be transformed to a stationary series by differencing.

Table 5.1 IDENTIFY statement check for white noise - cement price

Autocorrelation check for white noise									
Lag	Chi-Square	DF	Pr> ChiSq	Autocorrelations					
1-6	553.16	6	<.0001	0.956	0.912	0.870	0.833	0.797	0.760
7-12	865.84	12	<.0001	0.720	0.677	0.638	0.603	0.572	0.540
13-18	1024.77	18	<.0001	0.508	0.475	0.443	0.416	0.388	0.366
19-24	1102.71	24	<.0001	0.346	0.324	0.303	0.284	0.264	0.244

Since the series is nonstationary, the next step is to transform it to a stationary series by differencing. That is, instead of modeling the cement price series itself, the change in the cement price is modeled from one period to the next. If the period of differencing is set as 1, the autocorrelation plots for the differenced series are shown in Figure 5.3.

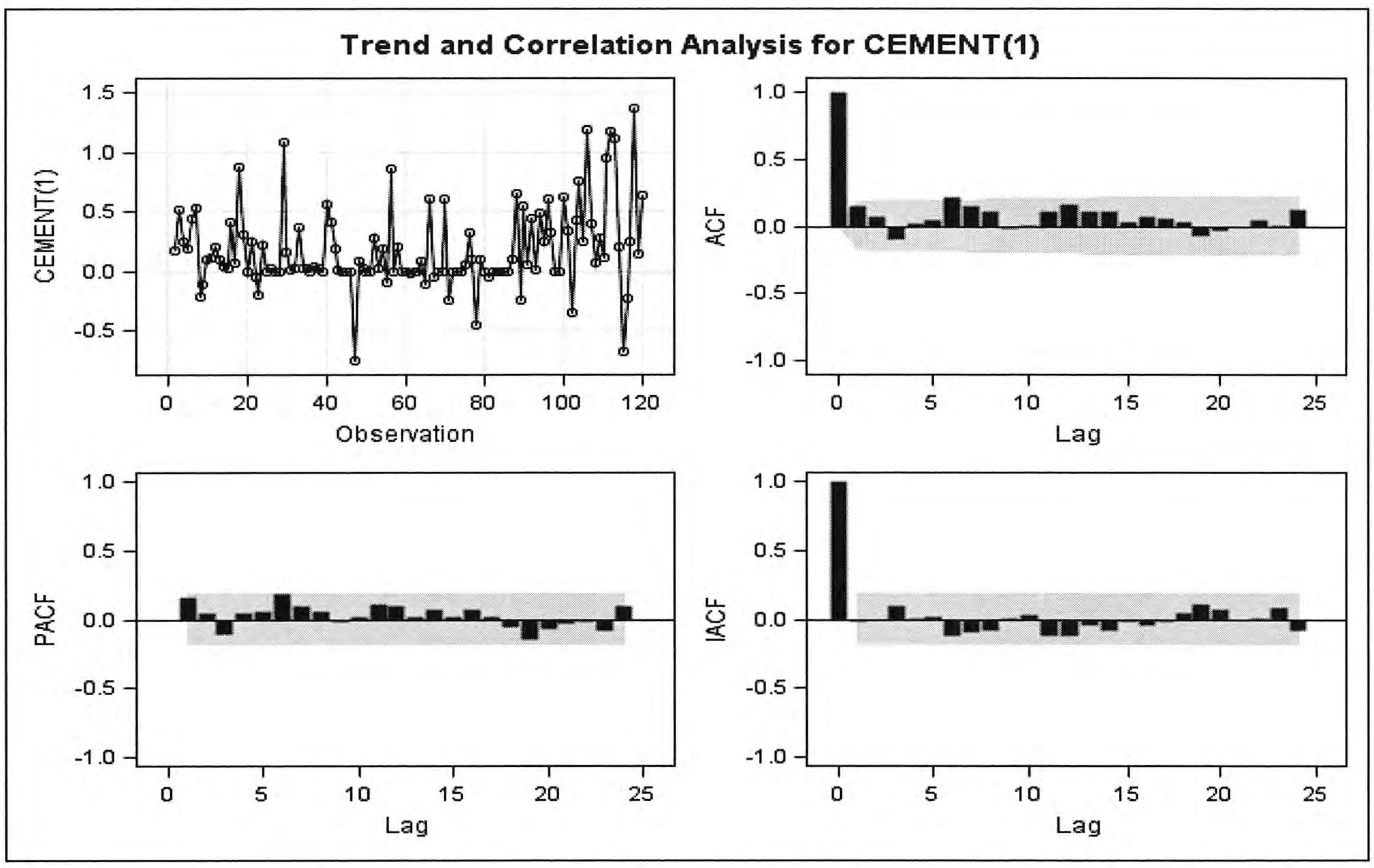


Figure 5.3 Correlation analysis of the change in cement price

The autocorrelation shown in Figure 5.3 decreases rapidly which indicates that the change in the cement price is now a stationary time series. The test for the white noise, shown in Table 5.2, indicates that the change in cement price is not autocorrelated (since the null hypothesis of no autocorrelation could not be rejected according to the p-values in Table 5.2). Thus, a *random walk with a drift* (ARIMA[0,1,0]) model is a good candidate model to fit to the series. The model has also been confirmed to be the best model by both the tentative order selection (The ARIMA procedure in SAS has diagnostic options to help tentatively identfy the orders of ARIMA processes) and the time series forecasting system in SAS (The Time Series Forecasting System provides a variety of tools for identifying potential forecasting models and for choosing the best fitting model, such as, the Series Viewer and Model Viewer). The order identification diagnostics in Table D-1 [Appendix D] gives the recommendations that an ARIMA (0,1,0) would be the best choice for a tentative model for cement price series based on 5% significance level.

Table 5.2 IDENTIFY statement check for white noise – differenced cement price

Autocorrelation check for white noise Period(s) of differencing=1									
Lag	Chi-Square	DF	Pr> ChiSq	Autocorrelations					
1-6	10.77	6	0.0957	0.156	0.068	-0.094	0.021	0.048	0.213
7-12	20.90	12	0.0518	0.152	0.109	-0.022	0.008	0.108	0.172
13-18	26.05	18	0.0986	0.108	0.118	0.033	0.081	0.055	0.031
19-24	29.94	24	0.1868	-0.076	-0.031	-0.009	0.054	0.001	0.128

Step 2: Model estimation and model diagnostic check

A random walk with the drift model is considered to predict the change in cement prices as an average change over one time period plus a random error. Table 5.3 shows the parameter estimates for this model specification. The mean term is labeled "MU"; and its estimated value is 0.17882.

Table 5.3 Estimated model for cement price

Conditional least squares estimation (Period(s) of differencing=1)				
Parameter	Estimate	Standard error	T value	Approx Pr > |t|
MU	0.17882	0.03177	5.63	<.0001

The diagnostic check of residual is shown in Figure D-1 and Figure D-2 in Appendix D. The residual and white noise test plots show that the hypothesis - the residuals are uncorrelated- cannot be rejected. As discussed before, even though the normality plot has a slight departure from normality, the model is still confirmed to be the best model by both the tentative order selection and the time series forecasting system in SAS. Thus, the model for cement price is identified as a "random walk with drift – ARIMA (0,1,0)". This model is specified as:

$$Y_t = 0.17882 + Y_{t-1} + \varepsilon_t \qquad (5.2)$$

5.1.2.2 Steel price

Step 1: Model identification

Figure 5.4 shows a time series of historical steel prices. Steel prices were highly volatile during 1997-2004, followed by a price spike from 2004 to 2007. The factors for the price increase include increase in material costs, a weak U.S. dollar, strong global demand, higher energy cost, the global consolidation of the steel industry, and others (ABARE, 2005; FHWA, 2007).

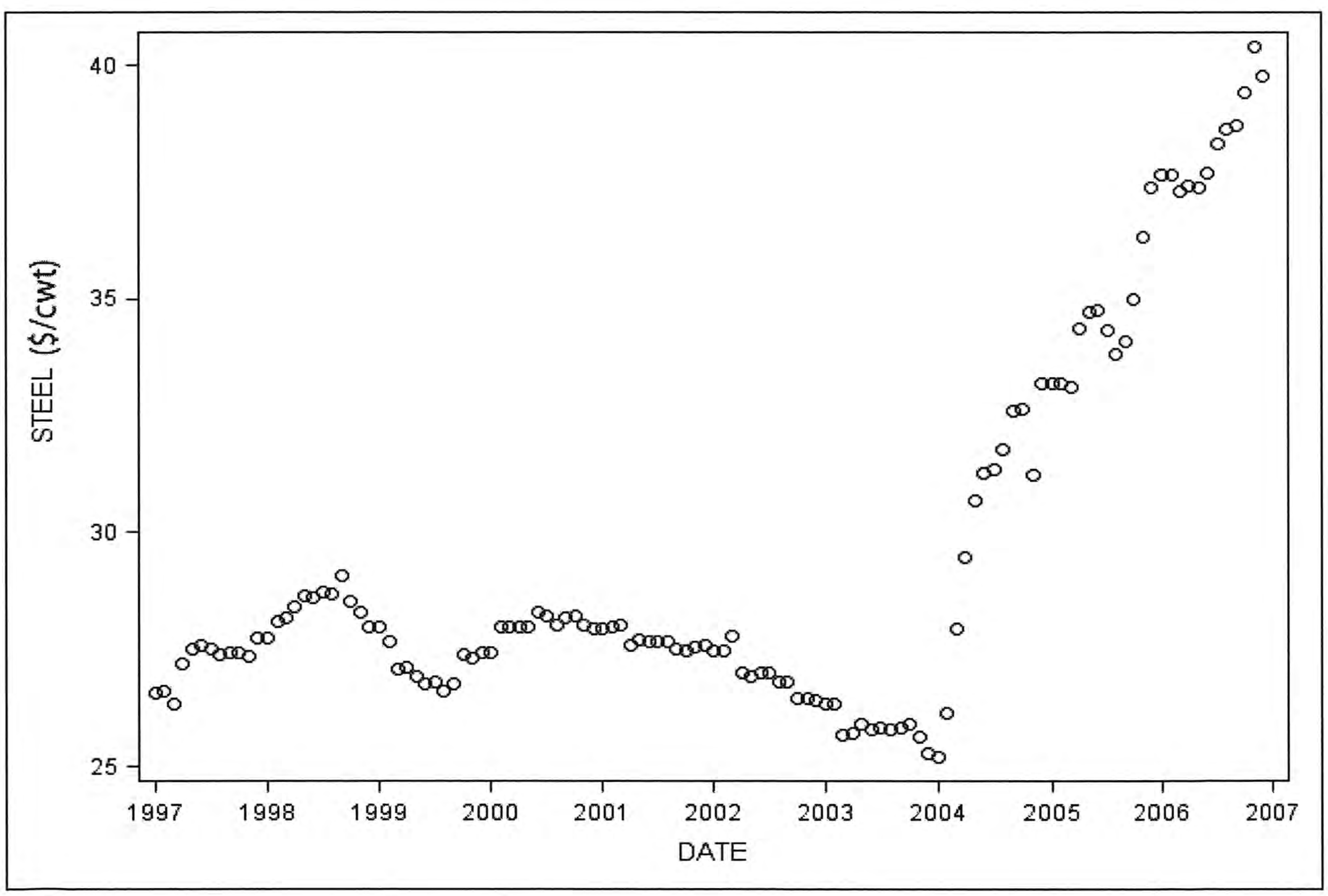


Figure 5.4 Historical steel price

Figure 5.5 shows the correlation analysis panel with time series plot of the series, the sample autocorrelation function plot (ACF) and the sample partial autocorrelation function plot (PACF). The ACF plot indicates that the steel price series is nonstationary since the ACF decreases very slowly.

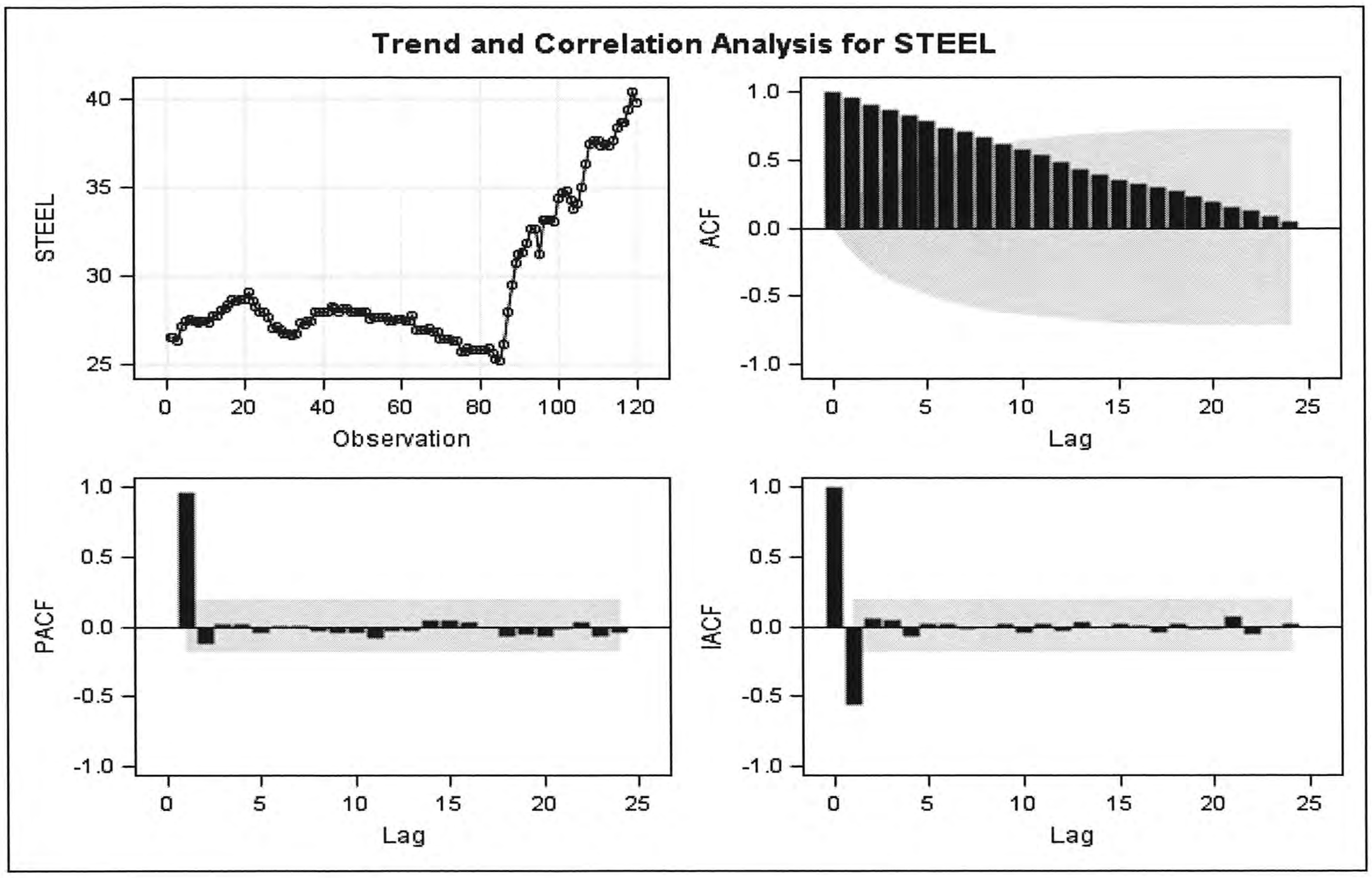


Figure 5.5 Correlation analysis of steel price

Table 5.4 shows that the white noise hypothesis is rejected with p-value less than 0.0001. Since the series is non-stationary, it needs to be transformed to a stationary one by differencing. If the period of differencing is set as 1, one can obtain the autocorrelation plots for the differenced series as shown in Figure 5.6.

Table 5.4 IDENTIFY statement check for white noise - steel price

Autocorrelation check for white noise									
Lag	Chi-Square	DF	Pr> ChiSq	Autocorrelations					
1-6	545.33	6	<.0001	0.960	0.912	0.867	0.825	0.783	0.742
7-12	831.93	12	<.0001	0.703	0.664	0.622	0.579	0.531	0.482
13-18	934.28	18	<.0001	0.433	0.391	0.355	0.326	0.299	0.268
19-24	955.20	24	<.0001	0.233	0.195	0.156	0.123	0.088	0.053

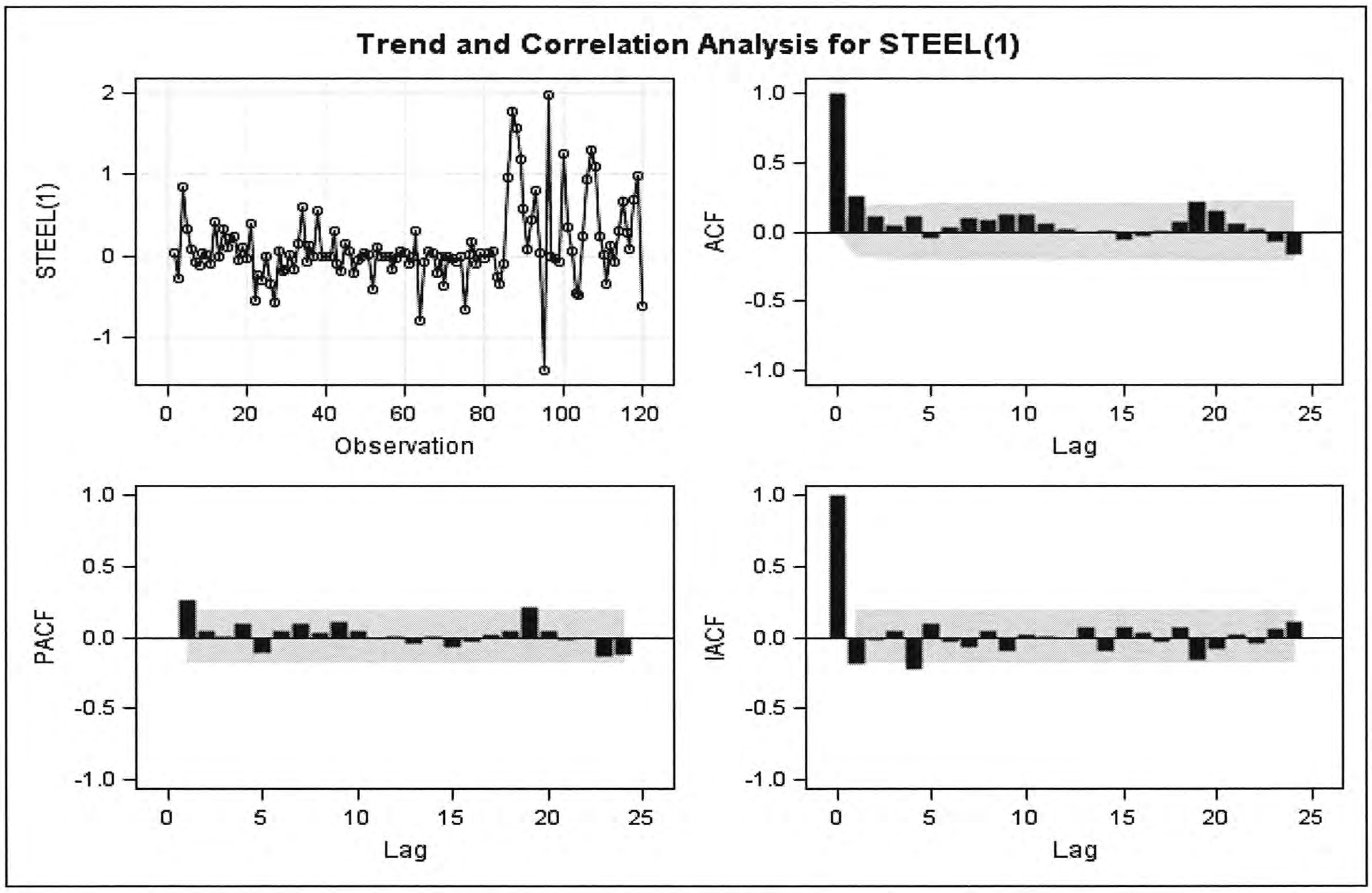


Figure 5.6 Correlation analysis of the change in steel price

The autocorrelation shown in Figure 5.6 decreases rapidly but PACF plot still indicates autoregression of lag 1. The check for white noise as shown in Table 5.5 indicates that the change in steel price might be slightly autocorrelated (p-value for the first six lags is 0.0607). The autocorrelations in Table 5.5 are corresponding to the lags shown in the first column of the table. For example, the autocorrealtions for lag 1, lag 2, lag 3, lag 4, lag 5, lag 6, are 0.261, 0.113, 0.050, 0.112, -0.041 and 0.028, respectively. Furthermore, the result from automatic best model selection process suggests the AR(1) model. The model has also been confirmed to be the most suited model by both the tentative order selection and the time series forecasting system in SAS. For example, after the series is differenced, the order identification diagnostics in Table D-2 [Appendix D] gives the recommendations that an ARIMA (1,1,0) would be a good choice for a tentative model for steel price series based on 5% significance level.

Table 5.5 IDENTIFY statement check for white noise – differenced steel price

Autocorrelation check for white noise Period(s) of differencing = 1									
Lag	Chi-Square	DF	Pr> ChiSq	Autocorrelations					
1-6	12.06	6	0.0607	0.261	0.113	0.050	0.112	-0.041	0.028
7-12	19.12	12	0.0857	0.095	0.091	0.132	0.121	0.064	0.025
13-18	20.41	18	0.3102	-0.007	0.003	-0.057	-0.035	0.011	0.067
19-24	35.88	24	0.0564	0.216	0.149	0.064	0.015	-0.070	-0.166

Step 2: Model estimation and model diagnostic check

An AR(1) model predicts the change in steel prices as an average change of one time period, plus some fraction of the previous change, plus a random error. Table 5.6 shows the parameter estimates and the goodness-of-fit statistics for this model. The mean term is labeled MU; and its estimated value is 0.1087. The autoregressive parameter is labeled "AR1,1"; this is the coefficient of the lagged value of the change in steel price, and its estimate is 0.26583.

Table 5.6 Final parameter estimates for AR(1) model - steel price

Conditional least squares estimation					
Parameter	Estimate	Standard Error	T value	Approx Pr > |t|	Lag
MU	0.10870	0.05813	1.87	0.0640	0
AR1,1	0.26583	0.09005	2.95	0.0038	1

Correlations of parameter estimates		
Parameter	MU	AR1,1
MU	1.000	-0.018
AR1,1	-0.018	1.000

Autocorrelation check of residuals									
Lag	Chi-Square	DF	Pr> ChiSq	Autocorrelations					
1-6	3.20	5	0.6697	-0.011	0.036	-0.012	0.127	-0.088	0.014
7-12	6.31	11	0.8518	0.073	0.041	0.090	0.086	0.033	0.014
13-18	7.10	17	0.9822	-0.022	0.014	-0.062	-0.030	0.007	0.017
19-24	18.06	23	0.7543	0.188	0.091	0.018	0.020	-0.037	-0.170

In Table 5.6, the autocorrelation check of residuals shows that the test statistics does not reject the no-autocorrelation hypothesis (because the p-values are large). This implies that the residuals are the white noise, and so the AR(1) model is an adequate model representation. There is no need to further develop more complex models as the benefits are minimal. A more visual diagnostic check of the residuals is shown in Figure D-3 and Figure D-4 [Appendix D].

Thus, the model identified for steel price is a "differenced first-order autoregressive model - IAR (1,1)" with the model equation stated below:

$$Y_t - Y_{t-1} = 0.1087 + 0.26583(Y_{t-1} - Y_{t-2}) + \varepsilon_t \qquad (5.3)$$

which is equivalent to:

$$Y_t = 0.1087 + 1.26583Y_{t-1} - 0.26583Y_{t-2} + \varepsilon_t \qquad (5.4)$$

5.1.2.3 Crude oil price

Step 1: Model identification

Figure 5.7 shows a time series of historical oil prices. Global oil price has been rising rapidly. Increased global consumption of fuels as well as limited crude oil production, and speculative demand has all contributed to this increase in oil price (ABARE, 2005; Xie *et al.*, 2006; FHWA, 2007).

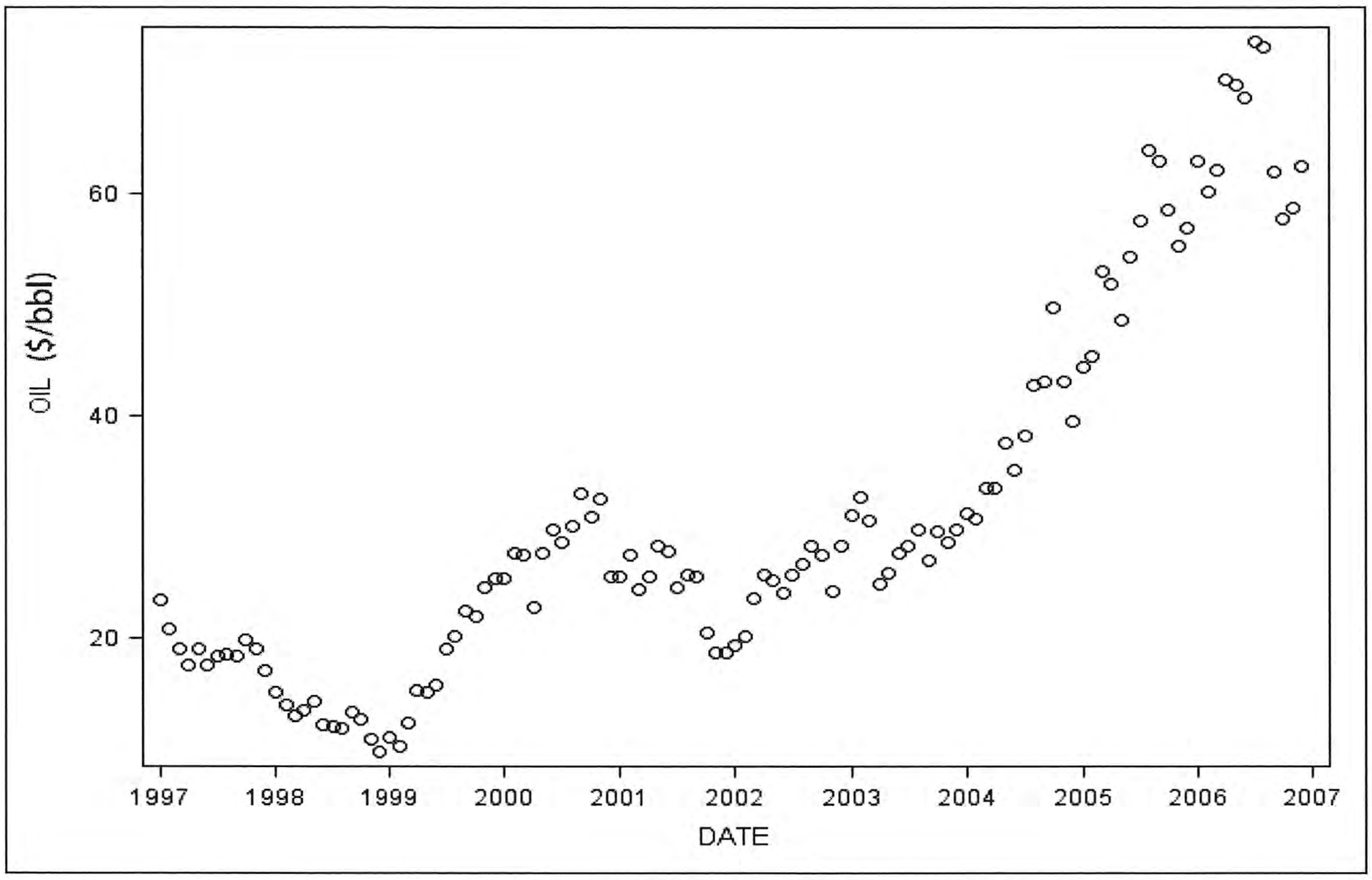


Figure 5.7 Historical oil price

Figure 5.8 shows the correlation analysis panel with time series plot of the series, the sample autocorrelation function plot (ACF) and the sample partial autocorrelation function plot (PACF). The ACF plot indicates that the oil price series is nonstationary since the ACF decreases very slowly.

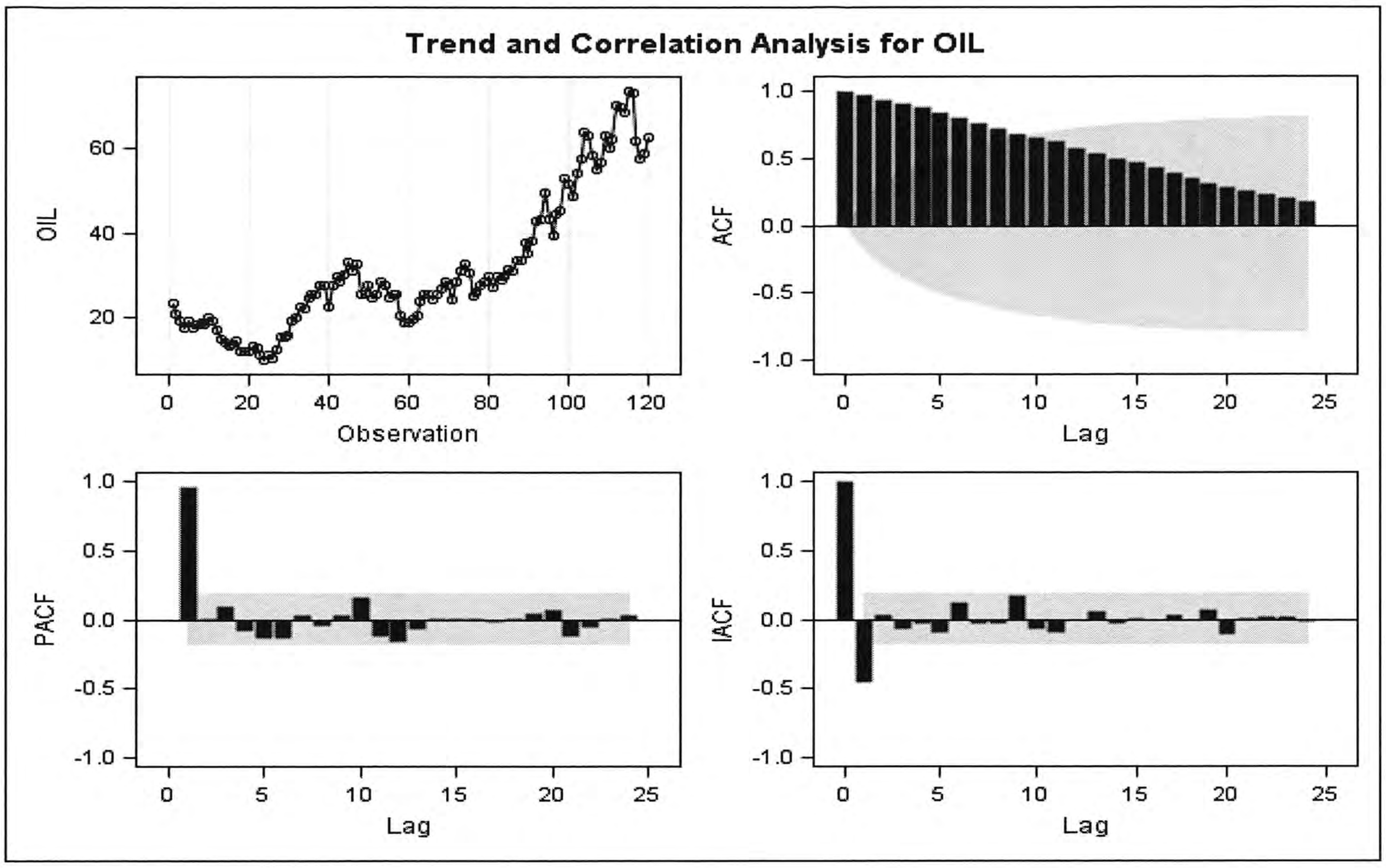


Figure 5.8 Correlation analysis of oil price

Table 5.7 shows that the white noise hypothesis is rejected as the p-value is less than 0.0001. Since the series is nonstationary, it needs to be transformed into a stationary one by differencing. If the period of differencing is set as 1, the autocorrelation plots for the differenced series are shown in Figure 5.9.

Table 5.7 IDENTIFY statement check for white noise - oil price

Autocorrelation check for white noise									
Lag	Chi-Square	DF	Pr> ChiSq	Autocorrelations					
1-6	599.96	6	<.0001	0.967	0.936	0.912	0.883	0.845	0.803
7-12	963.17	12	<.0001	0.765	0.726	0.688	0.660	0.625	0.581
13-18	1131.87	18	<.0001	0.538	0.502	0.466	0.425	0.385	0.349
19-24	1187.71	24	<.0001	0.317	0.292	0.262	0.230	0.200	0.176

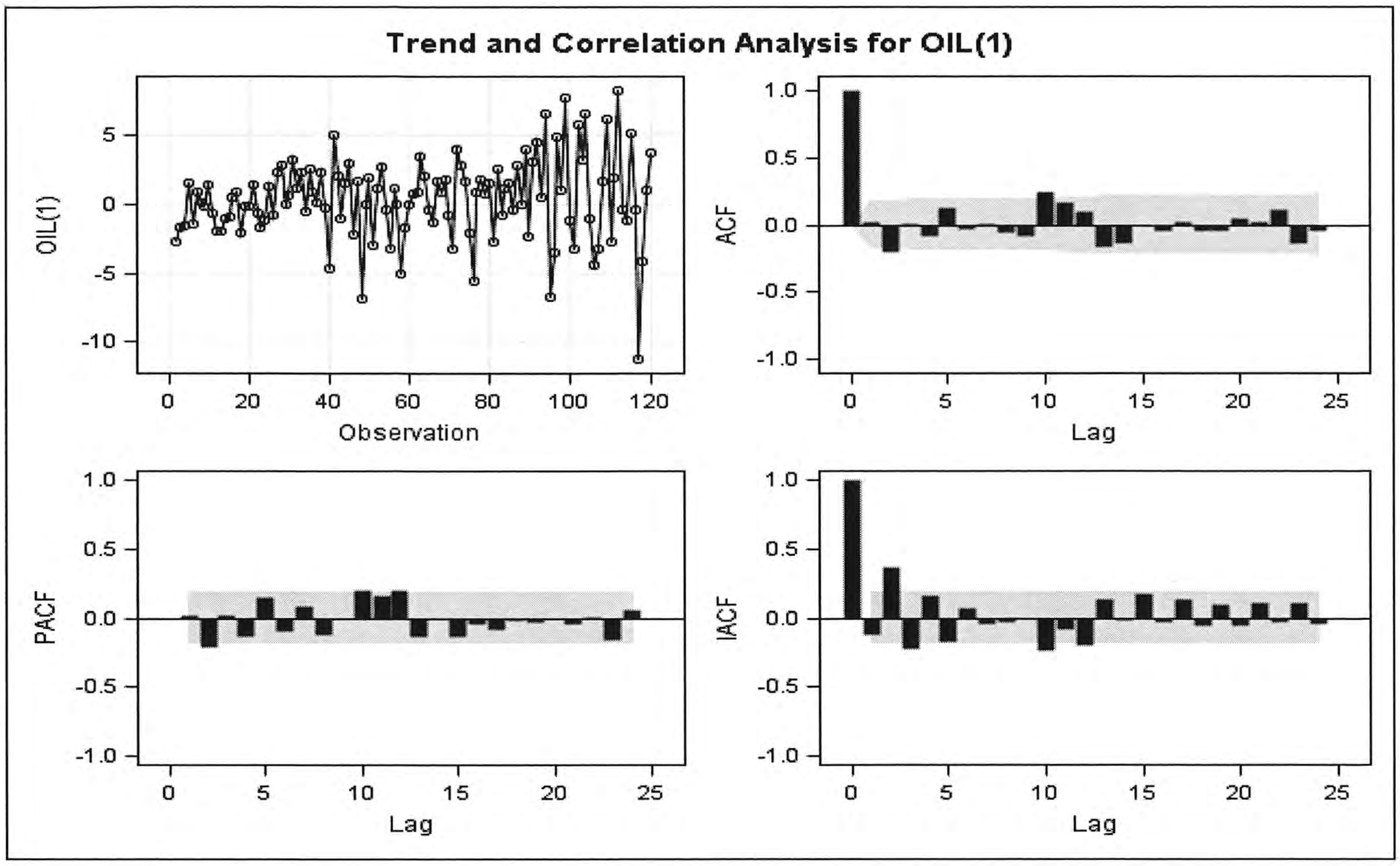


Figure 5.9 Correlation analysis of the change in oil price

Afer differencing, the autocorrelations shown in Figure 5.9 decrease rapidly. In addition, the check for white noise (Table 5.8) show low level of significance for lags. The results of estimates for "first-differenced" model I(1) is shown in Table 5.9.

Table 5.8 IDENTIFY statement check for white noise – differenced oil price

Autocorrelation check for white noise (Period(s) of differencing=1)									
Lag	Chi-Square	DF	Pr> ChiSq	Autocorrelations					
1-6	8.29	6	0.2174	0.022	-0.206	0.006	-0.090	0.123	-0.025
7-12	22.89	12	0.0287	0.005	-0.060	-0.078	0.252	0.162	0.105
13-18	29.61	18	0.0415	-0.163	-0.131	0.001	-0.049	0.022	-0.050
19-24	35.29	24	0.0643	-0.041	0.042	0.019	0.109	-0.141	-0.051

Table 5.9 Estimated I(1) model for oil price

Conditional least squares estimation				
Parameter	Estimate	Standard error	T value	Approx Pr > |t|
MU	0.32714	0.27405	1.19	0.2350

However, the IDENTIFY statement plots in Figure 5.8 suggest a mixed autoregressive and moving-average model, which adds both an autoregressive term and a moving-average term to the I(1) model. This model has also been confirmed to be the best model by both the tentative order selection and the time series forecasting system in SAS. For example, after the series is differenced, the order identification diagnostics in Table D-3 (Appendix D) recommends that the ARIMA(1,1,1) as the best choice for a tentative model for oil price series based on 5% significance level. Thus ARMA(1,1) model for the change in oil price is estimated next.

Step 2: Model estimation and model diagnostic check

The ARIMA(1,1,1) model predicts the change in oil price as an average change, plus some fraction of the previous change, and plus a random error, and plus some fraction of the random error in the preceding period. Table 5.10 shows the parameter estimates and the goodness-of-fit statistics. The mean term is labeled MU; and its estimated value is 0.3257. The autoregressive parameter is labeled "AR1,1"; this is the coefficient of the lagged value of the change in oil price and its estimate is -0.63244. The moving-average parameter estimate, labeled "MA1,1", is "-0.80513". Both the moving-average and the autoregressive parameters have significant *t* values.

Table 5.10 Final parameter estimates for ARIMA (1,1,1) model

Conditional least squares estimation													
Parameter			Estimate		Standard error		T value		Approx Pr > |t|			Lag	
MU			0.32570		0.29829		1.09		0.2772			0	
MA1,1			-0.80513		0.16142		-4.99		<.0001			1	
AR1,1			-0.63244		0.20901		-3.03		0.0031			1	
Correlations of parameter estimates													
Parameter					MA1,1				AR1,1				
MU					0.001				0.002				
MA1,1					1.000				0.938				
Autocorrelation check of residuals													
Lag	Chi-Square	DF	Pr> ChiSq	Autocorrelations									
1-6	4.91	4	0.2966	-0.092	-0.115	-0.027	-0.073	0.108	-0.015				
7-12	19.86	10	0.0306	-0.013	-0.021	-0.109	0.269	0.086	0.143				
13-18	25.77	16	0.0573	-0.183	-0.088	-0.016	-0.031	0.012	-0.028				
19-24	34.13	22	0.0477	-0.057	0.066	-0.023	0.144	-0.165	-0.011				

In Table 5.10, the autocorrelation check of residuals shows that the test statistics does not reject the no-autocorrelation hypothesis. This means the residuals are white noise, and so the ARIMA(1,1,1) model is an adequate model. The visual diagnostic check of residual is shown in Figure D-5 and Figure D-6 [Appendix D]. The residual and white noise test plots show that one cannot reject the hypothesis that the residuals are uncorrelated. The normality plots also show no departure from normality. The *t* values provide significance tests for the parameter estimates and indicate whether some terms in the model might be unnecessary. In this case, the t value for the AR and MA parameters are highly significant, but the t value for MU indicates that the mean term adds little to the model.

Thus, it is concluded that the ARIMA(1,1,1) model is adequate model representation for changes in crude oil price. Thus, the model identified for oil price is a "A 'mixed' model - ARIMA (1,1,1)" with the model equation shown below:

$$Y_t - Y_{t-1} = 0.3257 - 0.63244(Y_{t-1} - Y_{t-2}) + 0.80513\varepsilon_{t-1} + \varepsilon_t \quad (5.5)$$

Which is equivalent to:

$$Y_t = 0.3257 + 0.36756Y_{t-1} + 0.63244Y_{t-2} + 0.80513\varepsilon_{t-1} + \varepsilon_t \quad (5.6)$$

5.2 Correlations between Commodity Prices

Commodity prices are often correlated. For example, sharp oil price movements are likely to disturb aggregate economic activity including the supply/demand relationship for other commodities (Xie *et al.*, 2006). For example, energy represents about 50 percent of the production costs of cement, so any increase in the energy cost will affect the cement manufacturing. (The significant increase in fuel costs had increased not only the manufacturing price but also the cost of distribution [McGoldrick, 2006]) Also, steel production is energy intensive. The correlations between commodity prices can be supported by Table 5.11.

Table 5.11 shows the Pearson correlation and partial correlation, with p-values under the null hypothesis of zero correlation between prices of commodities. Pearson's correlation coefficient between the two variables is defined as the covariance of the two variables divided by the product of their standard deviations, while partial correlation measures the degree of association between two variables with the effect of a set of controlling variables removed, that is, the amount of correlation between two variables which is not explained by their mutual correlations with a specified set of other variables.

All the correlation coefficients show a strong positive and significant relationship. As one commodity price increases, the prices of the other two increase as well. For example, the coefficients of correlation and partial correlation between the prices of crude oil and steel is 0.87 and 0.77, respectively.

Table 5.11 Pearson correlation coefficients and partial correlation coefficients

Pearson correlation coefficients (N=120)		
	Steel	Oil
Cement	0.82706 (<.0001)	0.91995 (<.0001)
Steel	1.00000	0.87206 (<.0001)
Pearson partial correlation coefficients (N=120)		
	Steel	Oil
Cement	0.75762 (<.0001)	0.61542 (<.0001)
Steel	1.00000	0.76641 (<.0001)

In a univariate model (such as, ARIMA), estimation is conducted individually for each commodity price. In other words, the effect of correlation among the commodity prices as shown in Table 5.11 is ignored. Multivariate model (such as, VARMA), where a single model is fitted using historical time series of all commodities, is needed to account for the significant effect of correlation. The multivariate model incorporates all information, and estimates the dynamic interactions among multiple time series of commodity prices (Kamarianakis and Prastacos, 2003). For example, the risk can be significantly underestimated, if the prices of commodities show dependence on each other. Variance increases rapidly as the correlation among the risk factor (i.e., commodity price) increases.

5.3 VARMA Model

As the previous subsection shows, the effect of correlations could not be ignored when developing forecasting models. As reported in literature, many commodity prices share a tendency to move together over time, or are expected to move together in the long run to equilibrium (Myers, 1994; Ghosh *et al.*, 1999; Tuan, 2010). The co-movement can be due to (1) supply and demand shocks to any commodity that spill over into other related commodities causing a group of commodity prices to move together; (2) common macroeconomic shocks; and (3) market speculation and overreaction that cause spillovers between commodity markets (Myers, 1994). In the case where variables are related to each other, vector time series models are better representation than univariate model (e.g., ARIMA).

5.3.1 Steps for building VARMA model

This subsection provides the steps for building a VARMA model and discusses how the model considers the cross- and auto-correlations among and in the series.

Step 1: Tentative order selection

As in bulding an ARIMA model, the VARMA model selection procedure is based on diagnostics to help tentatively identfy the orders of a VARMA (p, q) process. Here, the minimum information criterion (MINIC) method is used (SAS Institute Inc., 2010), and can suggest various associated AR and MA orders.

Step 2: Unit root and cointegration tests

For situations where the stationarity of the time series is in question, the VARMAX procedure provides: a) Dickey-Fuller tests for the nonstationarity of each series to aid in determining the presence of unit roots, and b) Johansen cointegration test between series to aid in determining the presence of cointegration. If the stationarity condition is not satisfied, a differenced model or an error correction model might be more appropriate.

To formalize the co-movement among the commodity prices, cointegration needs to consider both the short-term and long-run dynamics in a multivariate system (Tuan, 2010). Even though individual variables may not be stationary, linear combination of them can be stationary when the variables are cointegrated. The long-term cointegrating vector implies that although short run prices vary, they would revert to their long-term equilibrium (Ardeni, 1989; Tuan, 2010).

Step 3: Model estimation and model diagnostic check

Based on the cointegration, the error correction model is then estimated to account for dynamic adjustments for long-run and short-run relationships among the series. Error-correcting allows long-run components of the variables to follow equilibrium constraints while the short-run components have a flexible dynamic specification (Engle and Granger, 1987). The error correction form of a cointegrated system has the advantage of separating the cointegration long-run or equilibrium relations from the short-term dynamics (Lutkepohl and Claessen, 1997).

The model checks and residual analysis include F test for AR and ARCH disturbance, and normality tests.

5.3.2 Model estimation and diagnostic check

Step 1: Tentative order selection

As previously mentioned, commodity prices co-vary. Figure 5.11 shows time series of historical cement, steel and oil prices from 1997 to 2006. Oil, steel and cement prices all increased sharply during the period from 2004 to 2006. Crude oil market experienced a surge in the price of crude oil from late 1998 to 2006. The price of cement tracks the growth in oil prices, as the production of cement is a highly fuel-intensive process (FHWA, 2007); in 2004, steel prices escalated and even exceeded the peak levels when the steel supply was tight.

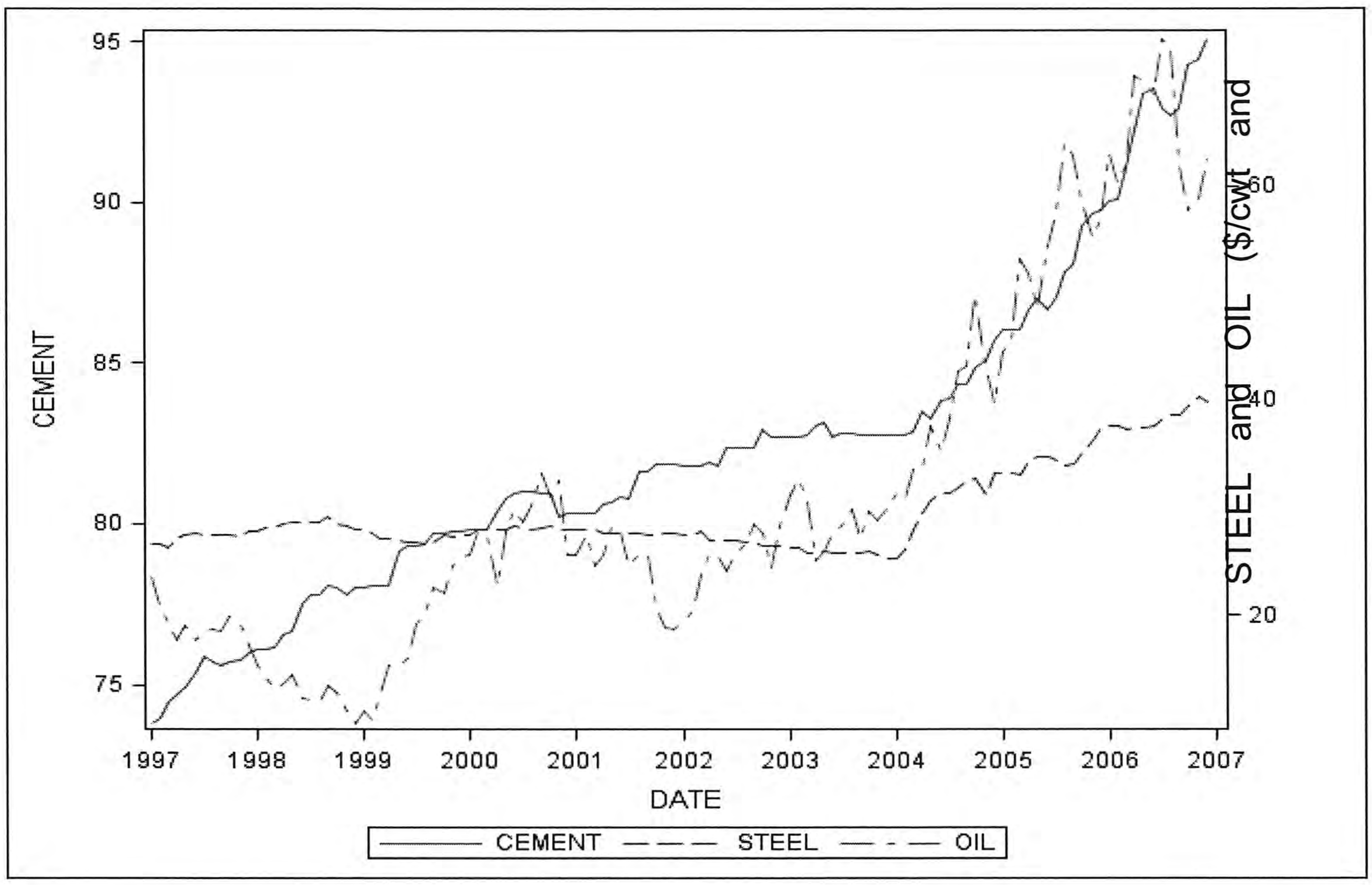


Figure 5.10 Plot of multiple time series

A VAR model with AR order 2 is suggested according to the result of tentative order selection, as well as the partial autoregression matrices, the partial cross-correlation matrices, and the partial canonical correlations (see Figure E-1, Figure E-2, Figure E-3, Figure E-4 with explanations in Appendix E).

Step 2: Unit root and cointegration tests

Table 5.12 shows the output for Dickey-Fuller tests for the nonstationarity of each series based on the null hypothesis that there is a unit root for individual series (that is, series is nonstationary). In Dickey-Fuller tests, three types of models are specified. These are zero-mean, single-mean and trend models. In Table 5.12, "Rho" and "Tau" represent the test statistics for unit rooting testing with their corresponding p-values. The p-values display that all series have a unit root, that is, non-stationary could not be rejected.

Table 5.12 Unit root tests

Dickey-Fuller unit root tests					
Variable	Type	Rho	Pr < Rho	Tau	Pr < Tau
Cement	Zero Mean	0.26	0.7440	4.37	0.9999
	Single Mean	1.76	0.9964	1.92	0.9998
	Trend	1.75	0.9996	0.61	0.9995
Steel	Zero Mean	0.46	0.7946	1.92	0.9866
	Single Mean	1.73	0.9962	0.93	0.9956
	Trend	-0.88	0.9890	-0.36	0.9878
Oil	Zero Mean	0.96	0.9040	1.00	0.9161
	Single Mean	-0.47	0.9283	-0.22	0.9315
	Trend	-9.96	0.4223	-2.37	0.3946

The Johansen cointegration rank test in Table 5.13 shows that the series are integrated at order 1. The null hypothesis is that the number (rank) of cointegrating vectors is less than or equal to r against the alternative hypothesis. For example, the first row tests rank $r = 0$ against $r > 0$; the second row tests $r = 1$ against $r > 1$. The last two columns in Table 5.13 explain the cointegration rank test with integrated order 1. The results indicate that there is a cointegrated relationship with cointegration rank 1 at the 5% significance level because the test statistic of 9.2489 is smaller than the critical value of 12.21. There is no evidence that the series are integrated order 2 at the 5% significance level (by looking at the row associated with r=1 and comparing the test statistic value and critical value pairs such as (81.58186, 12.21) and (5.81509, 4.14)). The cointegration relationship among the vector time series indicates that they are of long-run equilibrium.

Table 5.13 Cointegration rank test

Cointegration rank test for I(2)					
r\k-r-s	3	2	1	Trace of I(1)	5% CV of I(1)
0	189.63469	101.66136	39.65366	42.0797	24.08
1		81.58186	5.81509	9.2489	12.21
2			3.21875	0.7712	4.14
5% CV I(2)	24.08000	12.21000	4.14000		

Step 3: Model estimation and model diagnostic check

Based on the observation that the time series are cointegrated with rank 1, a VECM(2) form is fitted to the data. Further explanations are provided in Appendix E: Figure E-5 shows the estimates of the long-run parameter and the adjustment coefficient; Figure E-6 shows the parameter estimates and their significance.

The VECM(2) model fits the data well according to the diagnostic checks in Figure E-7 and Figure E-8. (Further statement and explanation are provided in Appendix E.) Table 5.14 provides the VAR(2) representation (the fitted series and forecast trend plots are shown in Figure E-9 [Appendix E]).

Table 5.14 Infinite order AR representation

Infinite order AR representation				
Lag	Variable	Cement	Steel	Oil
1	Cement	1.02744	-0.02631	-0.00404
	Steel	0.15799	1.23130	-0.02996
	Oil	0.81295	0.94063	1.04051
2	Cement	-0.03876	0.06105	0.00700
	Steel	-0.16181	-0.21958	0.03096
	Oil	-0.81131	-0.94565	-0.04093

The VECM(2) form in Table 5.14 can be rewritten as the following second-order vector autoregressive model below. It explains the relationship between the commodity prices, and how they affect each other. For example, the cement price at time t is determined by the sum of: (1) positive components: cement price at time lag $t-1$ and oil price at time $t-2$; (2) negative components: steel and oil price at time $t-1$, and cement and steel price at time $t-2$; and (3) the error term.

$$\begin{pmatrix} cement \\ steel \\ oil \end{pmatrix}_t = \begin{pmatrix} 1.0274 & -0.0263 & -0.004 \\ 0.158 & 1.2313 & -0.03 \\ 0.813 & 0.9406 & 1.0405 \end{pmatrix} \begin{pmatrix} cement \\ steel \\ oil \end{pmatrix}_{t-1} + \begin{pmatrix} -0.0388 & -0.0611 & 0.007 \\ -0.1618 & -0.2196 & 0.031 \\ -0.8113 & -0.9457 & -0.0409 \end{pmatrix} \begin{pmatrix} cement \\ steel \\ oil \end{pmatrix}_{t-2} + \varepsilon_t \tag{5.7}$$

5.4 Summary

This section introduces two main time series models for simulating (forecasting) commodity prices. This simulation is used for determining the optimal level of trigger barrier in escalation contracts. In the first subsection, univariate time series model is discussed and estimated; in the second subsection, the strong positive correlations between commodity prices are identified and presented. This implies that the correlations of commodity prices could not be ignored when developing optimal risk hedging strategies. Finally, in the third subsection, vector time series model is developed to address the concerns about the effect of correlations. Next section formulates and presents the multi-objective optimization problem, along with the solution methods.

6. MULTI-OBJECTIVE AND SINGLE-OBJECTIVE OPTIMIZATION

This section presents both a single-objective and a multi-objective optimization models that could assist highway agencies in developing optimal risk hedging strategies using escalation clauses with barriers. The principle difference between a single-objective and multi-objective optimization is outlined in the first subsection. The formulation of optimization models is presented in the second subsection, followed by a discussion about the optimization solution methods, such as genetic algorithm, in the third subsection.

The optimal risk hedging problem considers two conflicting objectives, that is, pay now in the initial bid price that is inflated for the risk premium versus pay later in the risk exposure during construction. Barrier levels for commodity prices that are considered as decision variables, balance between these two objectives.

There are two general approaches to address this optimization problem. The first approach moves one of these two objectives to the constraint set, which needs to be pre-established. This method can be rather arbitrary. In this case, an optimization method would return a single solution rather than a set of solutions that can be examined for trade-offs.

However, decision makers often prefer a set of good solutions that consider multiple objectives simultaneously. This is the second approach. For that, an entire Pareto optimal solution set is determined, where the solutions are nondominated with respect to each other. The Pareto optimal solution sets are often preferred to single solutions since the final solution left to the decision-maker is to make the tradeoff (Konak *et al.*, 2006). Thus, the optimization problem in this study can be approached in the following two ways: (1) Minimize risk premium by tolerating potential future losses; or (2) Minimize both the risk premium and the future exposure loss, and conduct tradeoff afterwards. The former is known as a single-objective optimization problem with loss constraints, while the latter is known as a multi-objective optimization problem.

6.1 Single-Objective and Multi-Objective Optimization

The fundamental difference between single-objective and multi-objective optimization problems is that the solution in single-objective optimization is the single optimum solution, whereas in multi-objective optimization, a number of optimal solutions arise because of the trade-offs between the conflicting objectives (Deb, 2001).

Single-objective optimization in this study is considered as a degenerate case of the multi-objective formulation (Deb, 2001). From the author's perspective, this is not only because its outcome has just one decision solution, but also due to the subjective nature of the constraint before the search algorithm is initiated. Consider a single-objective optimization case - "minimize risk premium by tolerating potential future losses". The upper bound on the tolerance level of risk exposure (e.g., the maximum loss that the agency would like to afford) is required. However, if the upper bound is not chosen appropriately, the feasible set might be empty, that is, there may be no solution to the corresponding single-objective optimization problem for a pre-determined tolerance level. Thus, a suitable range of values for the input parameter should be known beforehand in order to avoid this situation.

The principle for the multi-objective optimization procedure is shown in Figure 6.1 (Deb, 2001). It consists of two steps:

Step 1: Find multiple trade-off optimal solutions with a wide range of values for objectives. Each trade-off solution corresponds to a specific order of importance of the objectives. For example, an agency sets a two-objective optimization problem – minimize both the initial project cost and the future risk exposure. Then, a chosen algorithm is used to solve this optimization problem, resulting in a set of optimal solutions. Thus, the task of step 1 is to find as many different trade-off solutions as possible;

Step 2: Choose one of the obtained solutions using higher-level information. Once a well distributed set of trade-off solutions is found, in step 2 one needs to choose a solution. For example, since the result from step 1 gives a set of optimal solutions according to different levels of initial project cost and risk exposure, the agency's risk preference, which represents the attitude to taking on risk, will act as the higher-level information. This information is used to evaluate and compare each of the obtained trade-off solutions. This higher-level information helps the agency choose one optimal solution at a specific level of risk exposure and the corresponding initial project cost.

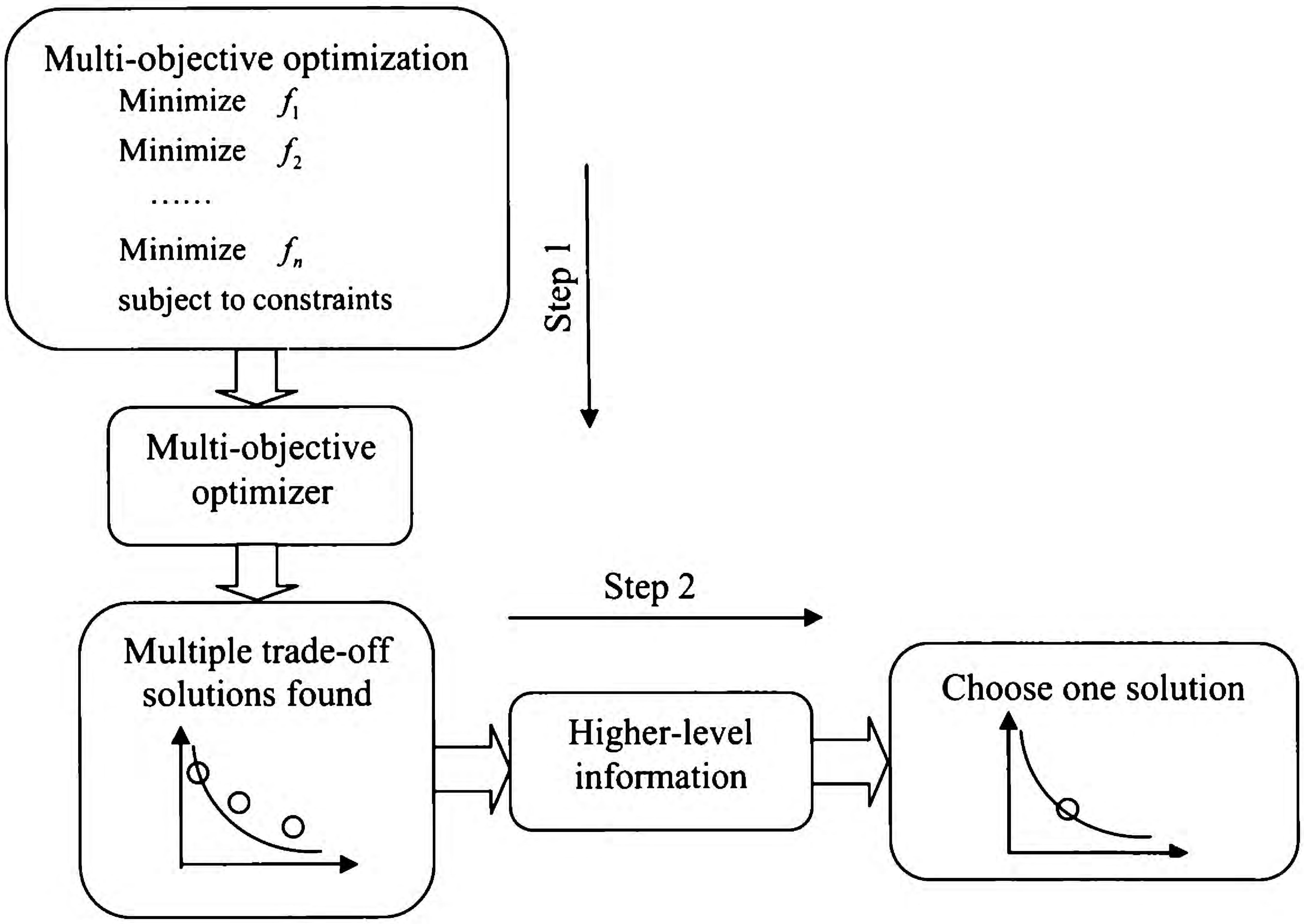


Figure 6.1 Schematic of a multi-objective optimization procedure

In the case of multi-objective optimization, there exists a set of Pareto-optimal solutions or nondominated solutions (Srinivas and Deb, 1994). Nondominated solutions imply that there is no solution which is the "best" among the solution set in terms of all the objectives. The choice of the "best" solution requires higher-level information which is non-technical, qualitative and experience-driven from a practical perspective. In this report, the prior knowledge or decision-makers' subjective selection would be the agencies' specified risk preferences in terms of their tolerance levels of risk exposure in the future if escalation clause is added.

6.2 Formulation of Optimization Models

Due to the previously identified issues, the report suggests using multi-objective optimization for optimal hedging of commodity cost risks, that is, "make a trade-off between the risk premium minimization and the future exposure loss minimization". The primary purpose of this subsection is to formulate this problem.

The multi-objective optimization problem has two conflicting objectives to be minimized: *Objective 1*—minimize the "unexpected losses" (see the definition in subsection 2.3) due to fluctuations in commodity prices if an escalation clause is added; and *Objective 2*—minimize the initial project cost (bidding cost), that is, minimizing the risk premium that is added by contractors. The formulation of optimization on both control-item level and project level are shown as follows.

Control-item level:

The multi-objective optimization formulation on control item-level is stated as in Equation 6.1:

$$\begin{aligned} &Obj\ 1:\ \text{minimize} \quad f_m(x)=\beta_m(x)\times CVaR_m(x)=\frac{F_m(x)}{N}\times\frac{S_m(x)}{F_m(x)} \quad m=1,\cdots,M \\ &Obj\ 2:\ \text{minimize} \quad g_m(x)=x\times ERP_m \\ &\qquad\quad \text{subject to} \quad x^L\le x\le x^U \end{aligned} \tag{6.1}$$

where x represents the decision variable – barrier level; x is restricted to be within a lower bound x^L and a upper bound x^U; M stands for the number of control items; N represents the number of scenarios for simulation; ERP_m represents the estimated risk premium for control item m (a constant calculated using the regression model discussed in Section 4) that is included in the unit bid without adding escalation clause; $F(x)$ counts the number of the simulated losses greater than VaR; $S(x)$ is the sum of losses greater than VaR.

As shown in objective 1, $f(x)$ represents the "unexpected loss" which is a product of $CVaR(x)$ (i.e., expected loss over distribution tail) and the probability of loss greater than VaR - $\beta(\mathrm{x})$. Here, $CVaR(x)$ and $\beta(x)$ are both dependent on the barrier level x. The motivation for specifying such a new "synthetic" objective function lies in the fact that "unexpected loss" characterizes both the conditional risk exposure and the probability that this risk realizes (losses exceed VaR).

To highway agencies, minimizing an initial project budget is equivalent to minimizing a risk premium, as the risk premium directly depends on the barrier of the escalation clause. It is assumed that the risk premium is linearly and positively related to the trigger barrier (see Section 4). The higher the barrier level (percentage of crude oil spot price is higher than the initial spot price), the lower the risk that the owner takes, and vice versa. The barrier is constrained from 0 percent to 100 percent, where the highest barrier is assumed to be a 100 percent increase from the initial price. Note that owner's risk preference regarding both the initial setting of VaR and the risk tolerance levels of "unexpected loss" affect the result of the decision variable (barrier level) in this multi-objective optimization problem.

Additionally, a single-objective optimization formulation on a control-item level is presented in Equation 6.2.

$$\begin{aligned} &\text{minimize} \quad g_m(x) = x \times ERP_m \qquad m = 1,\cdots,M \\ &\text{subject to} \quad f_m(x) = \beta_m(x) \times CVaR_m(x) = \frac{F_m(x)}{N} \times \frac{S_m(x)}{F_m(x)} \le \text{TL} \\ &\qquad\qquad x^L \le x \le x^U \end{aligned} \tag{6.2}$$

where all the functions and parameters in Equation 6.2 are the same with Equation 6.1, except *TL* is the maximum loss tolerance pre-determined by an agency. The comparison between the results of both single-objective and multi-objective optimization will be discussed in Section 7 using a case study data.

Project level:

Firstly, the multi-objective optimization formulation on the project level with only one single barrier level for all the control items is stated as in Equation 6.3:

$$\begin{aligned} &Obj\ 1: \text{minimize} \quad f(x) = \beta(x) \times CVaR(x) = \frac{F(x)}{N} \times \frac{S(x)}{F(x)} \\ &Obj\ 2: \text{minimize} \quad g(x) = x \times TRP \\ &\qquad\quad \text{subject to} \quad x^L \le x \le x^U \end{aligned} \tag{6.3}$$

where *TRP* is the total risk premium for the project; all the other functions and parameters in Equation 6.3 are the same as the ones in Equation 6.1, except that they are on the project level, that is, the simulated losses for the project is the sum of the losses for each control item. Then, $F(x)$ and $S(x)$ count the number of the project losses greater than "project *VaR*" (the predetermined *VaR* for the project); $S(x)$ is the sum of project losses greater than "project *VaR*".

Secondly, the multi-objective optimization formulation on project level with multiple barrier levels for different risk factors is stated as in Equation 6.4:

$$
\begin{aligned}
&Obj\ 1:\ \text{minimize} \quad f(x) = \beta(x) \times CVaR(x) = \frac{F(x)}{N} \times \frac{S(x)}{F(x)} \qquad i = 1, \cdots, I \\
&Obj\ 2:\ \text{minimize} \quad g(x) = \sum_{i=1}^{m} x_i \times ERP_m \qquad m = 1, \cdots, M \qquad\qquad (6.4) \\
&\qquad\quad \text{subject to} \quad x^L \le x_i \le x^U
\end{aligned}
$$

where the decision variables x_i represents the optimal barrier level for a risk factor (i.e., price of a commodity); I stands for the number of risk factors that are considered in the optimization; $g(x_i)$ represents the sum of the risk premium for each control item; all the other functions and parameters in Equation 6.4 are the same with the ones in Equation 6.3.

6.3 Solving Multi-Objective Optimization Problems

General multi-objective optimization problem (MOP) solution methods range from classical approaches to population based techniques. Most classical (i.e., point-by-point) algorithms, such as, direct and gradient-based methods, use a deterministic procedure in approaching the optimum solution, there are some common difficulties with most classical techniques, such as, the convergence to an optimal solution depends on the chosen initial solution. In addition, since nonlinearities and complex interactions among problem variables often exist in real-world problems, the search space usually contains more than one optimal solution. While solving these problems, when classical methods get attracted to any of the locally optimal solutions, there is no escape (Deb, 2001).

To solve MOP in an acceptable timeframe, evolutionary algorithms were developed. Evolutionary algorithm (EA) is a class of stochastic optimization methods that simulate the process of natural evolution. The origins of EAs for solving MOP can be traced back to the late 1950s (Veldhuizen and Lamont, 2000). Since the 1970s, several evolutionary methodologies have been proposed, mainly genetic algorithms, evolutionary programming, and evolution strategies. These algorithms have been proven to be a general, robust and powerful search mechanism. EAs are especially suited well to multi-objective optimization problem as they are able to capture multiple Pareto-optimal solutions in a single simulation run. Also, they can easily deal with concave Pareto fronts (Zitzler, 1999; Coello Coello, 2006). Other stochastic optimization techniques can also be used to generate the Pareto set (such as, ant colony optimization). However, these solutions very often do not guarantee to identify the optimal trade-offs (Abraham and Jain, 2005).

Genetic algorithms (GAs) have been the most popular EA approach to multi-objective design and optimization problems (Deb, 1999a; Konak *et al.*, 2006). GA has been identified to outperform conventional optimization methods especially when applied to difficult real-world optimization problems with non-convex, discontinuous, and multi-modal solutions spaces (Zheng *et al.*, 2004).

The multi-objective model is evaluated using the "gamultiobj" toolbox in Matlab. The multi-objective GA function "gamultiobj" uses a controlled elitist genetic algorithm (a variant of NSGA-II [Srinivas and Deb, 1994]). An elitist GA always favors "individuals" with better fitness value (rank) whereas, a controlled elitist GA favors "individuals" that can help increase the diversity of the population even if they have a lower fitness value. It is very important to maintain the diversity of population for convergence to an optimal Pareto front. This is done by controlling the elite members of the population as the algorithm progresses (MathWorks, 2011). The NSGA-II has become a benchmark against which other multi-objective evolutionary algorithms are compared to (Coello Coello, 2006), and it is implemented in this report.

6.4 Effect of Correlations on Optimal Hedging of Commodity Risks

Strong correlations between commodity prices have been observed in Section 5. Since the volatile commodity prices influence the future losses, the effect of the correlations on optimal hedging of commodity cost risks (such as, the impact on choosing optimal barrier level for escalation clause) should be investigated. This will be approached by comparing the optimization results under two situations: 1) univariate, and 2) vector representation of commodity prices. It is noted that the formulation of the multi-objective optimization remains the same, the only difference is in the change of commodity pricing models (e.g., ARIMA vs. VARMA).

6.5 Summary

This section presents the formulation and solutions approach to both a single-objective and a multi-objective optimization model. Genetic algorithm is considered as the approach to optimization problems. The principles of the difference between single-objective and multi-objective optimizations are discussed. While the formulations of optimization are presented, the genetic algorithm application to multi-objective optimization is discussed. Next section presents a case study based on a real highway construction project in Texas to show agencies the optimal hedging strategies dealing with commodity price risks.

7. CASE STUDY

This section presents a case study to illustrate the overall methodology. The case study is based on an actual TxDOT project let in 2004. The project is described in the first subsection. The analysis of the effects of different contract specifications on the risk premium and the future exposure is presented in the second subsection. This includes project characteristics data, contract specifications, and bid data for all control items. Finally, in the third subsection, the optimal risk hedging strategies for commodity prices are formulated and discussed. This includes implementation of both multi-objective and single-objective optimization models, and the consideration of the effect of correlations between risk factors (i.e., commodity prices).

7.1 Project Description

This TxDOT project was let in 2004. The bidding information and project characteristics is collected using thirty-four control items TxDOT reports in the database. The aggregation of bid data was done using project unique contract control-section-job (CCSJ) number. This project CCSJ number was 038001064 and the project was located in Lubbock (northwest part of Texas as highlighted in Figure 7.1). The selected lowest bid contractor was Granite Construction Company.

The project was to convert a non-freeway road to a freeway type of road. The major characteristics of the project are summarized in Table 7.1. This project was large in that the total bid was $200 million and it lasts over two years. The project consists of nineteen control items which are listed in Table 7.2.

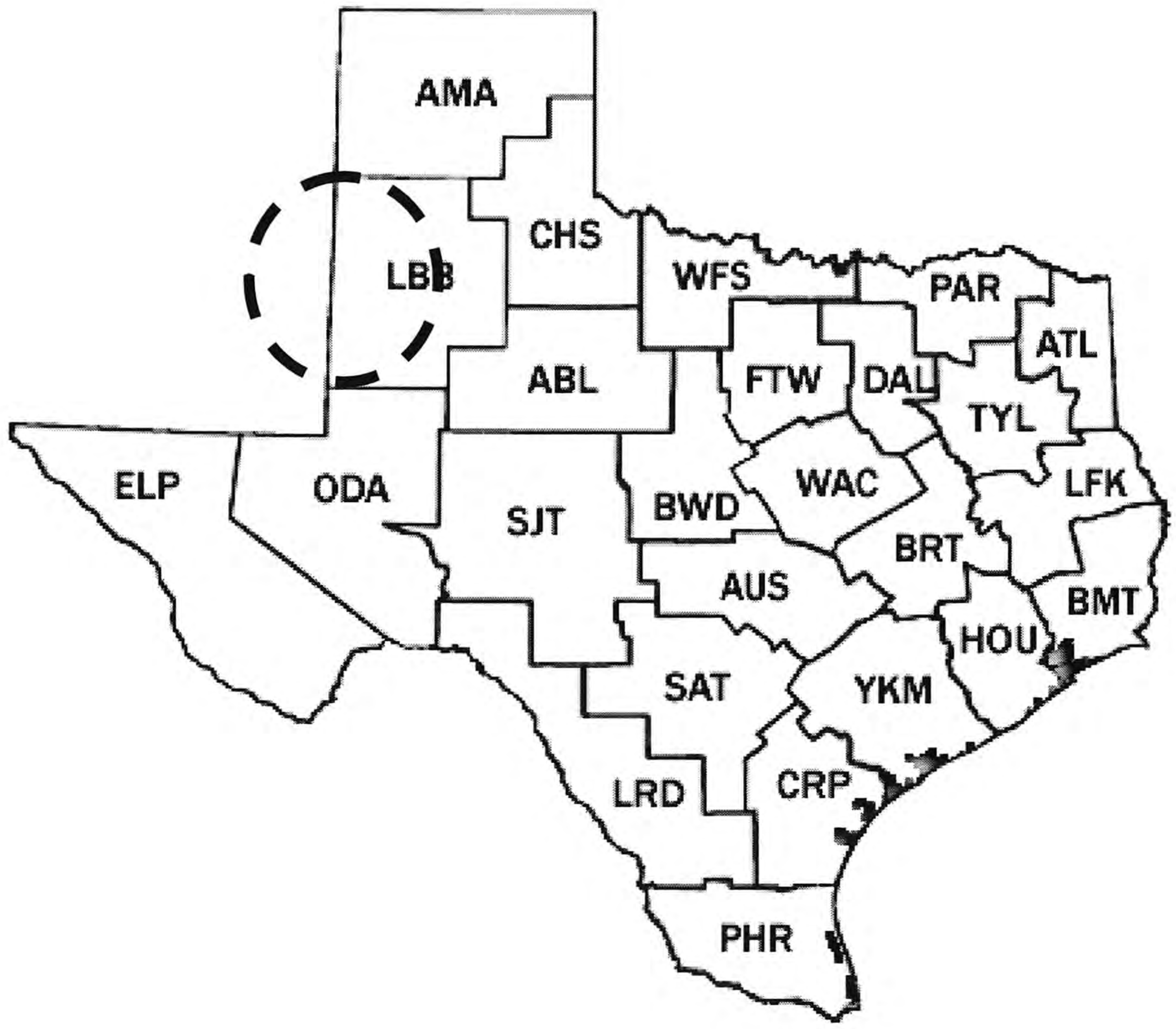


Figure 7.1 Project location at Texas

Table 7.1 Project description (1)

Letting date	County	Duration (day)	Number of bidders	Roadway	Total bid	Length (mi)
12/07/2004	Lubbock	860	4	US 82	$191,404,376	4.277

Table 7.2 Project description (2) – control items

Control item	Item	Item Description	Quantity	Unit	Bid price
ROADWAY EXCAVATION	**01100501**	**EXCAVATION (RDWY)**	**1576453**	**CY**	**3**
	01100503	EXCAVATION (SPECIAL)	500	CY	20
ROADWAY EMBANKMENT	**01320506**	**EMBANK (DENS CONT)(TY B)(CL 3)**	**893984**	**CY**	**3**
	01320512	EMBANK (DENS CONT)(TY D)(CL 3)	350223	CY	3
	01320521	EMBANK (ORD COMP)(TY C)(CL 3)	1008	CY	20
	01320546	EMBANK (TY D)(ORD COMP)(RANDALL CLAY)	9430	CY	5
	01320547	EMBANK (TY D)(DC)(SGMP)	16977	CY	5
SURFACE TREATMENT ASPHALT	03160862	ASPH (AC-20-5TR)	29330	GAL	1.5
SURFACE TREATMENT AGGREGATE	03160774	AGGR (TY PB GR7)	559	CY	80
HOT MIX ASPHALTIC CONCRETE	31460786	HOT MIX (TY D)(SURF)(PG 70-28)	1481	TON	50
	31460787	**HOT MIX (TY B)(BASE)(PG 70-28)**	**129083**	**TON**	**40**
CONTINUOUS REINFORCED CONCRETE PAVEMENT	03600503	CONC PAV (CONT REINF HY STL)(8 ")	47162.17	CY	121.51
	03600505	**CONC PAV (CONT REINF HY STL)(10")**	**77640.66**	**CY**	**107.99**
CLASS A CONCRETE	04200582	CL A CONC (ENCASEMENT)	56	CY	150
CLASS C CONCRETE	04200559	CL C CONC (FLUME)	36.6	CY	800
	04200692	CL C CONC (ABUT)(HPC)	2257	CY	600
	04200693	CL C CONC (BENT)(HPC)	5877.5	CY	650
CLASS S CONCRETE	04200518	CL S CONC (SLAB)	208.2	CY	500
	04200520	CL S CONC (SHEAR KEY)	75.7	CY	150
	04200747	CL S CONC (SLAB) (HPC)	1485.6	CY	500
BRIDGE RAIL(RIGID)	04500505	RAIL (TY T501)	22439.7	LF	27
	04500506	RAIL (TY T502)	900	LF	27
	04500531	RAIL (TY C411)(MOD)	7068.9	LF	150
	04500538	RAIL (TY T501)(MOD)	4614	LF	32
	04500683	RAIL (TY PR1)	57	LF	60
	04500695	RAIL (TY C411)	852.4	LF	130
	04500702	RAIL (PEDESTRIAN)(SPL)	2732.8	LF	250
	04500719	RAIL (TY T4)	142	LF	100
	04500828	RAIL (CONC WALL RAIL)	2442	LF	85
BRIDGE SLAB	04220504	REINF CONC SLAB (HPC)(CL S)	399811	SF	12
METAL FOR STRUCTURES	04420502	STRUCT STL-HS	7226100	LB	1.3
	04420646	STR STL (ARMOR JOINT)(SPL)	46537	LB	2.5
	04420654	STR STL (ARMOR JT) (SPL)(SIDEWALK)	1560	LB	5
	04420664	STR STL (SCUPPER)	684	LB	2.5
REGULAR BEAMS	**04250507**	**PRESTR CONC BEAM (TY IV)**	**24641.99**	**LF**	**105**
	04250594	PRESTR CONC U-BEAM (U40)	7849.48	LF	240

	04250598	PRESTR CONC BEAM (U54)	1583.57	LF	240
BOX BEAMS	04250520	PRESTR CONC BOX BEAM (4 B 20)	720.09	LF	125
	04250543	PRESTR CONC BOX BEAM (4 B 40)	1995.12	LF	240
DRILLED SHAFTS	04160503	DRILL SHAFT (18 ")	2442	LF	50
	04160504	DRILL SHAFT (30 IN)	4654	LF	70
	04160506	DRILL SHAFT (36 IN)	16707	LF	90
	04160508	DRILL SHAFT (48 IN)	1866	LF	150
	04160509	DRILL SHAFT (54 IN)	374	LF	180
	04160510	DRILL SHAFT (60 IN)	76	LF	210
	04160514	DRILL SHAFT (NON-REINF)(12IN)(SIGN MTS)	70	LF	50
	04160515	DRILL SHAFT (24 IN)(SIGN MTS)	10	LF	100
	04160517	DRILL SHAFT (36 IN)(SIGN MTS)	500	LF	150
	04160519	DRILL SHAFT (48 IN)(SIGN MTS)	17	LF	300
	04160520	DRILL SHAFT (54 IN)(SIGN MTS)	301	LF	350
	04160521	DRILL SHAFT (24 IN)	936	LF	60
	04160524	DRILL SHAFT (60 IN)(HIGH MAST POLE)	448	LF	400
CORRUGATED METAL PIPE	04600503	CMP (GAL STL 18 IN)	581	LF	25
REINFORCED CONCRETE PIPE	04640503	RC PIPE (CL III)(18 ")	244	LF	40
	04640505	RC PIPE (CL III)(24 ")	14237	LF	50
	04640507	RC PIPE (CL III)(30 IN)	1495	LF	70
	04640509	RC PIPE (CL III)(36 IN)	1350	LF	90
	04640510	RC PIPE (CL III)(42 IN)	865	LF	130
	04640511	RC PIPE (CL III)(48 IN)	196	LF	150
	04640520	RC PIPE (CL IV)(24 IN)	835	LF	60
CONCRETE REPAIR	04320501	RIPRAP (CONC)(CL B)	33.2	CY	400
	04320517	RIPRAP (CONC)(CL B)(DITCH LINING)	4783.7	CY	300
	04320524	RIPRAP (CONC)(CL B)(4 IN)	1324	CY	300
	04320529	RIPRAP (CONC)(CL B)(5 IN)	1718	CY	350
RETAINING WALLS	04230501	RETAINING WALL (MSE)	145743	SF	25
	04230511	RETAINING WALL (SOIL NAILED)	14954	SF	35
	04230523	RETAINING WALL (TIEBACK)	244282	SF	43.25

As described in the project description, the duration of the project (shown in Table 7.1) and quantities of the highlighted control items (shown in the fourth column of Table 7.2) to be investigated are comparably large. Also as observed from the historical time series in Figure 5.11, all the investigated commodity prices (cement, steel, and oil prices) were volatile and forecasted to keep increasing when the project was let. In fact, as the contract length affects the risk, i.e., the longer the contract duration, the more significant volatility becomes, both contract duration and indicated market-volatilities of commodity prices likely affected unit bid prices. Hence, TxDOT could have used the escalation clauses on the specific control items to reduce the premium and take on the risk. This case study aims to investigate what would have happened if they did, and what should have been the optimal strategy, of course, only using the information that was available at that time.

The five bid items (highlighted in Table 7.2) of this project are chosen for the analysis. The cost of these five control items accounts for 30.26 percent of the total cost of the project.

Currently TxDOT does not allow for the adjustment clauses. Based on the models developed in Section 4, the risk premiums are calculated and shown in Table 7.3. The estimated risk premiums in Table 7.3 come from the estimation results in subsection 4.2.3. The total risk premium for these five bid items was estimated to be \$3,139,367. Value at Risk for each item to compute "unexpected loss" in this study is arbitrarily specified (this should be specified by agencies according to their higher-level information, such as risk preferences).

Table 7.3 Five selected items for the case study

Control item	Quantity	Unit bid price	Risk factor	Calculated risk premium
Excavation	1576453	3	Oil price	\$613,248
Embankment	893984	3	Oil price	\$259,598
Hot mix asphaltic concrete	129083	40	Oil price	\$604,387
Continuous reinforced concrete pavement	77640.66	107.99	Cement price	\$881,523
Regular beams	24641.99	105	Steel price	\$780,611

7.2 Effect of Barriers on Risk Premium and Future Exposure

Let's consider the excavation control item. The contractor added approximately \$610,000 (based on the model) in anticipation of the changes in oil prices. If TxDOT decided to allow for escalation clauses and select "Barrier level = x", then the question is what losses should be expected in the future. This can be investigated by simulating oil prices.

Figure 7.2 shows the interactions among losses, premiums and barrier levels. Based on the estimated coefficient in subsection 4.2.3.1, every 1 percent increase in the expected change will result in (on average) an 8.4 percent increase in unit bid prices of excavation. Then, with simulating the future spot price of oil, the "expected losses" (the average loss of the whole distribution of simulated losses) given different levels of barriers can be simulated for excavation item. Further, the "unexpected losses" (a product of *CVaR* and the probability of loss greater than *VaR*, that is, the average of the losses greater than VaR) can be obtained as well, if a VaR level is pre-determined. As indicated by Figure 7.2, 1) the "unexpected loss" accounts for the most part of the losses, since it focuses on the worst cases of the "expected losses"; 2) there is an optimal barrier for risk hedging given the risk preferences of the agency, that balances between initial cost and the future exposure. Hence, at the time of letting, TxDOT should have evaluated the effectiveness of adding escalation clauses with barrier levels, based on the information available at that time.

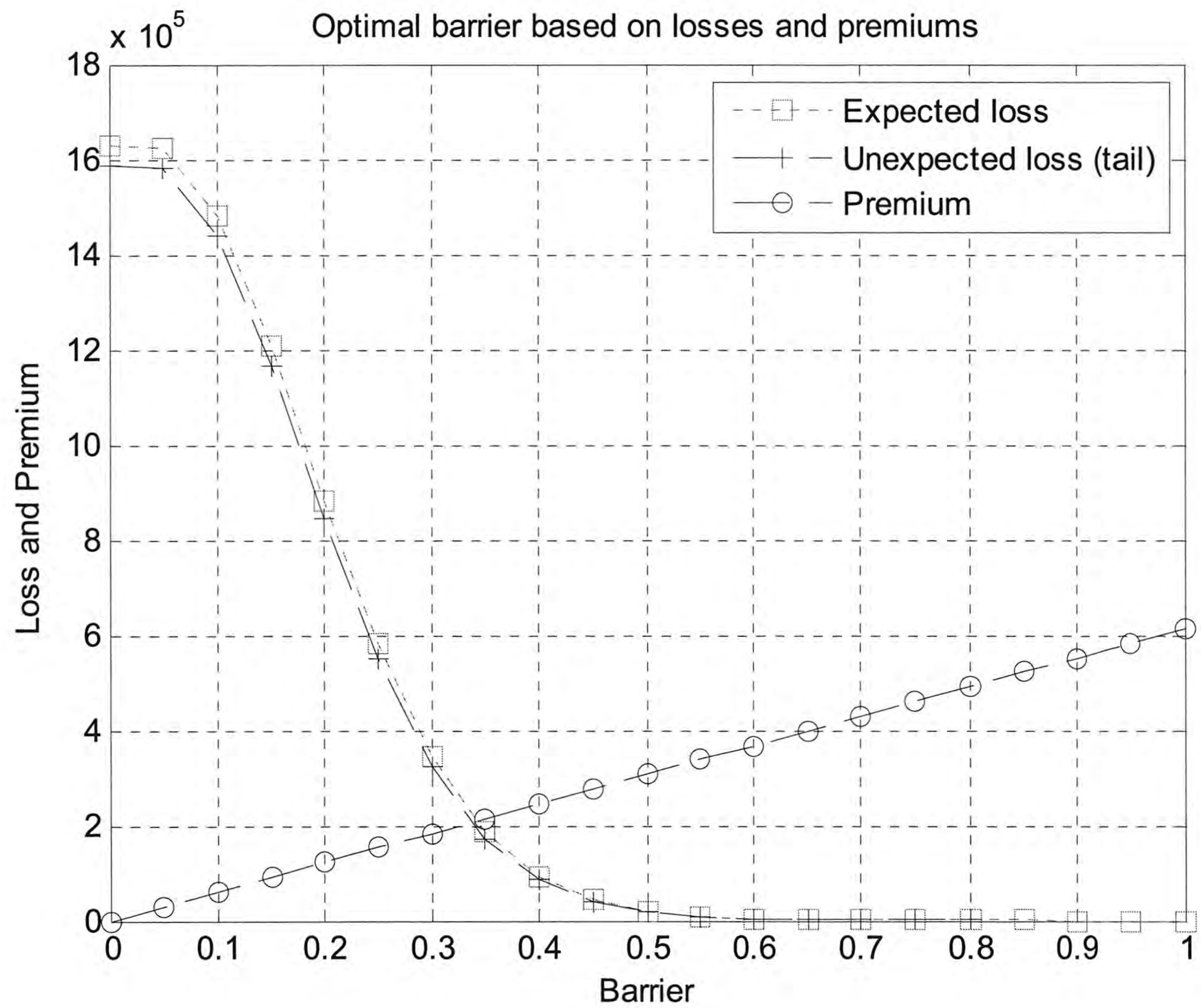


Figure 7.2 The effect of barriers on risk premium and exposure

7.3 Optimal Risk Hedging

Given the information in Table 7.3, the case study further investigates the following: (1) The losses for different control items and barrier levels, from both independent-risk and correlated-risk perspectives; (2) The optimal solution sets for control items, if they are optimized independently (i.e., optimizing the barrier levels for each control item, and simulating the commodity prices using ARIMA models); and (3) The overall optimal solution set from project-level perspective that considers correlated risks and optimizes the barriers on the project-level (i.e., optimizing the total project risks including the premiums for each item, and simulating commodity prices using VARMA model).

Control item-level analysis:

The simulated "expected losses" and the independent optimal solution set for each item are shown at first. The "expected losses" (if escalation clauses are added) are simulated based on univariate time series (TS) forecasting for each control item. The sample of simulation for "expected losses" is 100,000. If escalation clauses with triggers are added, the objective 1 and objective 2 of the multi-objective optimization are "minimize the 'unexpected loss'" and "minimize the trigger barrier (that is, to minimize the risk premium of the bids)", respectively. The parameters for multi-objective optimization in this section are: Sample=10,000, population size=60, Pareto front fraction=0.7. Additionally, the results of single-objective optimization problem, as discussed in Section 6, are presented to show why they are claimed as degenerative cases.

1) Excavation:

The "expected losses" simulation for excavation based on oil prices forecasted using univariate TS (ARIMA) is shown in Figure 7.3.

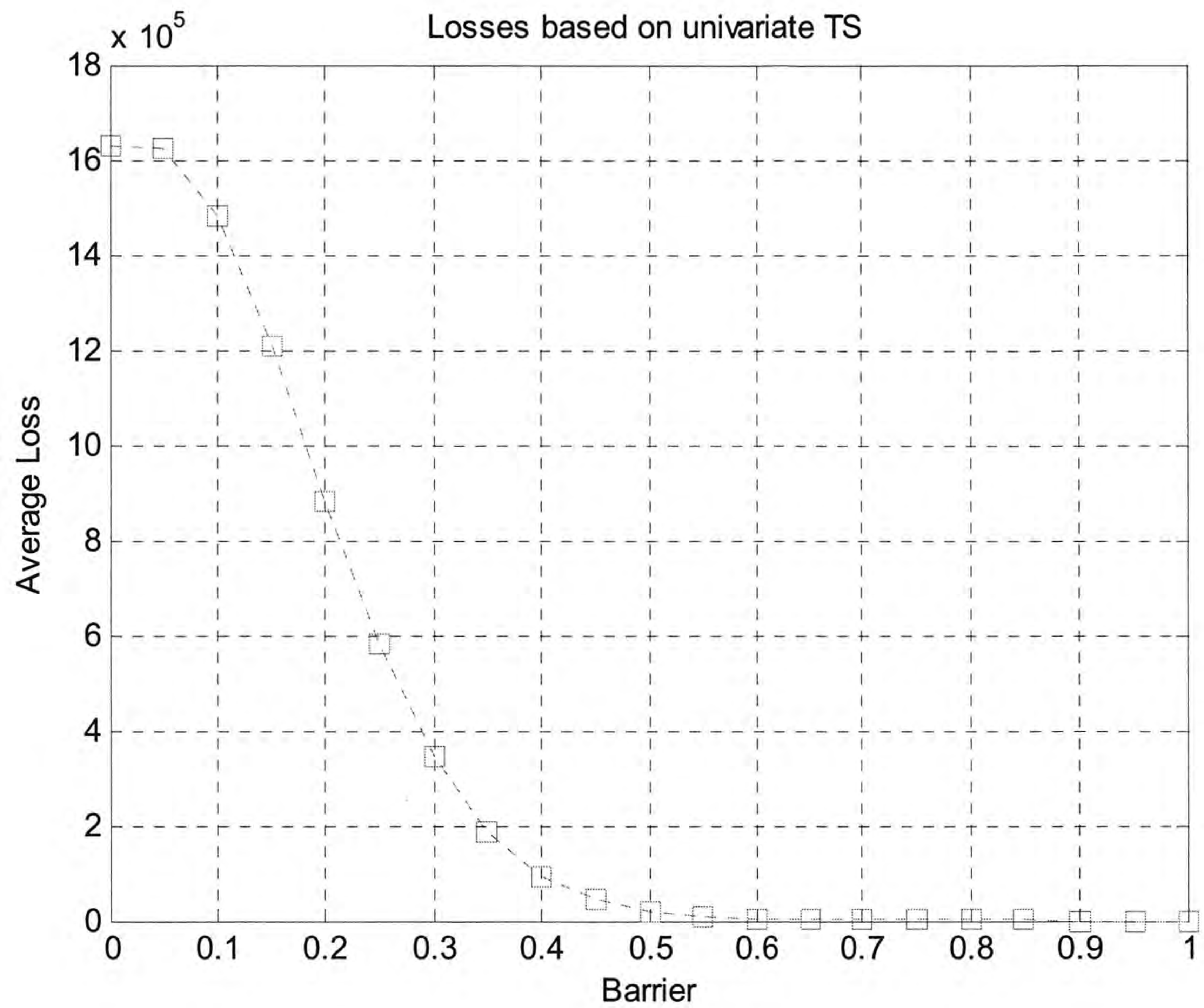


Figure 7.3 "Expected losses" for excavation

Figure 7.3 shows that the average loss converges to zero if the barrier level is greater than 50 percent. This also implies that there is no need for the agency to set the barrier over 50 percent if the optimization for control items is considered individually.

The result of multi-objective optimization for excavation with simulated oil prices based on univariate time series (ARIMA) is shown in Figure 7.4.

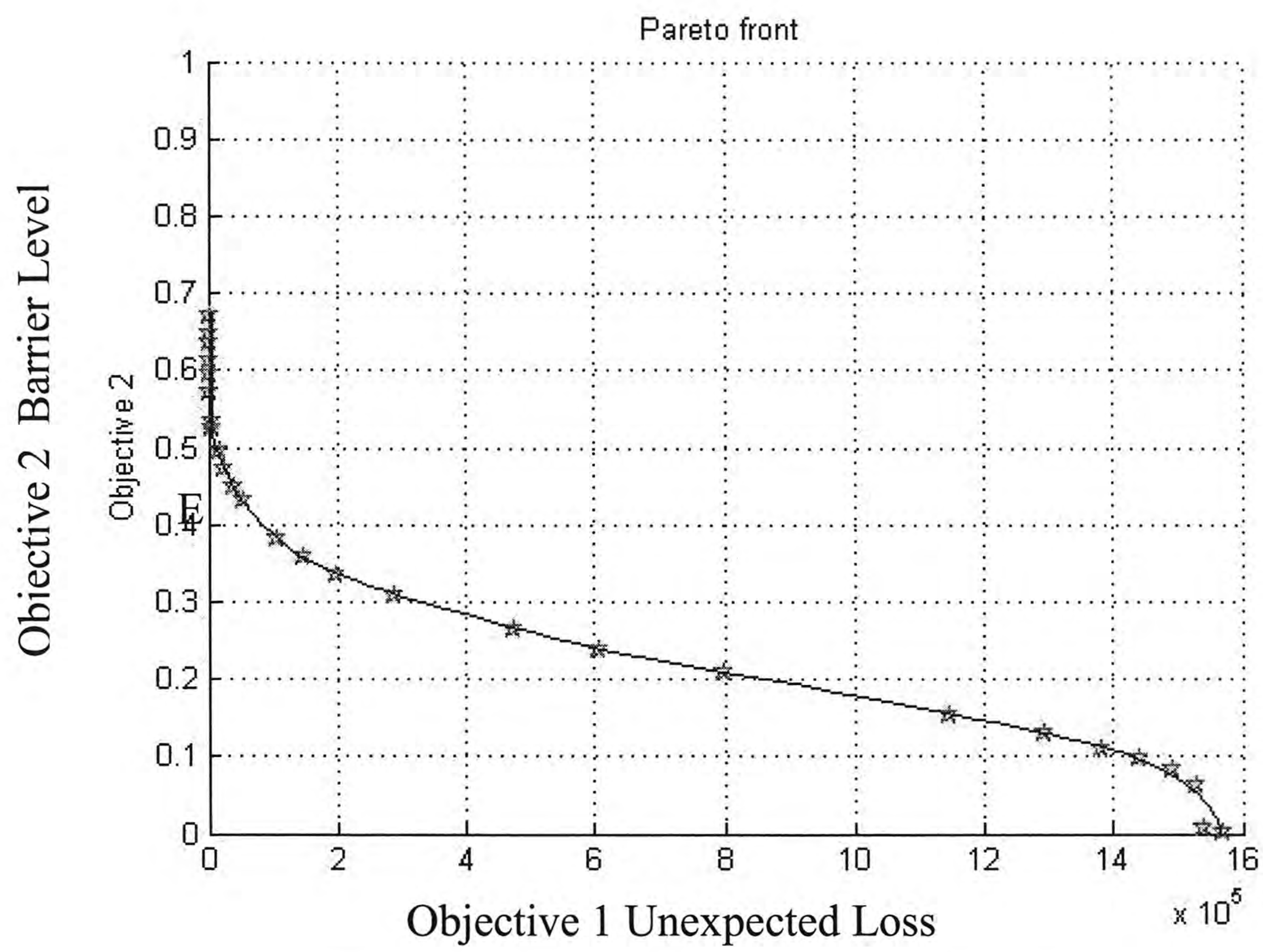


Figure 7.4 Pareto front for excavation (1)

The results shown in Figure 7.4 correspond to the overall optimal contracting strategy, where a Pareto front is associated with the decision variable – the barrier level x. For example, the decision-making at the point E corresponds to setting a trigger barrier level of 40 percent now (that is, the owner would probably pay 40 percent of the total risk premium now) while accepting a future "unexpected loss" of $80,000. Furthermore, it can be observed that as the cost of risk premium increases, the 'unexpected loss' (product of beta and CVaR) decreases, and vice versa. Further discussion about the implications of the Pareto front will be presented in the subsection 7.4.

2) Embankment and Hot Mix Asphaltic Concrete (HMAC):

As excavation, embankment and HMAC control items have the same risk factor – oil price, the optimization results of embankment and HMAC are similar to the results for excavation control item. The "expected losses" simulations and the Pareto front for embankment are shown in Figure F-1 and Figure F-2 (Appendix F); The "expected losses" simulations and the Pareto front for HMAC are shown in Figure F-3 and Figure F-4 (Appendix F).

3) Continuous Reinforced Concrete Pavement (CRCP):

The “expected losses” simulation for CRCP based on cement prices forecasted using univariate TS (ARIMA) is shown in Figure 7.5.

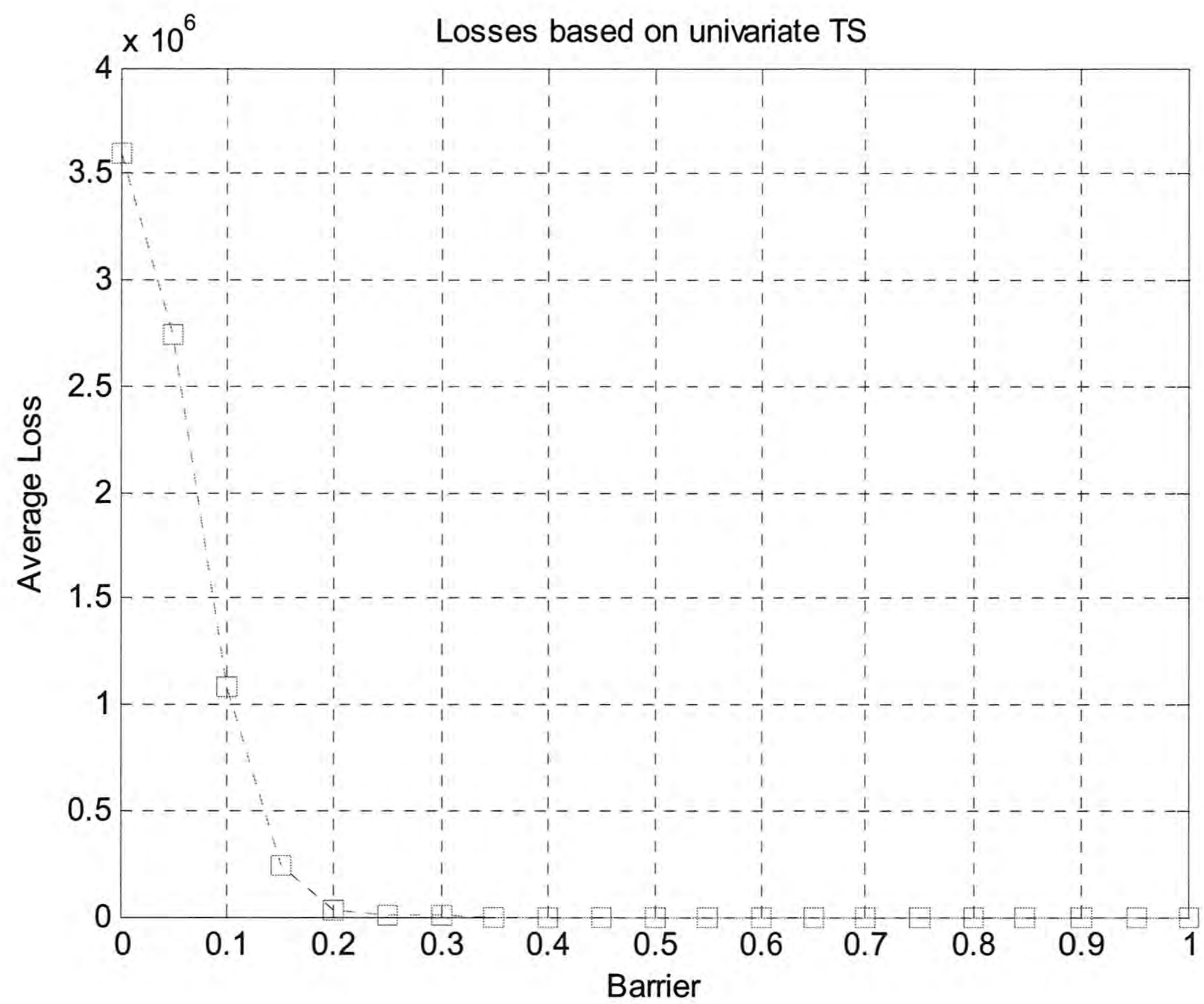


Figure 7.5 “Expected losses” for CRCP

Figure 7.5 shows that the simulated prices of steel will not increase to the level which is 25 percent higher than the original price. This is because the simulated losses become to be zero for the barriers which are greater than 25 percent. This may be because of the fact that cement price has experienced lower level of volatility compared to oil price. If the trigger barrier is set at 25 percent, then the TxDOT would have significantly reduced the risk premium, and at the same time, would not have been exposed to the loss during the construction.

The Pareto fronts of multi-objective optimization for CRCP control item with simulated cement prices based on univariate TS are shown in Figure 7.6.

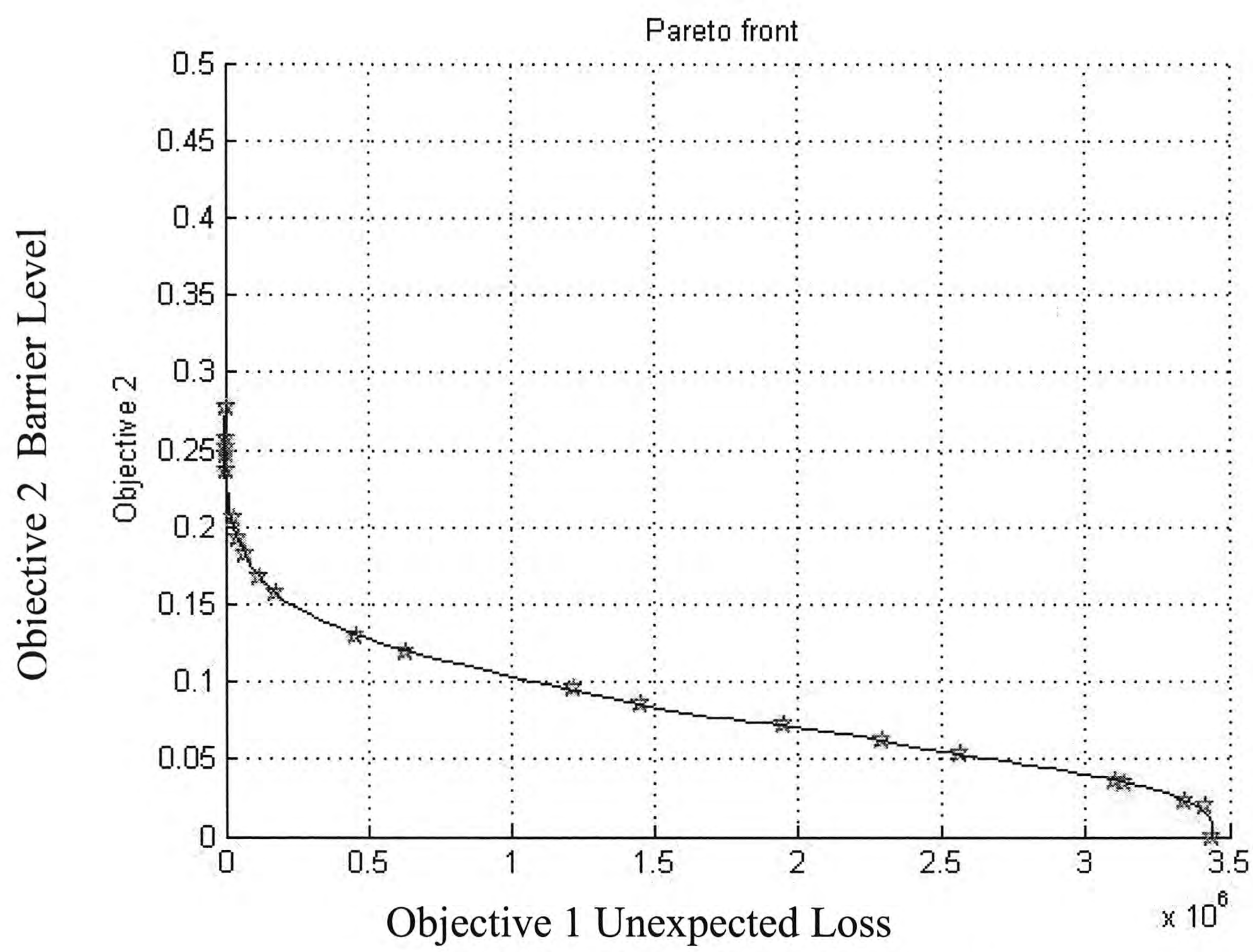


Figure 7.6 Pareto front for CRCP

Figure 7.6 also indicate the same trend as Figure 7.5. The "unexpected losses" converge to zero if the trigger barrier is set at 25 percent.

4) Regular Beams:

The "expected losses" simulation for regular beams based on steel prices forecasted using univariate TS (ARIMA) is shown in Figure 7.7.

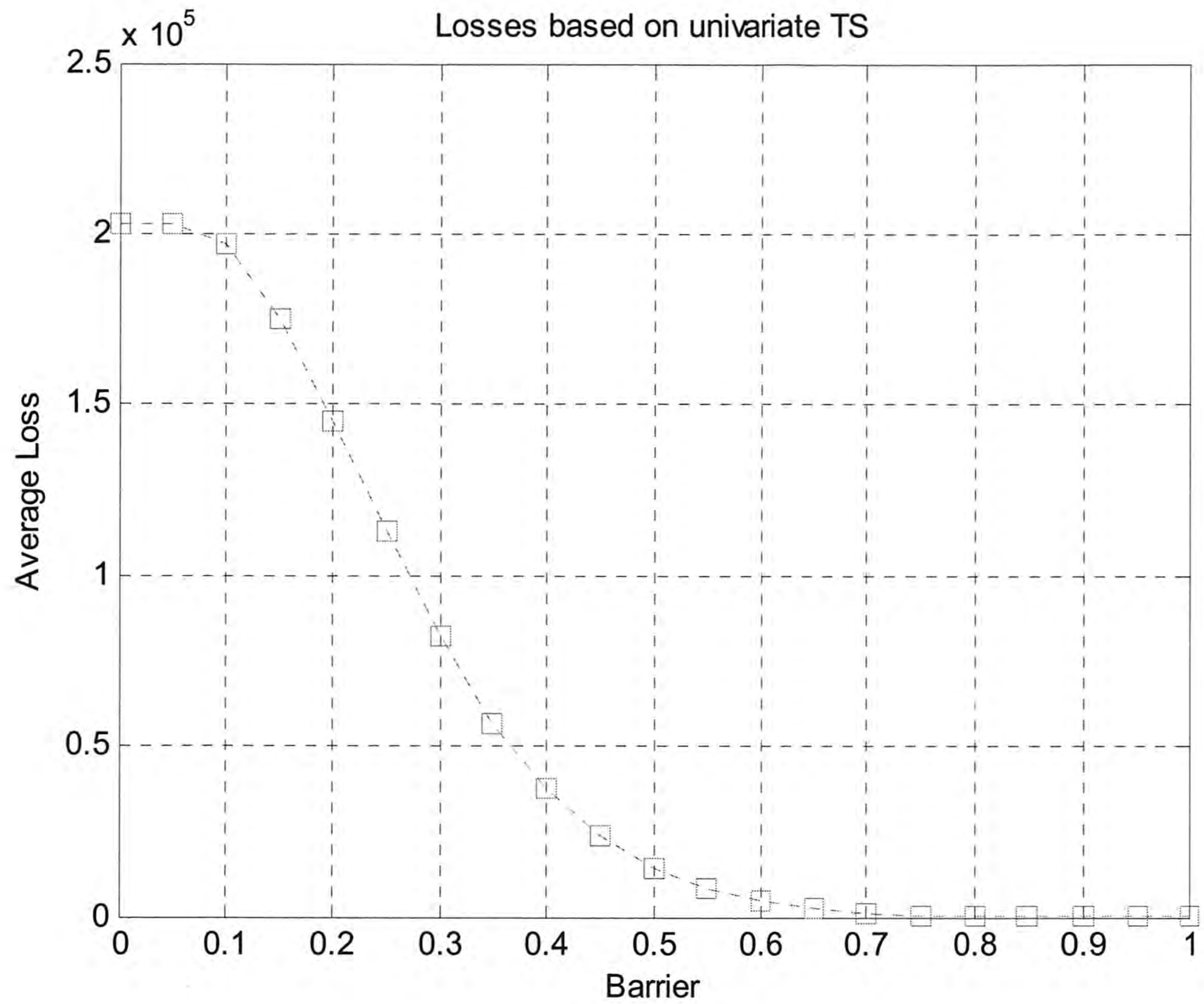


Figure 7.7 "Expected losses" for regular beams

Figure 7.7 shows that the average losses converge to zero if the barrier level is set to be over 70 percent, while Figure 7.8 shows the Pareto front for multi-objective optimization based on simulated steel prices using univariate time series. The front indicates that the "unexpected loss" would approach to zero if the barrier level is set to be around or over 70 percent. However, the Pareto front with barrier levels at 50 percent would be a "better" choice for risk hedging. This is not only because adding an escalation clause with barriers considerably reduces the initial cost (setting a barrier level at 50 percent reduces the risk premium for about 50 percent), but also because of the maximum "unexpected losses" are only $10,000.

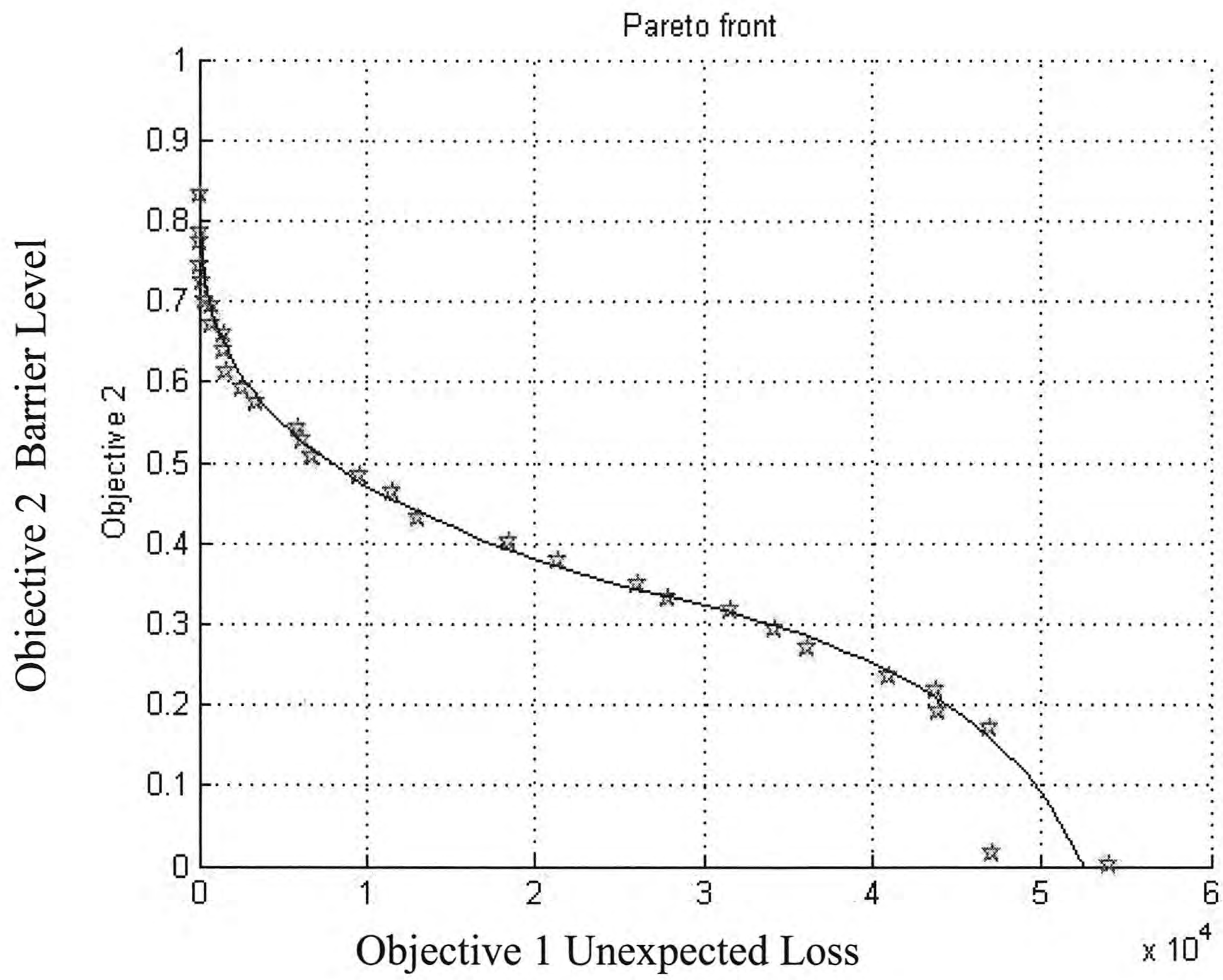


Figure 7.8 Pareto front for regular beams

The results of single-objective optimization are shown below in Table 7.4. Parameters for single-objective optimization use the default setting in the "ga" solver of optimization tool in Matlab (Mathworks, 2011). As discussed in Section 6, single-objective optimization requires prior knowledge about the constraint – the tolerance level of "unexpected loss" in this case. Again, this information is the subjective decision by TxDOT according to their risk preferences; however, in this example, the constraints are specified beforehand. Thus, if the upper bound is not chosen appropriately, the obtained feasible set might be empty. The result for each item shown in Table 7.4 is just a point solution from the Pareto front of multi-objective optimization as shown above.

Table 7.4 Optimal barriers of single-objective optimization for individual control items

Control item	Tolerance of "unexpected loss" ($)	Optimal barrier
Excavation	300,000	31.2%
Embankment	200,000	31%
HMAC	200,000	37.5%
CRCP	800,000	28%
Regular beams	40,000	30.4%

<u>Project-level discussion:</u>

Since the risks among all control items may be correlated, it is more realistic to conduct the multi-objective optimization on the project-level. The optimizations with a single barrier level for all control item (decision variable) and the optimization with multiple barrier levels for each control items are both discussed and compared next.

If TxDOT sets only one barrier for all the prices of commodities, then Figure 7.9 presents the overall optimal solution sets from the project perspective. The total risk premium ($3,139,367) is assumed to be the sum of the premium of individual control item as shown in the fifth column of Table 7.3. The Value-at-Risk ($2,500,000) used in the optimization is considered as the sum of VaRs of each control item.

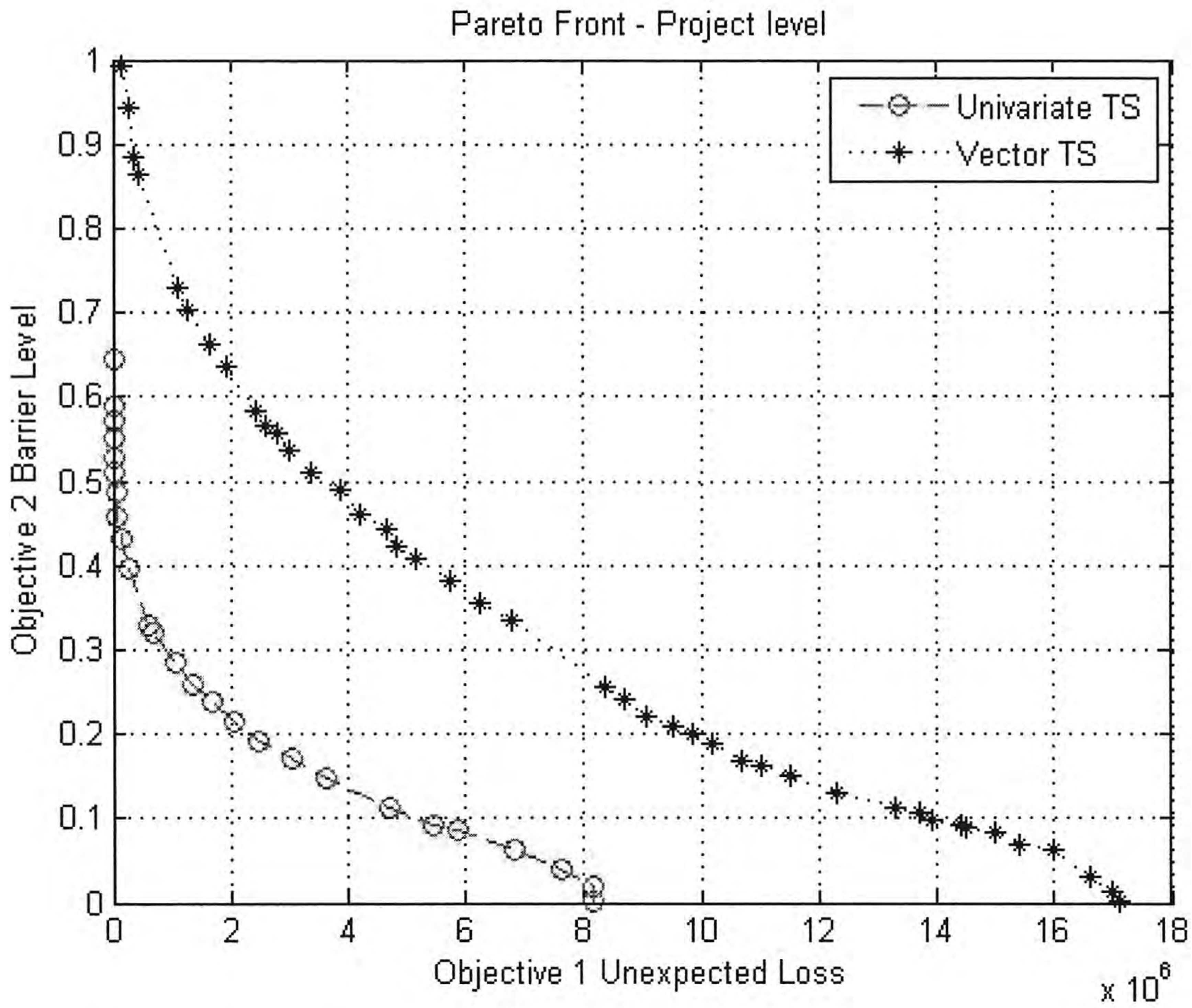


Figure 7.9 Pareto fronts on the project level – single barrier

Figure 7.9 further shows that the overall potential risk exposure of this project is considerably under-estimated if the commodity prices are forecasted independently. The effect of correlation between the prices of commodities was captured using the vector TS model developed in Section 5. It can be observed from Figure 7.9 that setting a conservative barrier level of approximate 45 percent corresponds to taking close-to-zero "unexpected loss" under the univariate TS model, but a $4,500,000 under the vector TS model.

While the example above investigates the optimal solutions with a single barrier level (decision variable) for all control items of the project, the project-level optimal solution sets with multiple barriers may provide better strategy as shown in Figure 7.10. The result indicates that optimization with multiple barriers can considerably reduces both the risk premium of the total project and the risk exposure during construction. For example, in Figure 7.10, if different escalation clauses are set for different commodities, taking on the risk of "unexpected loss" of $1,000,000 will require the agency to pay a risk premium of $1,000,000; TxDOT has to pay only $2,000,000 for the risk premium for close-to-zero "unexpected loss", compared to a risk premium of $2,700,000 based on "one barrier for all items" policy. Thus, it is suggested that

TxDOT considers the multi-objective optimization with multiple barriers to hedge against the risks of volatile prices of commodities.

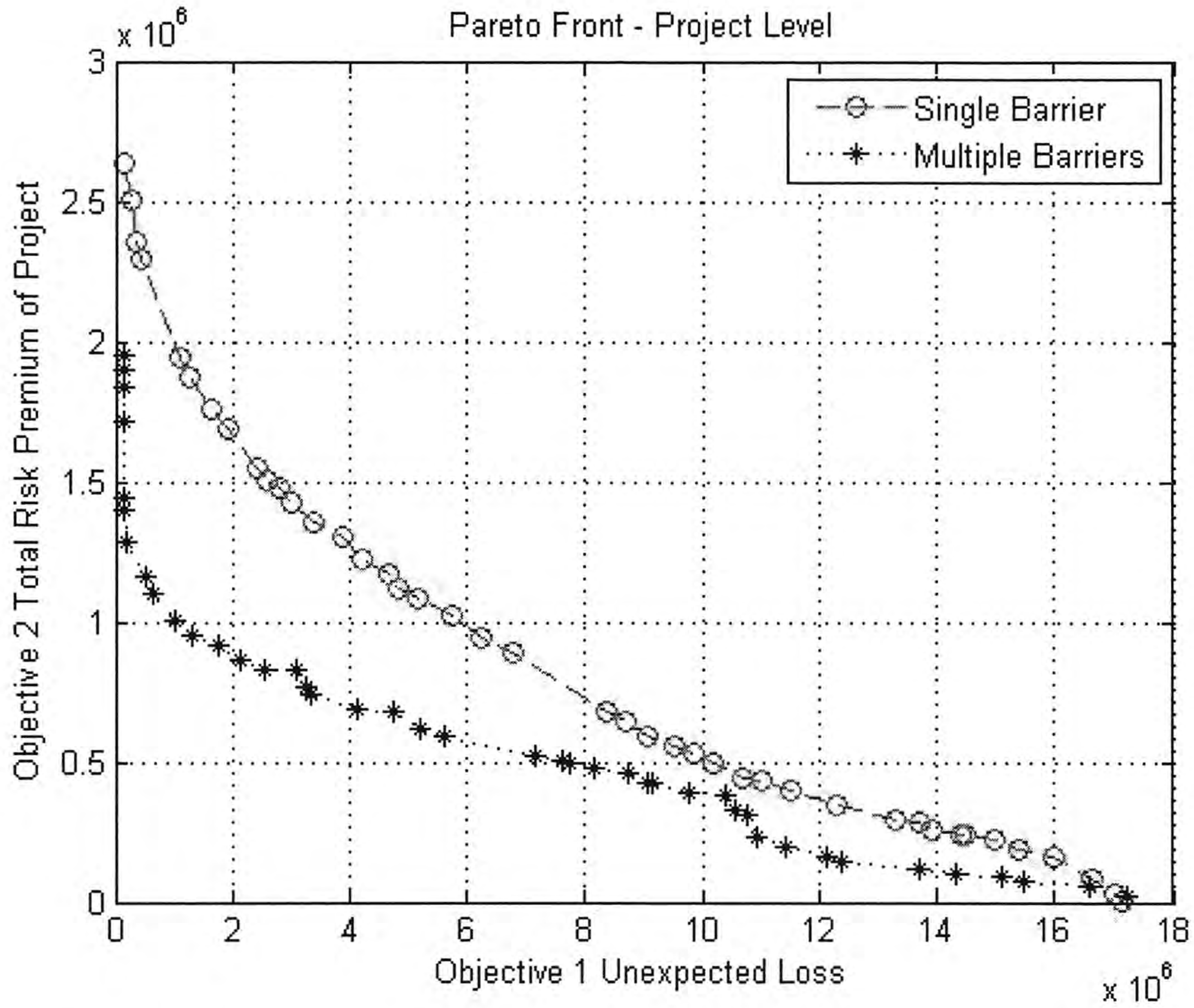


Figure 7.10 Pareto fronts on the project level – multiple barriers

It should be noted that all the Pareto fronts obtained above, either on control-item level or on project level, are associated with "unexpected loss"; the Pareto fronts associated with "CVaR" are presented in Appendix F, from Figure F-5 to Figure F-11.

The major findings for this case study are summarized as follows:

Control item-level:

1) For control items (excavation, embankment and hot mix asphaltic concrete) having the same risk factor - oil price, the optimization results based on the ARIMA models indicate that the trigger barrier setting around or just over 50 percent for oil prices will result in close-to-zero average and "unexpected losses". This means agencies can save half of the risk premium while taking on a close-to-zero risk;

2) Due to the comparably small volatility of cement price, the "unexpected losses" of regular beams are close to zero for the trigger barriers over 25 percent based on the ARIMA model. This means, for the control items (e.g. CRCP), where cement price is identified as the

most significant risk factor, adding an escalation clause with trigger barriers of 25 percent will reduce the bids while taking on a level of risk which is considerably lower.

Project-level:

1) The effect of correlations (between prices of commodities) should not be ignored. The potential risk exposure of the project is considerably under-estimated if the commodity prices are forecasted independently.

2) It is essential to consider multi-objective optimization with multiple barrier levels using the vector time series model when making optimal decisions to hedge against the risks from volatile prices of commodities. This is because the solutions from multi-variate optimization can considerably reduce the total cost of a project.

7.4 Implications

The developed models have a wide applicability for TxDOT. Following are some of the key insights from the models, using Figure 7.11 for excavation as an example.

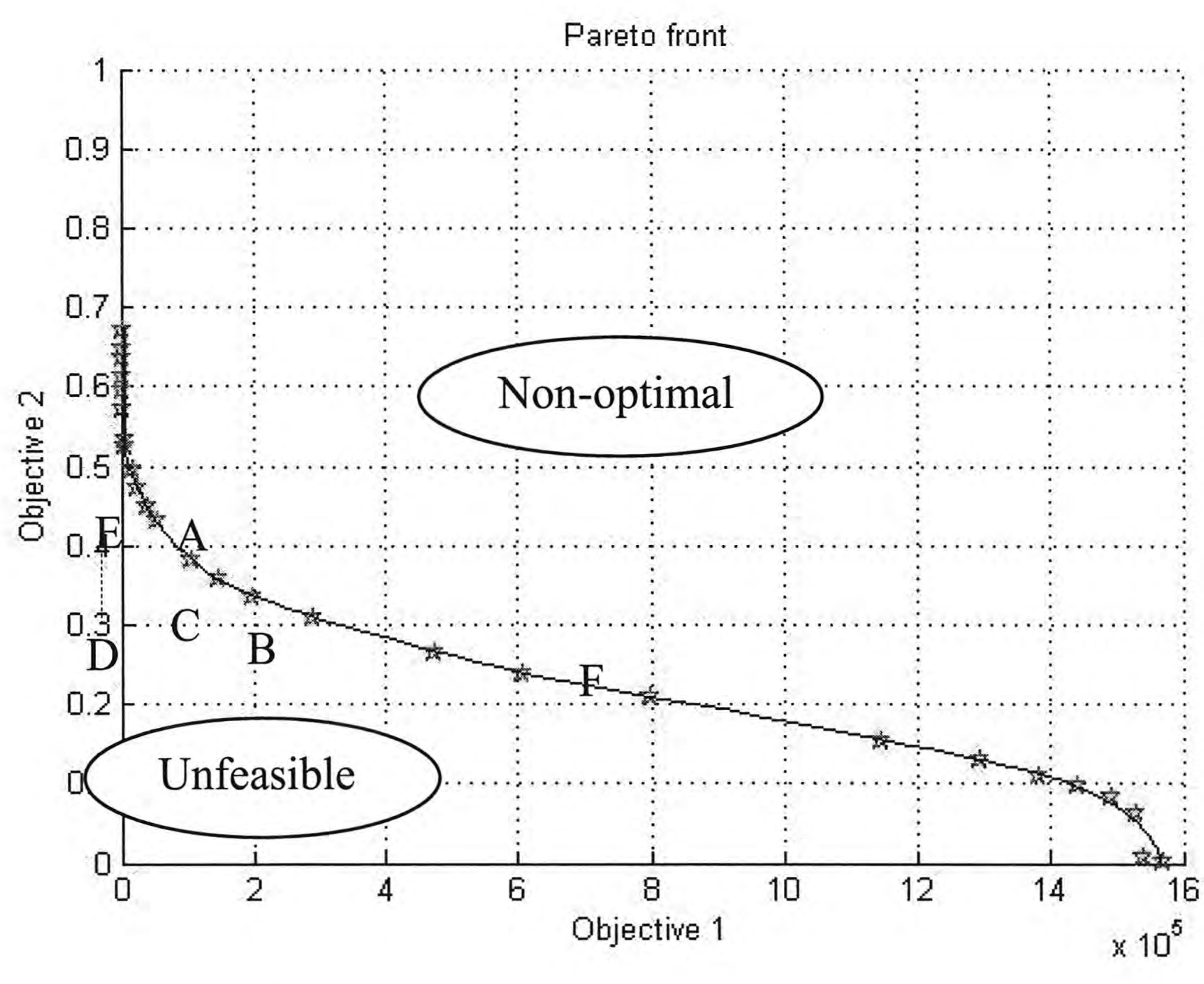


Figure 7.11 Pareto front for excavation (2)

TxDOT should be aware that an owner's risk preference affects the choice of contracting. The resulting multi-objective optimization problem considers a number of trade-offs in optimal solutions. If the decision-making process is governed by a "risk-seeking" policy, the decision-maker could consider optimal solutions in the bottom-right corner of the Pareto front, for example, the optimal solutions with "unexpected loss" greater than $800,000 and a trigger barrier less than 20 percent (the optimal frontier to the right side of point F); for a "risk averse" posture, the owner should consider optimal solutions in the upper-left corner of the Pareto front, for example, the optimal solutions with a trigger barrier greater than 35 percent and "unexpected loss" less than $200,000 (the optimal frontier to the left side of point C); and for a "risk neutral" attitude, the middle part should be considered.

If a decision-maker receives the bids in the "non-optimal" area (area above the Pareto front), then this implies that it would pay an excessive risk premium for the corresponding unexpected risk exposure level, or the model has over-estimated the unexpected risk exposure at the chosen barrier level. For example, consider point A. The corresponding "unexpected loss" is $200,000; the owner should realize that for that amount of exposure, the optimal trigger barrier is not 40 percent (point A) but 35 percent (point C). Alternatively, the agencies should keep the 40 percent barrier level, but pay only $80,000 (point E).

The area below the Pareto front indicates a set of unfeasible solutions given that the cost of risk is determined by the premium pricing model developed in this research. Still, the agency could receive the bids that enter the unfeasible region. What that means is that the contractor has lowered the premium in comparison with the overall construction market. For example, if the owner chooses a trigger barrier of 40 percent but receives the bid that corresponds to point D, this implies that the contractor has discounted the risk premium from $245,300 (point E) to $184,000 (point D) (the calculated risk premium for excavation is $613,248) while the agency still takes on a less unexpected loss of $80,000 (point E).

7.5 Summary

This section illustrates the overall methodology using a real TxDOT project data. The first subsection provides the main description of this project, including the project characteristics and the bidding information. It also suggests TxDOT that the escalation clause should be considered for this project. The second subsection discusses the effect of barrier levels on risk premiums and the future exposures. The third subsection presents the whole process of the optimal risk hedging of commodity prices, both on the control-item level and the project-level. It includes the risk premium estimation, the discussion about the observations and major findings from both multi-objective and single-objective optimization. The fourth subsection shows the implications. In the following section, a summary of the research findings is presented.

8. SUMMARY AND RECOMMENDATIONS FOR FUTURE STUDY

This section summarizes major findings, discusses limitations, and suggests direction for further study. The section is organized in two subsections. In the first subsection, a summary of the report work is presented, while in the second subsection, the limitations of developed methodology are presented, and the directions for future work are suggested.

8.1 Summary

This report investigates risk premiums for commodity prices in highway construction contracts and develops optimal risk hedging model based on agencies' risk preferences. More specifically, the objective of the model is to provide an optimal way to manage commodity-based risk in contracting using an escalation clause with trigger barriers. Such an escalation clause can be used to specify the amount of risk the agency would like to be exposed to during construction via the barrier levels. In other words, it allows balancing between (1) initial payment in the form of risk premium before a contract begins, and (2) future risk exposure during construction. The developed framework also accounts for correlations of commodity risks when balancing the objectives.

The optimal risk mitigation actions are conditional on the owner's risk preferences specified using CVaR-based measures ("unexpected losses"). The solution approach to the problem is based on a multi-objective optimization formulation (or a single-objective degenerate case) and genetic algorithms as a solution approach. The key insights of this study are as follows: (1) trade project cost now for possible cost exposure later by using escalation clause contracts with trigger barriers; (2) use CVaR-based "unexpected loss" to specify an agency's risk preferences; and (3) specify the appropriate level of VaR at the beginning and utilize the Pareto optimal set to determine the acceptable trade-offs.

The overall methodology framework contains three main topics. The first topic is to develop a model to price the unit bid item and the risk. It is essential to determine the price of risk in order to determine optimal risk hedging strategies. Section 4 explains the relationship between the unit bid prices of selected control items and the risk factors - cement price, steel price and oil prices, as well as estimates the impact volatile commodity price have on the unit bid prices (i.e., risk premiums). Weighted least square regression models are conducted to price the risk for unit bid prices.

The second topic is about the contract design using escalation clauses with barrier levels. The relationship between risk premiums and pre-defined barriers is investigated in Section 4. Different prior-defined barrier levels directly influence the future "unexpected losses". The losses are based on simulated commodity prices. Section 5 provides time series models for simulating commodity prices, including both ARIMA model and VARMA model that accounts for the correlation effects.

Finally, the third topic of this report is the multi-objective optimization where agencies' risk preferences (e.g., willingness to take on the risk) are specified using risk measures. Section 6 discusses both the single-objective and multi-objective optimization formulation for developing optimal risk hedging strategies.

8.2 Directions for Future Research

While this report work has presented a comprehensive methodological framework to determine the optimal risk hedging, it is by no means perfect. The presented model would benefit from more data. Firstly, if the data become available, one can develop Pareto fronts applicable to their markets. Secondly, the risk premium estimation analysis with more explanatory variables, such as, the size of contractor, would be more convincing and yield the results that would have higher statistical significance. If the data covering the chosen barrier levels and corresponding bids become available, the barrier levels can be used as an independent variable in the regression models, to see its effect on the unit bid prices.

Some identified problems requiring further research are presented next. Firstly, the relationship between risk premiums and trigger barriers is assumed to be linearly related in this report. More research work is needed to demonstrate or investigate this relationship. Secondly, on a portfolio level, the developed methodology framework can be used to support further analysis of managing cost risks in a multi-project environment. For example, strategies associated with diversifying risk retention and transferring policies for minimizing project portfolio risks can be investigated.

REFERENCES

AASHTO (2008a) *Survey on the Use of Price Adjustment Clauses*, Fall 2008. AASHTO Subcommittee on Construction, Washington, D.C.

AASHTO (2008b) *Guide Specifications for Highway Construction.* American Association of State Highway and Transportation Officials, Washington D.C.

AASHTO (2009) *Survey on the Use of Price Adjustment Clauses*, Fall 2009. AASHTO Subcommittee on Construction, Washington, D.C.

ABARE (2005) Forecasts and issues. *Australiancommodities*, **12**(1), 5-264.

Abraham, A., Jain, L. (2005) Evolutionary multiobjective optimization. In Abraham, A., Jain, L., and Goldberg, R. (Eds.), Evolutionary Multiobjective Optimization: Theoretical Advances and Applications. Springer-Verlag, London, 1-6.

Akintoye, A.S., and MacLeod, M.J. (1997) Risk analysis and management in construction. *International Journal of Project Management*, **15**(1), 31-38.

Al-Harthy, M.H. (2007) Stochastic oil price models: Comparison and impact. *The Engineering Economist*, **52**, 269-284.

Al-Tabtabai, H., and Alex, A.P. (1999) Using genetic algorithms to solve optimization problems in construction. *Engineering, Construction and Architectural Management*, **6**(2), 121-132.

Ardeni, P.G. (1989) Does the law of one price really hold for commodity prices? *American Journal of Agricultural Economics*, **71**(3), 661-669.

Artzner, P., Delbaen, F., Eber, J.M., and Heath D. (1999) Coherent measures of risk. *Mathematical Finance*, **9**(3), 203-228.

Baker, M.P., Mayfield, E.S. and Parsons, J.E. (1998) Alternative models of uncertain commodity prices for use with modern asset pricing. *Energy Journal*, **19**(1), 115-148.

Baloi, D., and Price, A.D.F. (2003) Modelling global risk factors affecting construction cost performance. *International Journal of Project Management*, **21**, 261-269.

Barraza, G. A., Back, W. E., and Mata, F. (2000) Probabilistic monitoring of project performance using SS-curves. *Journal of Construction Engineering and Management*, **126**(2), 142-148.

Bernabe, A., Martina, E., Alvarez-Ramirez, J., and Ibarra-Valdez, C. (2004) A multi-model approach for describing crude oil price dynamics. *Physica A: Statistical Mechanics and its Applications*, **338**, 567-584.

Bessembinder, H., Coughenour, J., Seguin, P., and Smoller, M. (1995) Mean reversion in equilibrium asset price: Evidence from the futures term structure. *Journal of Finance*, **50**(1), 361-375.

Black, F. and Scholes, M.S. (1973) The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**(3), 637–654.

Box, G.E.P., and Cox, D.R. (1964) An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological),* **26**(2), 211-252.

Box, G. E. P., Jenkins, G. M., and Reinsel, G.C. (1994) *Time Series Analysis: Forecasting and Control*, 3rd Ed., Englewood Cliffs, NJ: Prentice-Hall.

Buchananan, W. K., Hodges, P., and Theis, J. (2001) Which way the natural gas price: An attempt to predict the direction of natural gas spot price movements using trader positions, *Energy Economics*, **23**(3), 279–293.

Carr, V., and Tah, J.H.M. (2001) A fuzzy approach to construction project risk assessment and analysis: construction project risk management system. *Advances in Engineering Software*, **32**, 847-857.

Caron, F., Fumagalli, M., and Rigamonti, A. (2007) Engineering and contracting projects: A value at risk based approach to portfolio balancing. *International Journal of Project Management*, **25**, 569-578.

Chinn, M.D., LeBlanc, M., and Coibion, O. (2001) The predictive characteristics of energy futures: Recent evidence for crude oil, natural gas, gasoline and heating oil. UCSC Economics Working Papr No. 490. Available at SSRN: http://ssrn.com/abstract=288844 or doi:10.2139/ssrn.288844.

Chan, W.T., D.K.H. Chua, and G. Kannan. (1996) Construction resource scheduling with genetic algorithms, *Journal of Construction Engineering and Management*, **122**(2), 125-132.

Cheng, G. and Wilmot, C.G. (2008) Louisana highway construction cost trend after hurricane Katrina and Rita. TRB 2008 Annual Meeting, Washington D.C. January 2008.

Christensen, B.J. and Prabhala, N.R. (1998) The relation between implied and realized volatility. *Journal of Financial Economics*, **50**, 125-150.

Clements, M.P. and Hendry, D.F. (2008) *A Companion to Economic Forecasting*. Malden, MA: Wiley-Blackwell.

Coello Coello, C. A. (2005). Recent trends in evolutionary multiobjective optimization. In Abraham, A., Jain, L., and Goldberg, R. (Eds.), Evolutionary Multiobjective Optimization: Theoretical Advances and Applications. Springer-Verlag, London, 7-32.

Coello Coello, C.A. (2006) Evolutionary multiobjective optimization: A historical view of the field. *IEEE Computational Intelligence Magazine*, **1**(1), 28-36.

Contreras, J., Espinola, R., Nogales, F.J., and Conejo, A.J. (2003) ARIMA models to predict next-day electricity prices. *IEEE Transactions on Power Systems*, **18**(3), 1014-1020.

Corrado, C. J. and Miller, T. W. (1996) A note on a simple, accurate formula to compute implied standard deviations. *Journal of Banking and Finance*, **20**, 595–603.

Damnjanovic, I., and Zhou, X. (2009) Impact of crude oil market behavior on unit bid prices: The evidence from the highway construction sector. *Construction Management and Economics*, **27**, 881-890.

Dawood, N., and Sriprasert, E. (2006) Construction scheduling using multi-constraint and genetic algorithms approach. *Construction Management and Economics*, **24**, 19-30.

Deb, K. (1999a) An introduction to genetic algorithms. *Sadhana*, **24**, 293–315.

Deb, K. (1999b) Multi-objective genetic algorithms: Problem difficulties and construction of test problems. *Evolutionary Computation*, **7**(3), 205-230.

Deb, K. (2001) *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons. Chichester, U.K.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, **6**(2), 182-197.

Deb, K. and Gupta, H. (2005) Searching for robust pareto-optimal solutions in multiobjective optimization. *Proceedings of the third International Conference on Evolutionary Multi-Criterion Optimization (EMO-05)*, 150–164.

Deshpande, V.P., Damnjanovic, I.D., and Gardoni, P. (2010) Reliability-based optimization models for scheduling pavement rehabilitation. *Computer-Aided Civil and Infrastructure Engineering*, **25**, 227-237.

Dikmen, I., Birgonul, M.T., and Han, S. (2007) Using fuzzy risk assessment to rate cost overrun risk in international construction projects. *International Journal of Project Management*, **25**, 494-505.

Emsley, M. W., Lowe, D. J., Duff, A. R., Harding, A. and Hickson, A. (2002) Data modeling and the application of a neural network approach to the prediction of total construction cost. *Construction Management and Economics*, **20**, 465-472.

Engle, R.F., and Granger, C.W.J. (1987) Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, **55**(2), 251-276.

Feng, C.W., Liu, L., and Burns, S.A. (1997) Using genetic algorithms to solve construction time-cost trade-off problems. *Journal of Computing in Civil Engineering*, **11**(3), 184-189.

FHWA (1980) *Development and Use of Price Adjustment Contract Provisions*. Technical Advisory T5080.3. U.S. Department of Transportation.

FHWA (2004) *Guidelines on Preparing Engineer's Estimate, Bid Reviews and Evaluation*. January 20, 2004. http://www.fhwa.dot.gov/programadmin/contracts/ta508046.cfm

FHWA (2006) *Price trends for federal-aid highway construction*. U.S. Dept. of Transportation, Rep. No. FHWA-IF-06-048, Washington, D.C.

FHWA (2007) *Growth in Highway Construction and Maintenance Costs*. U.S. Department of Transportation. Report number: CR-2007-079. Washington, D.C.

Figlewski, S. (1997) Forecasting volatility. *Financial Markets, Institutions & Instruments*. **6**(1), 1-88.

Gallagher, J., and Riggs, F. (2006) *Material price escalation: Allocating the risks*. Construction Briefings, Thomson/West, Eagan.

Ghosh, A., Saidi, R., and Johnson, K.H. (1999). Who moves the Asia-Pacific stock markets-US or Japan? Empirical evidence based on the theory of cointegration. *The Financial Review*, **34**, 159-170.

GNB. (2007) Cross cutting management issues. Measures, Markers and Mileposts, The Gray Notebook for the quarter ending September 30, Washington State Department of Transportation. www.wsdot.wa.gov/biz/construction/GNB/3q2007graybook.pdf

Guzel, A. (2003) Optimal Commodity Procurement. Department of Information Systems, Statistics, and Management Science, University of Alabama, Alabama, Report in 2003.

Hartgen, D. T. and Talvitie, A. P. (1995) Comparative performance of transportation systems: Pitfalls and prospects. Proc., World Conf. on Transit Research, Sydney, Australia.

Hartgen, D. T., Bowman, V. B., and Horner, M. W. (1997) Baseline market forecasts for state highway construction engineering, maintenance and administration programs 1997-2001. Transportation Publication Number 151, Center for Interdisciplinary Transportation Studies, University of North Carolina, Charlotte, N. C.

Hastak, M., and Shaked, A. (2000) ICRAM-1: Model for international construction risk assessment. *Journal of Management in Engineering*, **16**(1), 59-69.

Jaafari, A. (2001) Management of risks, uncertainties and opportunities on projects: time for a fundamental shift. *International Journal of Project Management*, **19**, 89-101.

Kamarianakis, Y. and Prastakos, P. (2003) Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches, Proceedings of the Transportation Research Board 82nd Annual Meeting, Washington, DC.

Kangari, R. (1995) Risk management perceptions and trends of U.S. construction. *Journal of Construction Engineering and Management*, **121**(4), 422-429.

Kangari, R., and Riggs, L.S. (1989) Construction risk assessment by linguistics. *IEEE Transactions on Engineering Management*, **36**(2), 126-131.

Kartam, N.A., and Kartam, S.A. (2001) Risk and its management in the Kuwaiti construction industry: A contractors' perspective. *International Journal of Project Management*, **19**, 325-335.

Kawaller, I.G., Koch, P.D., and Peterson, J.E. (1994) Assessing the intraday relationship between implied and historical volatility. *The Journal of Futures Markets*, **14**(3), 323-346.

Kim, G.H., An, S.H., and Kang, K.I. (2004) Comparison of construction cost estimating models based on regression analysis, neural networks and case-based reasoning. *Building and Environment,* **39**, 1235–1242.

Kinney, W.R. (1978) ARIMA and regression in analytical review: An empirical test. *The Accounting Review*, **53**(1), 48-60.

Konak, A., Coit, D.W., and Smith, A.E. (2006) Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering and System Safety*, **91**, 992-1007.

Koppula, S. D. (1981) Forecasting engineering costs: Two case studies. *Journal of the Construcion Division*, ASCE, **107**(4), 733-743.

Kotze, A.A. (2005) Stock price volatility: a primer. *Financial Chaos Theory*, January 2005, http://www.quantonline.co.za.

Leu, S.S., and Yang, C.H. (1999) A genetic-algorithm-based resource-constrained construction scheduling system. *Construction Management and Economics*, **17**, 767-776.

Leu, S.S., and Yang, C.H., and Huang, J.C. (2000) Resource leveling in construction by genetic algorithm-based optimization and its decision support system application. *Automation in Construction*, **10**, 27-41.

Li, S. (2005) A new formula for computing implied volatility. *Applied mathematics and computation*, **170**, 611-625.

Li, H., and Love, P.E.D. (1997) Improved genetic algorithms for time-cost optimization, *Journal of Construction Engineering and Management*, **123**(3), 233-287.

Lowe, D. J., Emsley, M. W. and Harding, A. (2006) Predicting construction cost using multiple regression techniques. *Journal of Construction Engineering and Management*, **132**(7), 750-758.

Lutkepohl, H., and Claessen, H. (1997) Analysis of cointegrated VARMA processes. *Journal of Econometrics*, **80**, 223-239.

Mansini, R., Ogryczak, W., and Speranza, M.G. (2007) Conditional Value at Risk and related linear programming models for portfolio optimization. *Annals of Operations Research*, **152**(1), 227-256.

MathWorks (2011) Global optimization toolbox. http://www.mathworks.com/products /global-optimization/index.html.

McFall, K. (2005). Crude oil prices hit new heights. *ENR Cost Report. 3Q*, 71.

McGoldrick, T. (2006) Material price increases: What can you do? *The Chemical Engineer*, No. 786, 18-19.

Mendell, B.C. (2006) Benchmarking construction cost changes in the U.S. south. American Reconstruction, September 2006, http://www.forisk.com/UserFiles/File/Construction%20Index%20final.pdf.

Merton, R.C. (1973) The theory of rational option pricing. *Journal of Economics and Management Science*, **4**, 141–183.

Montgomery, A.L., Zarnowitz, V. Tsay, R.S., and Tiao, G.C. (1998) Forecasting the U.S. unemployment rate, *Journal of the American Statistical Association*, **93**(442), 478-493.

Morana, C. (2001) A semiparametric approach to short-term oil price forecasting, *Energy Economics*, **23**(3), 325–338.

Myers, R.J. (1994) Time series econometrics and commodity price analysis: A review. *Review of Marketing and Agricultural Economics*, **62**, 167-181.

Nash, P.T., Jayawickrama, P., Senadherra, S., Borrelli, J. and Ashek Rana, A.S.M. (1995) Guidelines for Using Hydrated Fly Ash as a Flexible Base, Research Report 0-1365-1F, College of Engineering, Texas Tech University's, Lubbock.

NIST (2010) NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, June 23, 2010.

Nochai, R. and Nochai, T. (2006) ARIMA model for forecasting oil palm price. Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications, Universiti Sains Malaysia, Penang, June 13-15, 2006.

Olsson, R. (2008) Risk management in a multi-project environment: an approach to manage portfolio risks. *International Journal of Quality & Reliability Management*, **25**(1), 60-71.

Paek, J.H., Lee, Y.W., and Ock, J.H. (1993) Pricing construction risk: fuzzy set application. *Journal of Construction Engineering and Management*, **119**(4), 743-756.

Pandit, D., Anderson, S., Damnjanovic, I. and Nejat, A. (2009) Assessment of strategies and methods to reduce construction costs while maintaining Quality. Proceedings of 1st International Conference on Transportation Construction Management, Orlando, FL.

Pelet, M. (2003) Real options in petroleum: Geometric Brownian motion and Mean-Reversion with jumps. MS. report, Oxford University.

Postali, F.A.S. and Picchetti, P. (2006) Geometric Brownian motion and structural breaks in oil prices: A quantitative analysis. *Energy Economics*, **28**, 506-522.

Quenouille, M.H. (1957). *The Analysis of Multiple Time Series*. London: Griffin.

Rahman, M.M. and Kumaraswamy, M.M. (2002) Joint risk management through transactionally efficient relational contracting. *Construction Management and Economics*, **20**, 45-54.

Reed Construction Data. (2008) RS Means Building Construction Cost Data. 66th Annual Edition RS Means. Kingston, MA.

Roman, D., Darby-Dowman, K., and Mitra, G. (2007) Mean-risk models using two risk measures: A multi-objective approach. *Quantitative Finance*, **7**(4), 443-458.

Sarykalin, S., Serraino, G., and Uryasev, S. (2008) Value-at-Risk vs. Conditional Value-at-Risk in risk management and optimization, Hanover, MD, USA. Tutorials in Operations Research, INFORMS, 270-294.

SAS Institute Inc. (2010) *SAS/STAT® 9.22 User's Guide*. Cary, NC: SAS Institute Inc.

Senouci, A.B., and Eldin, N.N. (2004) Use of genetic algorithms in resource scheduling of construction projects. *Journal of Construction Engineering and Management*, **130**(6), 869-877.

Sheather, S. J. (2008) *A Modern Approach to Regression with R*. Springer, New York.

Srinivas N. and Deb, K. (1994) Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, **2**(3), 221-248.

Stevens, J.D. (1995) Cost estimating and forecasting for highway work in Kentucky. Research Rep. KTC 95-12, Kentucky Transportation Center, University of Kentucky, Lexington, Ky.

Stoll, H. R. (1969) The relationship between put and call option prices. *The Journal of Finance*, **24**(5), 801-824.

Tah, J.H.M., and Carr, V. (2000) Information modeling for a construction project risk management system. *Engineering, Construction and Architectural Management*, **7**(2), 107-119.

Tam, C.M., Tong, T.K.L., Cheung, S.O., and Chan, A.P.C. (2001) Genetic algorithm model in optimizing the use of labour. *Construction Management and Economics*, **19**, 207-215.

Tamaki, H., Kita, H., and Kobayashi, S. (1996) Multi-objective optimization by genetic algorithms: A review. *Proceedings of the 1996 IEEE International Conference on Evolutionary Computation*, Nagoya, Japan, 517–522.

Tiao, G.C. and Box, G.E.P. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association*, **76**, 802– 816.

Tong, T.K.L., C.M. Tam, and A.P.C. Chan. (2001) Genetic algorithm optimization in building portfolio management. *Construction Management and Economics*, **19**, 601-609.

Trigeorgis, L. (2002) *Real Options: Managerial Flexibility and Strategy in Resource Allocation.* The MIT Press, Cambridge, Massachusetts, London, England.

Trost, S. M. and Oberlender, G. D. (2003) Predicting accuracy of early cost estimates using factor analysis and multivariate regression. *Journal of Construction Engineering and Management*, **129**(2), 198-204.

Tuan, L.T. (2010) Phosphate fertilizers' domestic price movement in Vietnam, *International Journal of Business and Management*, **5**(5), 74-81.

Veldhuizen, D. A. V., & Lamont, G. B. (2000) Multiobjective evolutionary algorithms: Analyzing the state-of-the-art. *Evolutionary Computation*, **8**(2), 125-147.

Weiss, E. (2000) Forecasting commodity prices using ARIMA. *Technical Analysis of Stocks and Commodities*, **18**(1), 18-19.

Wilmot, C. G., and Cheng, G. (2003) Estimating future highway construction costs. *Journal of Construction Engineering and Management*, **129**(3), 272-279.

Wilmot, C. G. and Mei, B. (2005) Neural network modeling of highway construction costs. *Journal of Construction Engineering and Management*, **131**(7), 765-771.

WSDOT (2011) Trends in highway material costs. http://www.wsdot.wa.gov/biz/construction/constructioncosts2.html

Xie, W., Yu, L., Xu, S., and Wang, S. (2006) A new method for crude oil price forecasting based on support vector machines. *Lecture Notes in Computer Science*, **3994**, 444-451.

Zhang, G.P. (2003) Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, **50**, 159-175.

Zheng, D.X.M., Ng, S.T., and Kumaraswamy, M.M. (2004) Applying a genetic algorithm-based multiobjective approach for time-cost optimization. *Journal of Construction Engineering and Management*, ASCE, **130**(2), 168-176.

Zhi, H. (1995) Risk management for overseas construction projects. *International Journal of Project Management*, **13**(4), 231-237.

Zitzler, E. (1999) Evolutionary algorithms for multiobjective optimization: Methods and applications. Doctoral report ETH 13398, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland.

APPENDIX A

TXDOT BIDDING DATA

Table A-1 HCI Item Definition

Category	Element	Control Item	Unit
Earthwork	Excavation	Roadway Excavation	CY
	Embankment	Roadway Embankment	CY
Subgrade and base course	Lime treated subgrade or base	Lime	Ton
		Lime Treatment	CY
		Plant Mix	CY
	Cement treated subgrade or base	Cement	Ton
		Cement Treatment	CY
		Cement Trt Plant Mix	CY
	Asphalt treated base or foundation course	Asph. Trt. Plant Mix	Ton
	Flexible base	Flexible Base	CY
Surfacing	Surface treatment	Surface Treatment Asphalt	Gal
		Surface Treatment Aggregate	CY
	Bituminous mixtures	Hot Mix Asphaltic Concrete	Ton
	Concrete pavement	Continuous Reinforced Concrete Pavement	CY
		Jointed Reinforced concrete Pavement	CY
		Jointed Non-reinforced Concrete Pavement	CY
Structures	Structural concrete	Class A Concrete	CY
		Class C Concrete	CY
		Class S Concrete	CY
		Bridge Rail (Rigid)	LF
		Bridge Slab	SF
	Metal for structures	Metal for Structures	LB
	Precast prestressed conc structural members	Regular Beams	LF
		Box Beams	LF
	Foundations	Concrete Piling	LF
		Steel H Piling	LF
		Drilled Shafts	LF
	Drainage	Reinforced Concrete Pipe	LF
		Corrugated Metal Pipe	LF
		Reinforced Concrete Pipe(Sewer)	LF
		Concrete Box Culvert	LF
		Concrete Box Sewer	LF
	Riprap	Concrete Riprap	CY
	Retaining walls	Retaining Walls	SF

APPENDIX B

VOLATILITY OF COMMODITY PRICES

This appendix contains the computed implied volatility of crude oil prices, and historical volatilities of cement and steel prices. The formulas in the literature for calculating implied volatility of crude oil prices are listed, followed by the result and discussion about how the best formula is chosen.

Table B-1 Formulas for computing implied volatility in the literature

Formula	Equation	assumption	conclusion
1	$\sigma \approx \sqrt{\frac{2\pi}{T}}\frac{C}{S}$	$S = Xe^{-rt}$	The data does not meet this assumption.
2	$\sigma \approx \frac{2\sqrt{2}}{\sqrt{T}}z - \frac{1}{\sqrt{T}}\sqrt{8z^2 - \frac{6\alpha}{\sqrt{2}z}}$ $\alpha = \frac{\sqrt{2\pi}C}{S}$ $z = \cos[\frac{1}{3}\cos^{-1}(\frac{3\alpha}{\sqrt{32}})]$	$S = Xe^{-rt}$	This is more accurate than formula 1.
3	$\sigma \approx \frac{2\sqrt{2}}{\sqrt{T}}z - \frac{1}{\sqrt{T}}\sqrt{8z^2 - \frac{6\alpha}{\sqrt{2}z}}$ $z = \cos[\frac{1}{3}\cos^{-1}(\frac{3\alpha}{\sqrt{32}})]$ $\alpha = \frac{\sqrt{2\pi}}{1+\eta}[\frac{2C}{S} + \eta - 1]$	Nearly $S = Xe^{-rt}$ $\sigma \gg \sqrt{\frac{\lvert\eta - 1\rvert}{T}}$ $\eta = \frac{K}{S}$	In our data, σ is not far bigger than $\sqrt{\frac{\lvert\eta-1\rvert}{T}}$
4	$\sigma \approx \sqrt{\frac{2\pi}{T}}\frac{1}{S+K}\left[C - \frac{S-K}{2} + \sqrt{\left(C - \frac{S-K}{2}\right)^2 - \frac{(S-K)^2}{\pi}}\right]$	$S \neq Xe^{-rt}$	This formula fits the data.
5	$\sigma \approx \sqrt{\frac{2\pi}{T}}\frac{C-(S-K)/2}{S-(S-K)/2}$	Far Less accurate than formula 4.	
6	$\sigma \approx \frac{\alpha + \sqrt{\alpha^2 - \frac{4(\eta-1)^2}{1+\eta}}}{2\sqrt{T}}$ $\quad \alpha = \frac{\sqrt{2\pi}}{1+\eta}\left[\frac{2C}{S} + \eta - 1\right]$ $\eta = \frac{K}{S}$	For deep-in or out-of-money options.	

Implied volatilities have been calculated from all the 6 formulas in Table B-1 for 3 months' call option. Formula 5 is not used because it is far less accurate than formula 4. It turns out that formula 4 yields implied volatilities approximation that are nearly identical to the ones directly calculated from Black-Scholes model. The comparison is based on estimation error (the difference between the real volatility and estimated volatility). A more detailed calculation shows that the estimation error by formula 4 on average is 0.00668 less than that by formula 6. Note that average true volatility is about 39.1%. Thus formula 4 is on average about 1.71% more accurate than formula 6. Thus formula 4 is chosen as the best formula for calculating implied volatility of crude oil prices.

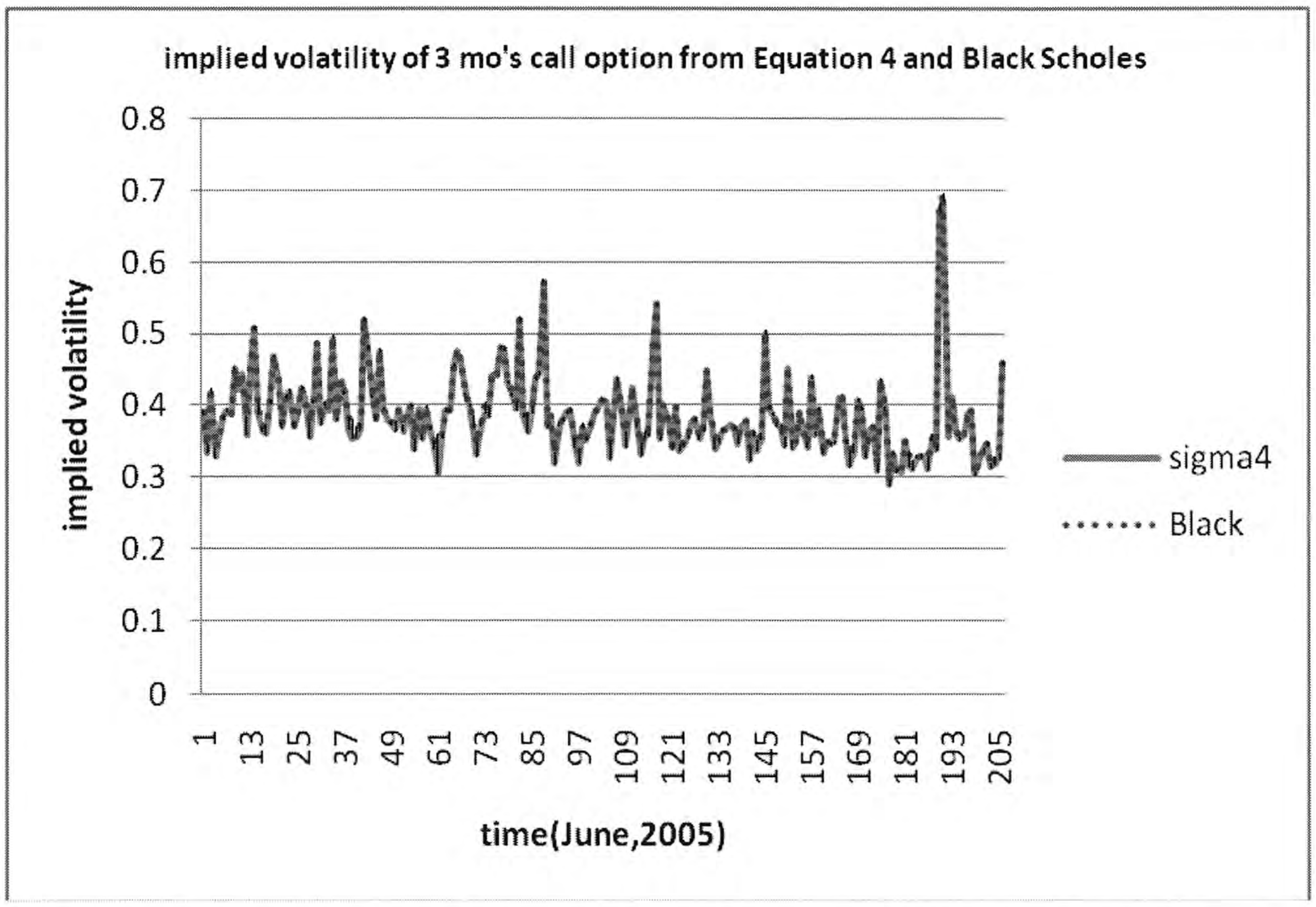


Figure B-1 Implied volatility of 3 mo's call option from Equation 4 and Black Scholes

Table B-2 Implied volatility of 3 month call option in 1998-2006

	1998	1999	2000	2001	2002	2003	2004	2005	2006
Jan	0. 4421	0. 6483	0. 2983	0. 4300	0. 4550	0. 3560	0. 3600	0. 4039	0. 1483
Feb	0. 5631	0. 6527	0. 2297	0. 3545	0. 4492	0. 4299	0. 3574	0. 4079	0. 3256
Mar	0. 6214	0. 7500	0. 2951	0. 4470	0. 4394	0. 4824	0. 3555	0. 3942	0. 2935
Apr	0. 5876	0. 5538	0. 3888	0. 4102	0. 3904	0. 4280	0. 3982	0. 3893	0. 2612
May	0. 4268	0. 5390	0. 2698	0. 2771	0. 3778	0. 3614	0. 3872	0. 3672	0. 2802
Jun	0. 7257	0. 5420	0. 2714	0. 2293	0. 3931	0. 3536	0. 4430	0. 3790	0. 2854
Jul	0. 5818	0. 3793	0. 3042	0. 3672	0. 3288	0. 3321	0. 3864	0. 3540	0. 2817
Aug	0. 5638	0. 3417	0. 3062	0. 2952	0. 3322	0. 3248	0. 3454	0. 3301	0. 2507
Sep	0. 5751	0. 3741	0. 3059	0. 4122	0. 3820	0. 3317	0. 3969	0. 4158	0. 3149
Oct	0. 5615	0. 3442	0. 4117	0. 5537	0. 3192	0. 2889	0. 4356	0. 3687	0. 3148
Nov	0. 6775	0. 2852	0. 2893	0. 5443	0. 4052	0. 3751	0. 5613	0. 3794	0. 2686
Dec	0. 7510	0. 2636	0. 4507	0. 5812	0. 3263	0. 3630	0. 5018	0. 3525	0. 2234

Table B-3 Historical volatility of cement prices in 1998-2006

	1998	1999	2000	2001	2002	2003	2004	2005	2006
Jan	0. 0105	0. 0126	0. 0139	0. 0134	0. 0110	0. 0108	0. 0072	0. 0117	0. 0159
Feb	0. 0106	0. 0127	0. 0139	0. 0134	0. 0110	0. 0108	0. 0072	0. 0117	0. 0156
Mar	0. 0096	0. 0127	0. 0139	0. 0134	0. 0110	0. 0108	0. 0073	0. 0120	0. 0162
Apr	0. 0102	0. 0123	0. 0146	0. 0119	0. 0108	0. 0112	0. 0101	0. 0118	0. 0180
May	0. 0102	0. 0173	0. 0085	0. 0107	0. 0111	0. 0109	0. 0105	0. 0098	0. 0192
Jun	0. 0134	0. 0145	0. 0085	0. 0107	0. 0126	0. 0108	0. 0105	0. 0119	0. 0165
Jul	0. 0125	0. 0144	0. 0085	0. 0108	0. 0125	0. 0107	0. 0105	0. 0118	0. 0214
Aug	0. 0116	0. 0144	0. 0086	0. 0153	0. 0078	0. 0107	0. 0112	0. 0130	0. 0224
Sep	0. 0110	0. 0147	0. 0081	0. 0153	0. 0078	0. 0108	0. 0111	0. 0125	0. 0224
Oct	0. 0114	0. 0145	0. 0082	0. 0154	0. 0101	0. 0078	0. 0117	0. 0158	0. 0229
Nov	0. 0125	0. 0138	0. 0134	0. 0108	0. 0109	0. 0072	0. 0115	0. 0158	0. 0231
Dec	0. 0125	0. 0140	0. 0134	0. 0109	0. 0109	0. 0072	0. 0121	0. 0158	0. 0229

Table B-4 Historical volatility of steel prices in 1998-2006

	1998	1999	2000	2001	2002	2003	2004	2005	2006
Jan	0. 0387	0. 0337	0. 0369	0. 0262	0. 0172	0. 0343	0. 0306	0. 1085	0. 0625
Feb	0. 0394	0. 0337	0. 0407	0. 0171	0. 0170	0. 0343	0. 0517	0. 1096	0. 0623
Mar	0. 0361	0. 0385	0. 0328	0. 0171	0. 0213	0. 0366	0. 0776	0. 0993	0. 0642
Apr	0. 0226	0. 0371	0. 0329	0. 0229	0. 0328	0. 0279	0. 0903	0. 0930	0. 0571
May	0. 0216	0. 0345	0. 0314	0. 0234	0. 0322	0. 0300	0. 0944	0. 0883	0. 0577
Jun	0. 0223	0. 0343	0. 0306	0. 0195	0. 0325	0. 0294	0. 0924	0. 0882	0. 0575
Jul	0. 0215	0. 0336	0. 0314	0. 0195	0. 0326	0. 0295	0. 0922	0. 0908	0. 0539
Aug	0. 0205	0. 0332	0. 0314	0. 0188	0. 0330	0. 0293	0. 0903	0. 0929	0. 0477
Sep	0. 0227	0. 0282	0. 0314	0. 0180	0. 0329	0. 0296	0. 0889	0. 0904	0. 0483
Oct	0. 0327	0. 0375	0. 0246	0. 0175	0. 0348	0. 0275	0. 0892	0. 0932	0. 0458
Nov	0. 0341	0. 0371	0. 0260	0. 0169	0. 0346	0. 0286	0. 1063	0. 0800	0. 0395
Dec	0. 0337	0. 0369	0. 0262	0. 0171	0. 0343	0. 0306	0. 1095	0. 0633	0. 0393

APPENDIX C

MULTIPLE LINEAR REGRESSION PROCESS

Step 1-1: Scatter plot for response variable and explanatory variables

Figure C-1 shows a scatter plot matrix of the response variable and four predictor variables while the other independent variables are dummy variables. The response variable and four predictor variables are each highly skewed. In addition, the predictors do not appear to be linearly related. Thus, we need to consider transformations of the response and the four predictor variables.

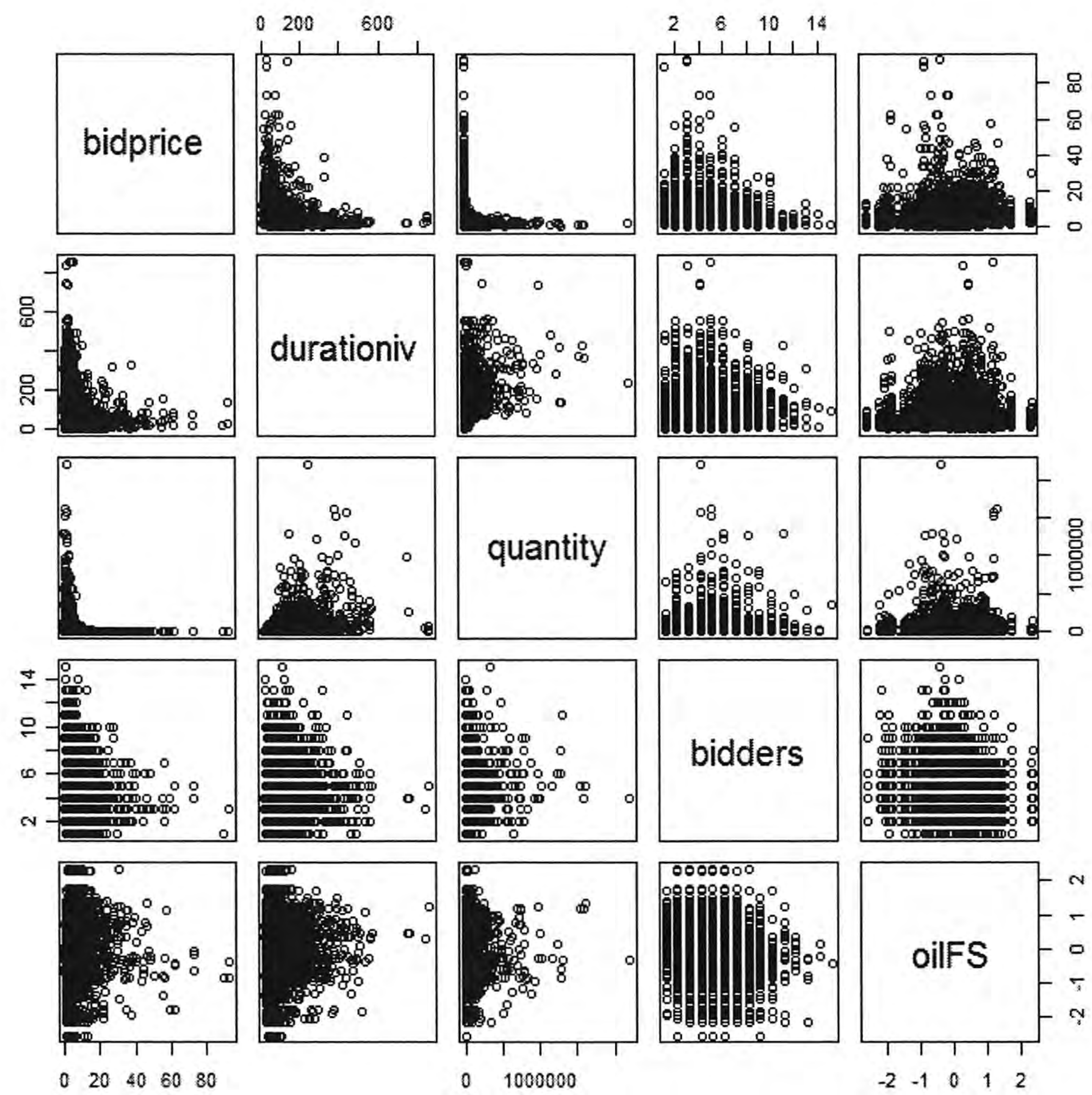
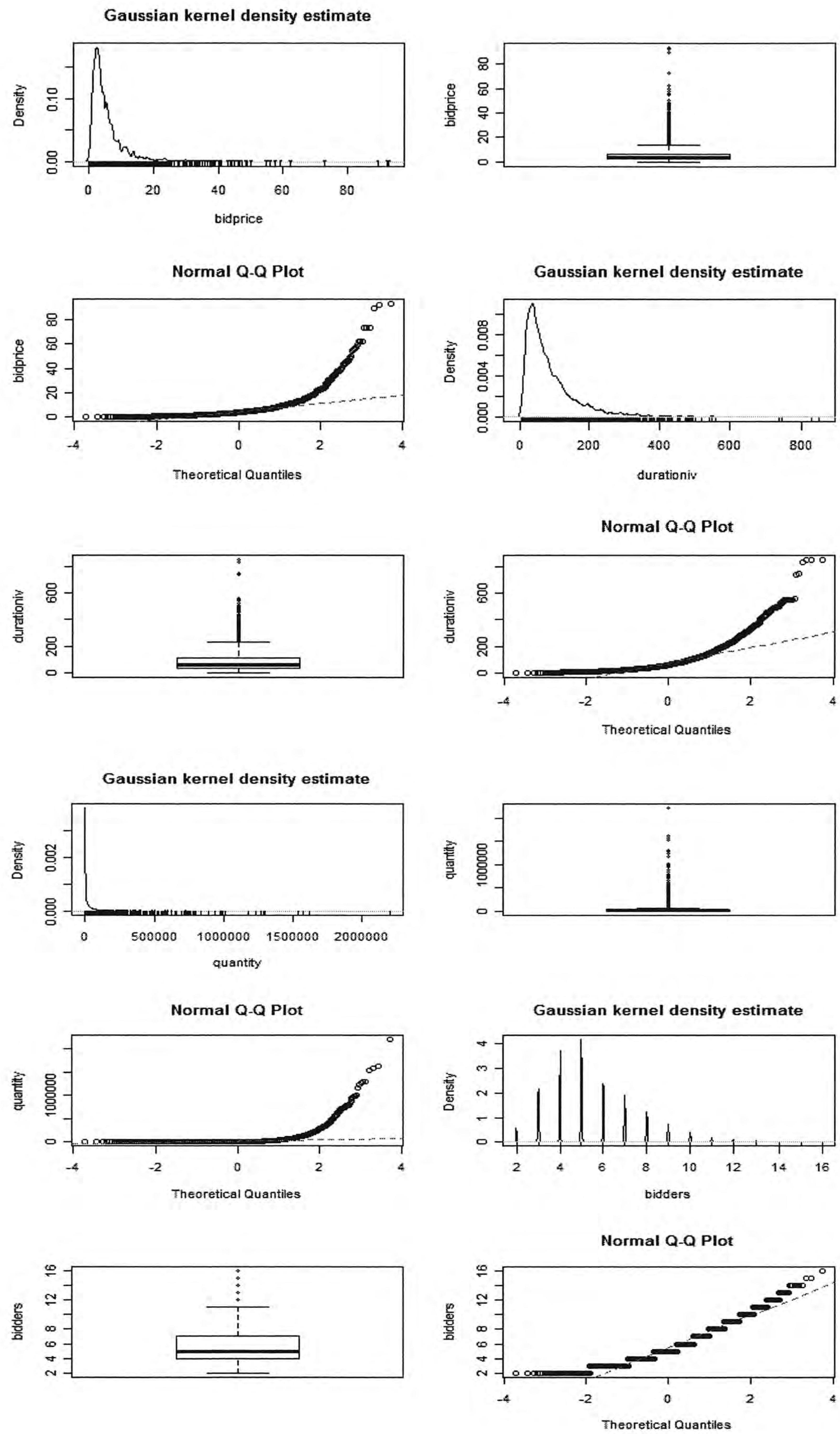


Figure C-1 A scatter plot matrix of untransformed data

Step 1-2: Box plots, normal QQ plots

Figure C-2 contains box plots, normal QQ plots and Gaussian kernel density estimates for the untransformed data. It is evident from Figure C-2 that the distributions of unit bid price, duration-based implied volatility, quantity and number of bidders are skewed. On the other hand the distribution of the difference of futures and spot prices is consistent with a normal distribution.


Gaussian kernel density estimate
Density
0.00
0.10
0
20
40
60
80
bidprice
bidprice
0
20
40
60
80
Normal Q-Q Plot
bidprice
0
20
40
60
80
-4
-2
0
2
4
Theoretical Quantiles
Gaussian kernel density estimate
Density
0.000
0.004
0.008
0
200
400
600
800
durationiv
durationiv
0
200
600
Normal Q-Q Plot
durationiv
0
200
600
-4
-2
0
2
4
Theoretical Quantiles
Gaussian kernel density estimate
Density
0.000
0.002
0
500000
1000000
1500000
2000000
quantity
quantity
0
1000000
Normal Q-Q Plot
quantity
0
1000000
-4
-2
0
2
4
Theoretical Quantiles
Gaussian kernel density estimate
Density
0
1
2
3
4
2
4
6
8
10
12
14
16
bidders
bidders
2 4 6 8
12
16
Normal Q-Q Plot
bidders
2 4 6 8
12
16
-4
-2
0
2
4
Theoretical Quantiles

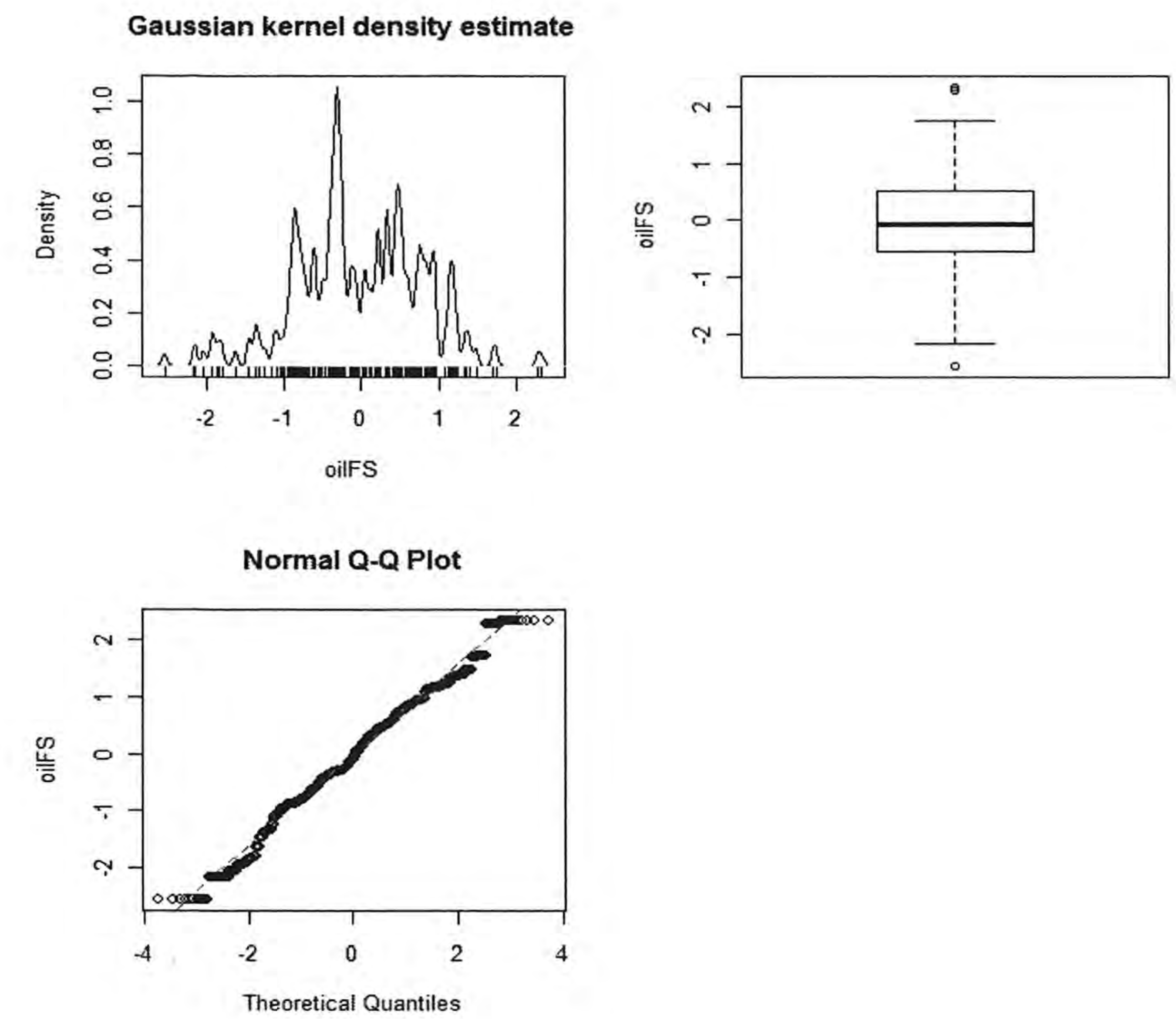


Figure C-2 Box plots, normal QQ-plots and kernel density estimates of untransformed data

Step 2: Data transformation using Box-Cox

Box-Cox method is considered to overcome problems due to nonlinearity. Given below is the ouput from R using Approach 1 which is the same as mentioned in model development I:

Output from R using Approach 1

```
             Est.Power Std.Err. Wald(Power=0) Wald(Power=1)
durationiv     0.0149   0.0135      1.0971        -72.7692
quantity       0.0270   0.0051      5.3326       -192.0118
bidders       -0.0299   0.0387     -0.7740        -26.6420
oilFS          1.0924   0.0213     51.3745          4.3442
                                    LRT  df  p.value
LR test, all lambda equal 0   2615.186  4    0
LR test, all lambda equal 1  34050.849  4    0
```

Using the Box-Cox method to transform the predictor variables toward normality results in taking natural logarithms of duration-based volatility, quantity and number of bidders while the difference between futures and spot prices of crude oil untransformed. Figure C-3 contains an inverse response plot which provides the closest fit of power of unit bid price is 0.06. The estimated optimal value 0.06 can be rounded to 0 which corresponds to natural logrithm transformation of unit bid price.

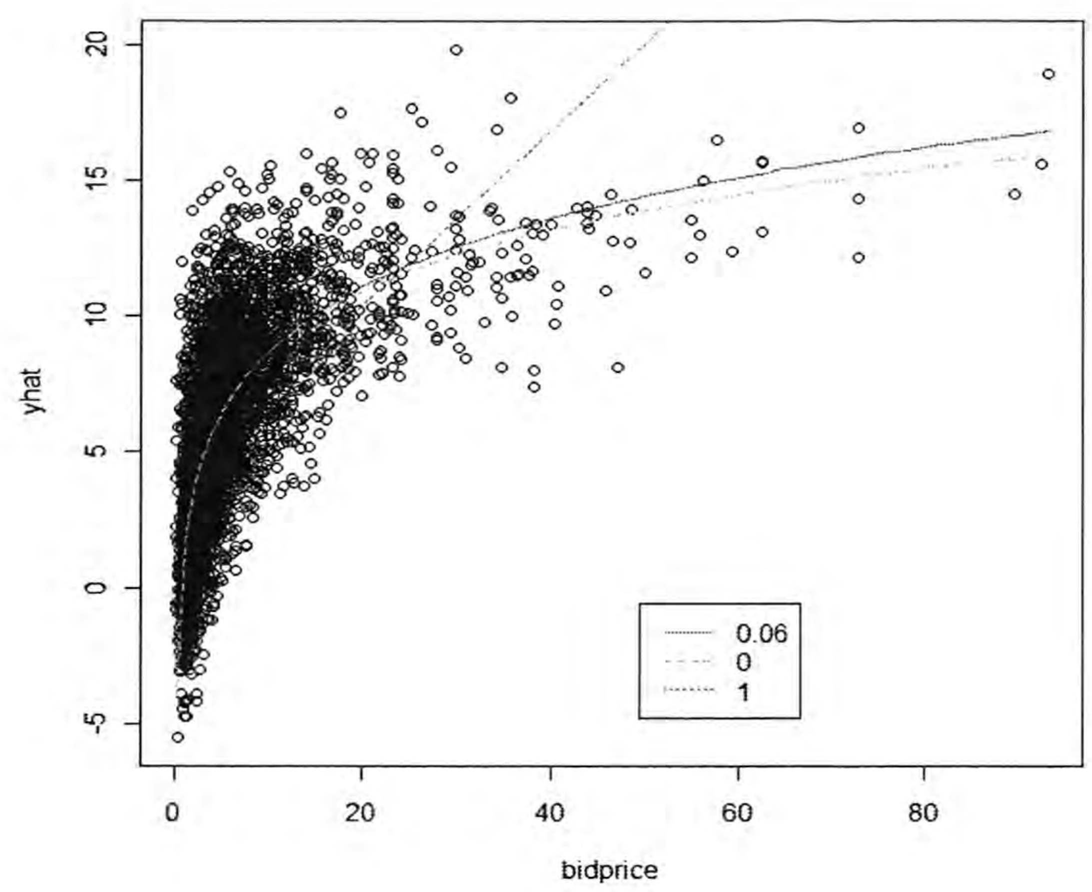


Figure C-3 Inverse response plot

Given below is the ouput from R using Approach 2 which is the same as mentioned in model development I:

Output from R using Approach 2

	Est.Power	Std.Err.	Wald(Power=0)	Wald(Power=1)
bidprice	-0.3436	0.0166	-20.7121	-80.9886
durationiv	0.0132	0.0136	0.9708	-72.6233
quantity	0.0108	0.0046	2.3432	-214.8733
bidders	-0.0385	0.0386	-0.9971	-26.8801
oilFS	1.0953	0.0212	51.6852	4.4951

	LRT	df	p.value
LR test, all lambda equal 0	3078.454	5	0
LR test, all lambda equal 1	44038.086	5	0

The estimated power for bid price is -0.34 which could also be rounded to 0. Thus, the two approaches agree in that they suggest that each variable be transformed using the log transformation except the difference between futures and spot prices of crude oil.

Step 3: Re-check step 1

Figure C-4 shows a scatter plot matrix of the log transformed response and some predictor variables (except the expected change of crude oil prices and dummy variables). The pairwise relationships in Figure C-4 are much more linear than those in Figure C-1. Their joint distributions are roughly ellipsoidal.

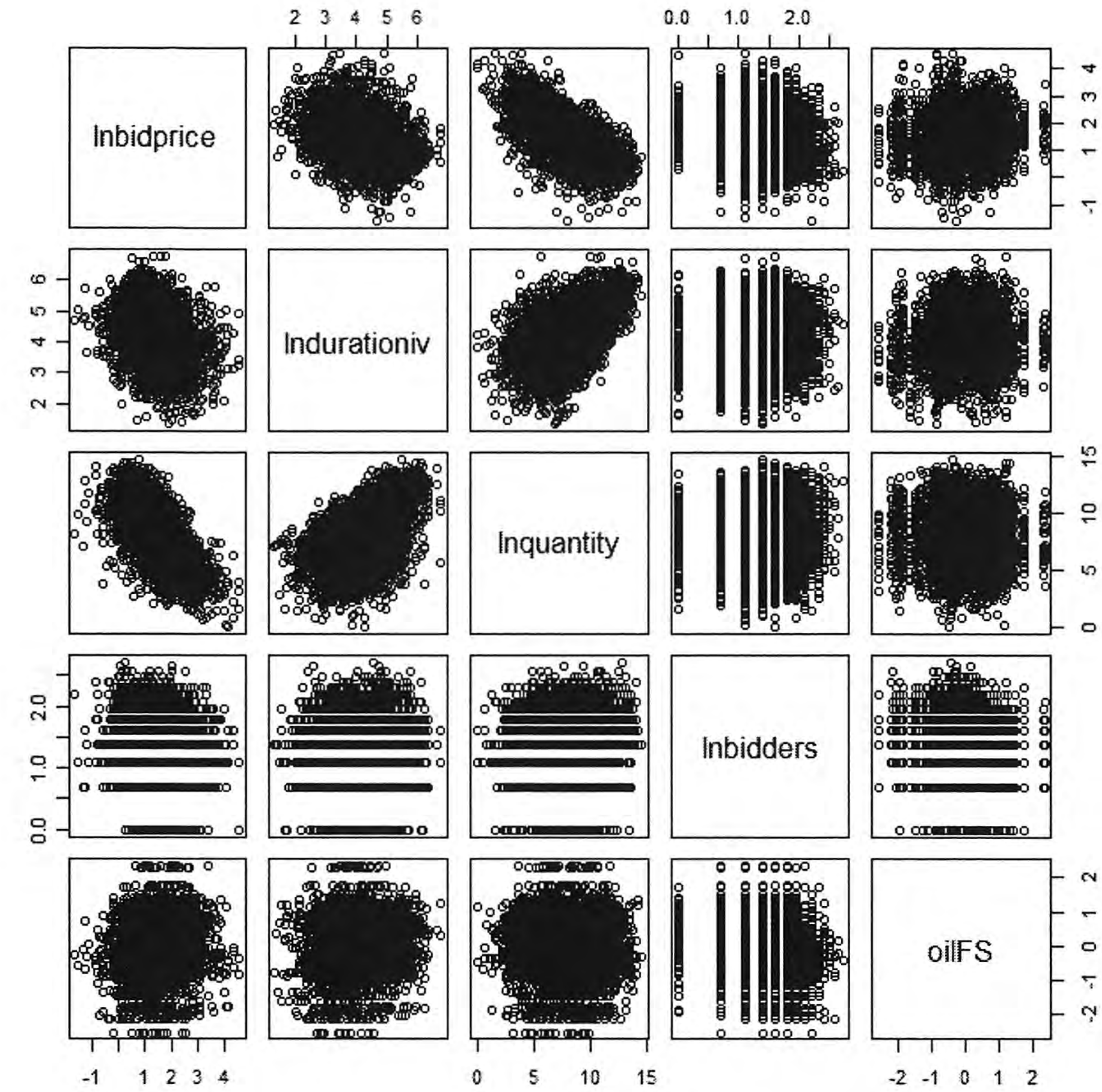


Figure C-4 Scatter plot matrix of transformed data

Figure C-5 contains the plots of the transformed version. It is evident that the log transformations have dramatically reduced skewness and produced variables which are more consistent with normally distributed data.

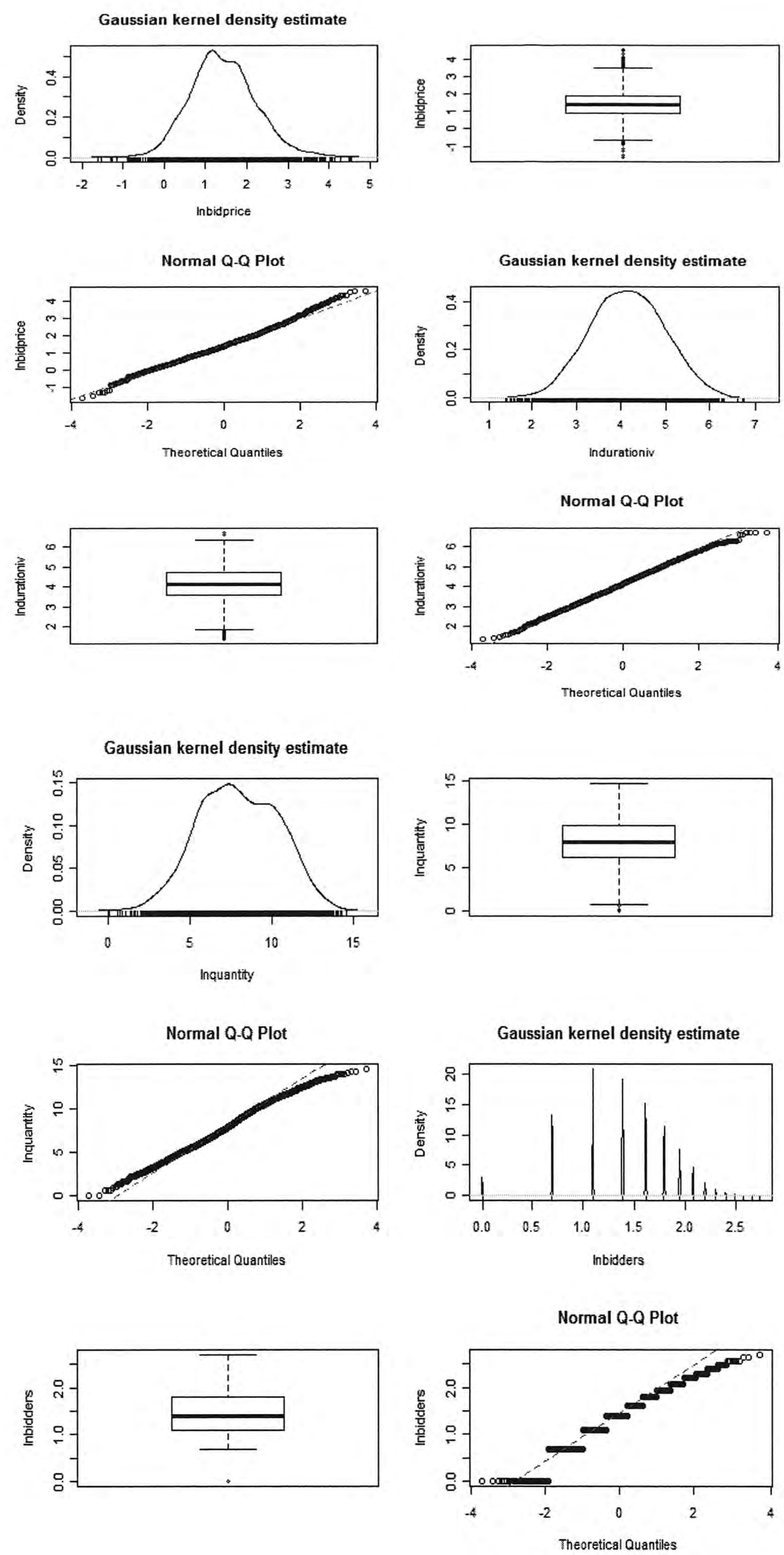


Figure C-5 Box plots, normal QQ-plots and kernel density estimates of transformed data

Step 4: Variable selection

We next consider a multiple linear regression model based on the log transformed data, namely,

$$\log(bidprice) = \gamma_0 + \gamma_1 \log(duration \times i.v.) + \gamma_2 \log(quantity) + \gamma_3 \log(bidders) + \gamma_4(\text{oilFS}) + \gamma_5 quarter1 \cdots + \gamma_7 quarter3 + \gamma_8 district1 \cdots + \gamma_{31} district24 \qquad \text{(C-1)}$$

Backward elimination based on Bayesian information criterion (BIC) is chosen as the variable selection method which aims to choose the subset of the predictors that is "best" in a given sense. BIC is based on likelihood theory when both the predictors and the response are normally distributed. The smaller the value of BIC, the better the model. BIC penalizes complex models more heavily than Akaike's information criterion (AIC), thus it favors simpler model than AIC. Given below are the variable selection results from backward AIC and backward BIC and regression results based on these two methods:

<u>Output from R: Backward selection based on AIC</u>

```
Final Step:  AIC=-3807.27
lnbidprice ~ lndurationiv + lnquantity + lnbidders + oilFS + city + quarter1 + quarter2 + quarter3 + d1 +
d3 + d4 + d5 + d7 + d9 + d10 + d11 + d12 + d13 + d15 + d16 + d18 + d19 + d20 + d21 + d22 + d24
               Df Sum of Sq    RSS     AIC
<none>                      2455.6 -3807.3
- d20           1      0.99 2456.6 -3807.2
- lndurationiv  1      1.08 2456.7 -3807.0
- city          1      1.25 2456.9 -3806.6
- d24           1      1.70 2457.3 -3805.7
- d1            1      1.74 2457.3 -3805.6
- d15           1      2.39 2458.0 -3804.2
- quarter3      1      2.50 2458.1 -3804.0
- d22           1      3.65 2459.3 -3801.6
- d13           1      4.04 2459.7 -3800.8
- quarter1      1      6.28 2461.9 -3796.1
- d3            1      6.42 2462.0 -3795.8
- d5            1      7.54 2463.1 -3793.4
- d16           1      8.89 2464.5 -3790.6
- d18           1     10.97 2466.6 -3786.2
- d9            1     10.99 2466.6 -3786.2
- d11           1     11.15 2466.8 -3785.8
- d7            1     13.14 2468.7 -3781.6
- d10           1     13.55 2469.2 -3780.8
- quarter2      1     14.02 2469.6 -3779.8
- d12           1     15.32 2470.9 -3777.1
- d21           1     29.65 2485.3 -3747.1
- oilFS         1     34.06 2489.7 -3738.0
- d4            1     37.07 2492.7 -3731.7
- lnbidders     1     74.99 2530.6 -3653.5
- d19           1     76.41 2532.0 -3650.6
- lnquantity    1   1607.38 4063.0 -1202.4
```

<u>Regression output from R: based on backward AIC</u>

```
lm(formula = lnbidprice ~ lndurationiv + lnquantity + lnbidders + oilFS + city + quarter1 + quarter2 +
quarter3 + d1 + d3 + d4 + d5 + d7 + d9 + d10 + d11 + d12 + d13 + d15 + d16 + d18 + d19 + d20 + d21 +
d22 + d24, weights = lnmoreone)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	3.445399	0.051758	66.568	< 2e-16	***
lndurationiv	0.016432	0.010923	1.504	0.132571	
lnquantity	-0.216448	0.003728	-58.061	< 2e-16	***
lnbidders	-0.228927	0.018254	-12.541	< 2e-16	***
oilFS	0.082075	0.009711	8.452	< 2e-16	***
city	-0.036307	0.022417	-1.620	0.105372	
quarter1	-0.086909	0.023948	-3.629	0.000287	***
quarter2	-0.123396	0.022757	-5.422	6.15e-08	***
quarter3	-0.049714	0.021701	-2.291	0.022013	*
d1	-0.091818	0.048087	-1.909	0.056264	.
d3	-0.175428	0.047827	-3.668	0.000247	***
d4	0.290475	0.032943	8.818	< 2e-16	***
d5	0.193940	0.048785	3.975	7.12e-05	***
d7	-0.215560	0.041065	-5.249	1.59e-07	***
d9	-0.217982	0.045399	-4.801	1.62e-06	***
d10	0.182584	0.034255	5.330	1.02e-07	***
d11	0.279387	0.057776	4.836	1.37e-06	***
d12	0.182338	0.032170	5.668	1.52e-08	***
d13	0.092582	0.031790	2.912	0.003603	**
d15	0.121878	0.054436	2.239	0.025205	*
d16	-0.180485	0.041810	-4.317	1.61e-05	***
d18	-0.196530	0.040981	-4.796	1.67e-06	***
d19	-0.507028	0.040053	-12.659	< 2e-16	***
d20	0.096070	0.066515	1.444	0.148702	
d21	0.271065	0.034372	7.886	3.78e-15	***
d22	-0.120908	0.043705	-2.766	0.005687	**
d24	0.100337	0.053210	1.886	0.059396	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6905 on 5150 degrees of freedom
Multiple R-squared: 0.5251, Adjusted R-squared: 0.5227
F-statistic: 219 on 26 and 5150 DF, p-value: < 2.2e-16

Output from R: Backward selection based on BIC

Final Step: AIC=-3664.69
lnbidprice ~ lnquantity + lnbidders + oilFS + quarter1 + quarter2 + d1 + d3 + d4 + d5 + d7 + d9 + d10 + d11 + d12 + d16 + d18 + d19 + d21 + d22

	Df	Sum of Sq	RSS	AIC
<none>			2467.8	-3664.7
- quarter1	1	4.14	2471.9	-3664.6
- d5	1	4.16	2471.9	-3664.5
- d1	1	4.26	2472.0	-3664.3
- d11	1	7.24	2475.0	-3658.1
- d10	1	7.37	2475.1	-3657.8

```
- d22        1     7.85 2475.6 -3656.8
- d12        1     9.34 2477.1 -3653.7
- d3         1    11.73 2479.5 -3648.7
- quarter2   1    12.50 2480.2 -3647.1
- d16        1    16.83 2484.6 -3638.1
- d18        1    17.69 2485.4 -3636.3
- d9         1    18.01 2485.8 -3635.6
- d7         1    21.31 2489.1 -3628.7
- d21        1    24.88 2492.6 -3621.3
- d4         1    28.33 2496.1 -3614.1
- oilFS      1    38.13 2505.9 -3593.9
- lnbidders  1    74.12 2541.9 -3520.0
- d19        1   101.47 2569.2 -3464.6
- lnquantity 1  2174.40 4642.2  -402.1
```

Regression output from R: based on backward BIC

```
lm(formula = lnbidprice ~ lnquantity + lnbidders + oilFS + quarter1 + quarter2 + d1 + d3 + d4 + d5 + d7
+ d9 + d10 + d11 + d12 + d16 + d18 + d19 + d21 + d22, weights = lnmoreone)
Residuals:
    Min       1Q   Median       3Q      Max
-4.59275 -0.41300  0.01789  0.43063  2.89643
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.495122   0.038035  91.891  < 2e-16 ***
lnquantity  -0.213586   0.003169 -67.409  < 2e-16 ***
lnbidders   -0.222449   0.017874 -12.445  < 2e-16 ***
oilFS        0.086026   0.009637   8.927  < 2e-16 ***
quarter1    -0.057923   0.019694  -2.941 0.003285 **
quarter2    -0.093361   0.018269  -5.110 3.33e-07 ***
d1          -0.137889   0.046241  -2.982 0.002877 **
d3          -0.227941   0.046043  -4.951 7.63e-07 ***
d4           0.230013   0.029893   7.695 1.69e-14 ***
d5           0.137935   0.046799   2.947 0.003219 **
d7          -0.259426   0.038871  -6.674 2.75e-11 ***
d9          -0.266644   0.043470  -6.134 9.21e-10 ***
d10          0.112088   0.028570   3.923 8.85e-05 ***
d11          0.212189   0.054537   3.891 0.000101 ***
d12          0.129248   0.029253   4.418 1.02e-05 ***
d16         -0.234910   0.039615  -5.930 3.23e-09 ***
d18         -0.234095   0.038505  -6.080 1.29e-09 ***
d19         -0.550830   0.037827 -14.562  < 2e-16 ***
d21          0.228133   0.031639   7.210 6.38e-13 ***
d22         -0.168556   0.041627  -4.049 5.21e-05 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.6918 on 5157 degrees of freedom
Multiple R-squared: 0.5227,    Adjusted R-squared: 0.521
F-statistic: 297.3 on 19 and 5157 DF,  p-value: < 2.2e-16
```

Step 5: Regression model and output

It is shown that BIC penalizes complex models more heavily than AIC, thus it results in simpler model than AIC. Duration-based volatility is chosen by AIC, but not by BIC. Backward elimination with BIC chooses the statistically significant predictors as follows: quantity, number of bidders, the difference between futures and spot prices, quarter 1 and quarter 2 of the year, 14 districts in Texas which are Abilene, Atlanta, Austin, Beaumont, Bryan, Corpus Christi, Dallas, El Paso, Fort Worth, Lufkin, Paris, Pharr, San Antonio and Tyler. If duration-based volatility is included, the regression output is as below, and the regression equation for excavation with some log transformed variables is shown in Equation C-2.

Regression output from R

```
lm(formula = lnbidprice ~ lndurationiv + lnquantity + lnbidders + oilFS + quarter1 + quarter2 + d1 + d3 +
d4 + d5 + d7 + d9 + d10 + d11 + d12 + d16 + d18 + d19 + d21 + d22, weights = lnmoreone)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.443017  0.047437  72.580  < 2e-16 ***
lndurationiv 0.021663  0.010878   1.991 0.046485 *
lnquantity  -0.217740  0.003719 -58.547  < 2e-16 ***
lnbidders   -0.226191  0.017922 -12.621  < 2e-16 ***
oilFS        0.084194  0.009698   8.682  < 2e-16 ***
quarter1    -0.057363  0.019712  -2.910 0.003629 **
quarter2    -0.093401  0.018281  -5.109 3.35e-07 ***
d1          -0.137831  0.046270  -2.979 0.002907 **
d3          -0.225298  0.046091  -4.888 1.05e-06 ***
d4           0.238731  0.030212   7.902 3.34e-15 ***
d5           0.137290  0.046829   2.932 0.003386 **
d7          -0.256516  0.038925  -6.590 4.84e-11 ***
d9          -0.266986  0.043497  -6.138 8.98e-10 ***
d10          0.111703  0.028591   3.907 9.47e-05 ***
d11          0.208297  0.054606   3.815 0.000138 ***
d12          0.130083  0.029273   4.444 9.02e-06 ***
d16         -0.235662  0.039641  -5.945 2.95e-09 ***
d18         -0.240427  0.038654  -6.220 5.37e-10 ***
d19         -0.549740  0.037855 -14.522  < 2e-16 ***
d21          0.230702  0.031643   7.291 3.55e-13 ***
d22         -0.167120  0.041658  -4.012 6.11e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6922 on 5157 degrees of freedom
Multiple R-Squared: 0.5236,    Adjusted R-squared: 0.5217
F-statistic: 283.4 on 20 and 5157 DF,  p-value: < 2.2e-16
```

$$\begin{aligned}\ln(bidprice) &= 3.443 + 0.022\ln(duration \times i.v.) - 0.218\ln(quantity) - 0.226\ln(bidders) \\ &+0.084 oilFS - 0.057 quarter1 - 0.093 quarter2 - 0.138 d1 - 0.225 d3 + 0.239 d4 \\ &+0.137 d5 - 0.257 d7 - 0.267 d9 + 0.112 d10 + 0.208 d11 + 0.13 d12 - 0.236 d16 - 0.24 d18 \\ &-0.55 d19 + 0.231 d21 - 0.167 d22\end{aligned} \quad \text{(C-2)}$$

Step 6: Model validation and regression diagnostics

After fitting a model, the associated regression diagnostics should be examined. The random pattern in Figure C-6 indicates that the model in Equation C-2 is a valid model for the unit bid price. The plot of transformed unit bid price against the fitted values in Figure C-7 provides further evidence that the model in Equation C-2 is a valid model because the straight line fit to this plot provides a reasonable fit.

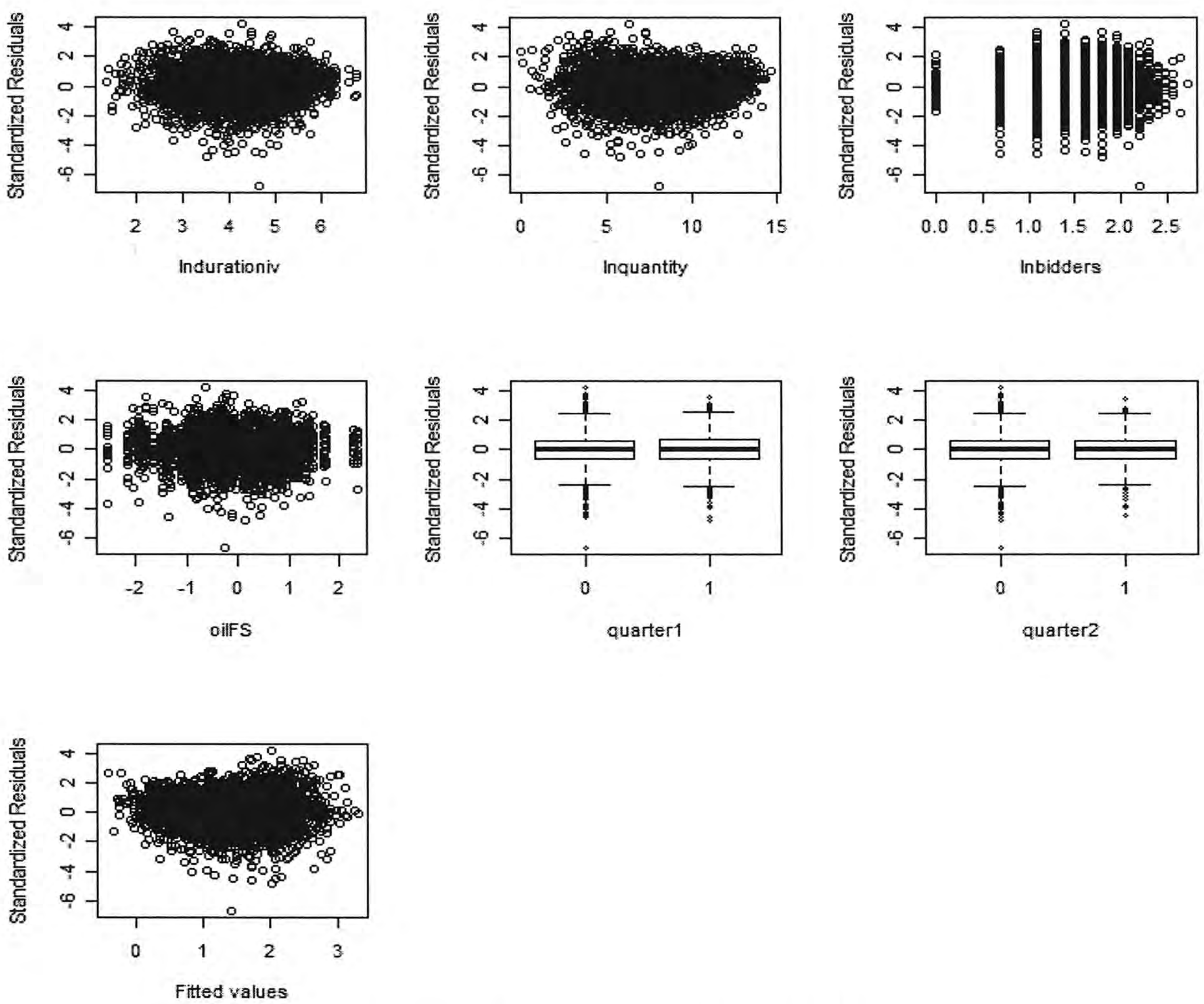


Figure C-6 Plots of residuals

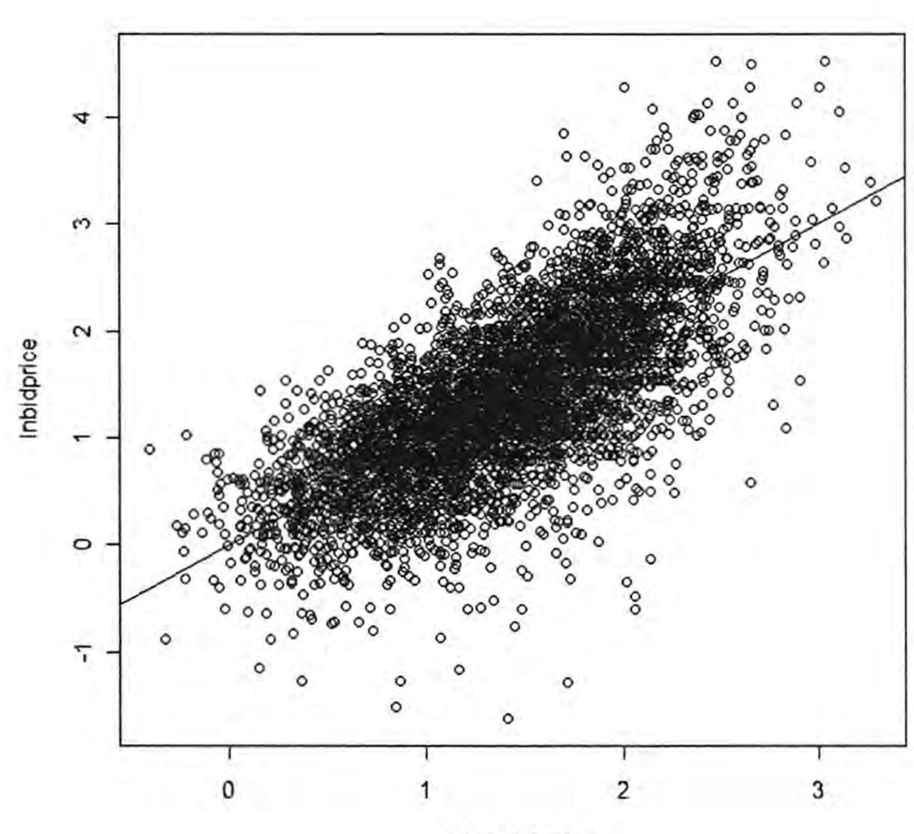


Figure C-7 A plot of transformed unit bid price against fitted values with a straight line added

Figure C-8 shows the diagnostic plots for the model. There is almost a straight line in the normal QQ plot. There are no outliers identified from Figure C-8. These plots further confirm that the model is a valid model for the data.

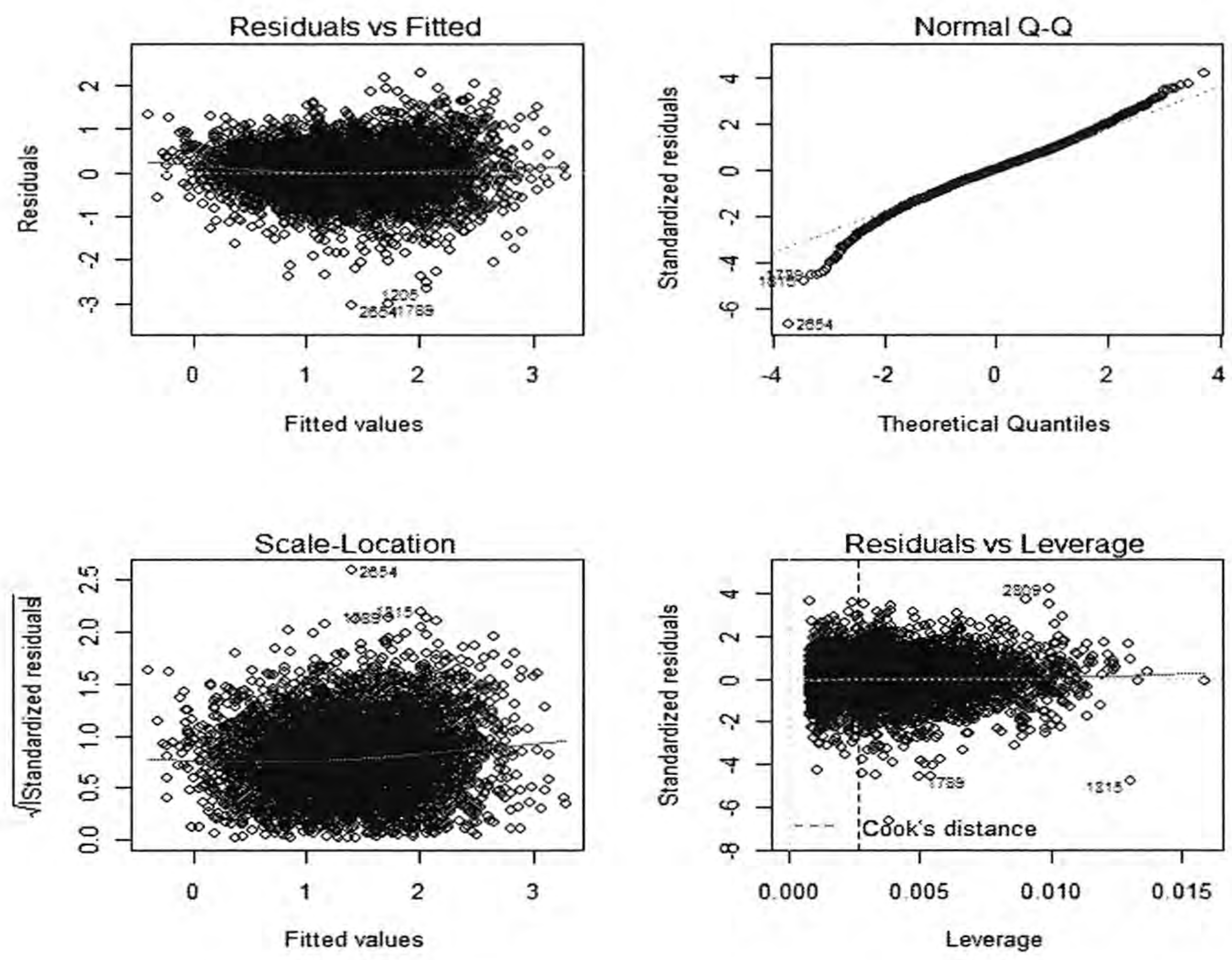


Figure C-8 Diagnostic plots from R

For model fit assessment, marginal model plots shown in Figure C-9 allow for the comparison of the fitted model with a nonparametric model fit with loess. The nonparametric estimates of each pair-wise relationship are marked as solid curves, while the smooths of the fitted values are marked as dashed curves. The two curves in each plot match very well thus providing further evidence that the model is a valid model.

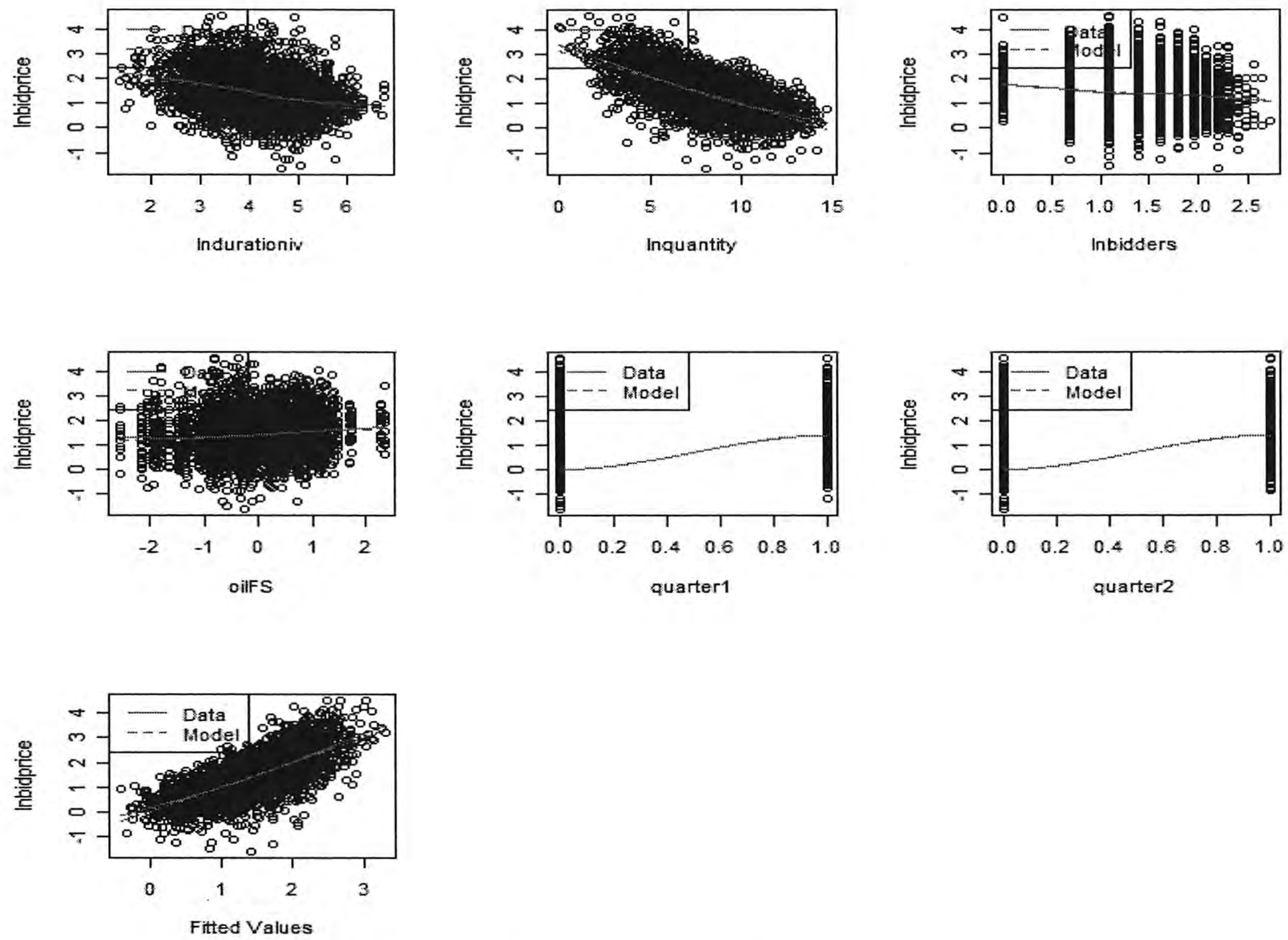


Figure C-9 Marginal model plots

Added variable plots from Figure C-10 below enable us to visually assess the effect of each predictor, having adjusted for the effects of the other predictors. It tells that the variable duration-based implied volatility is not that highly significant as the other independent variables.

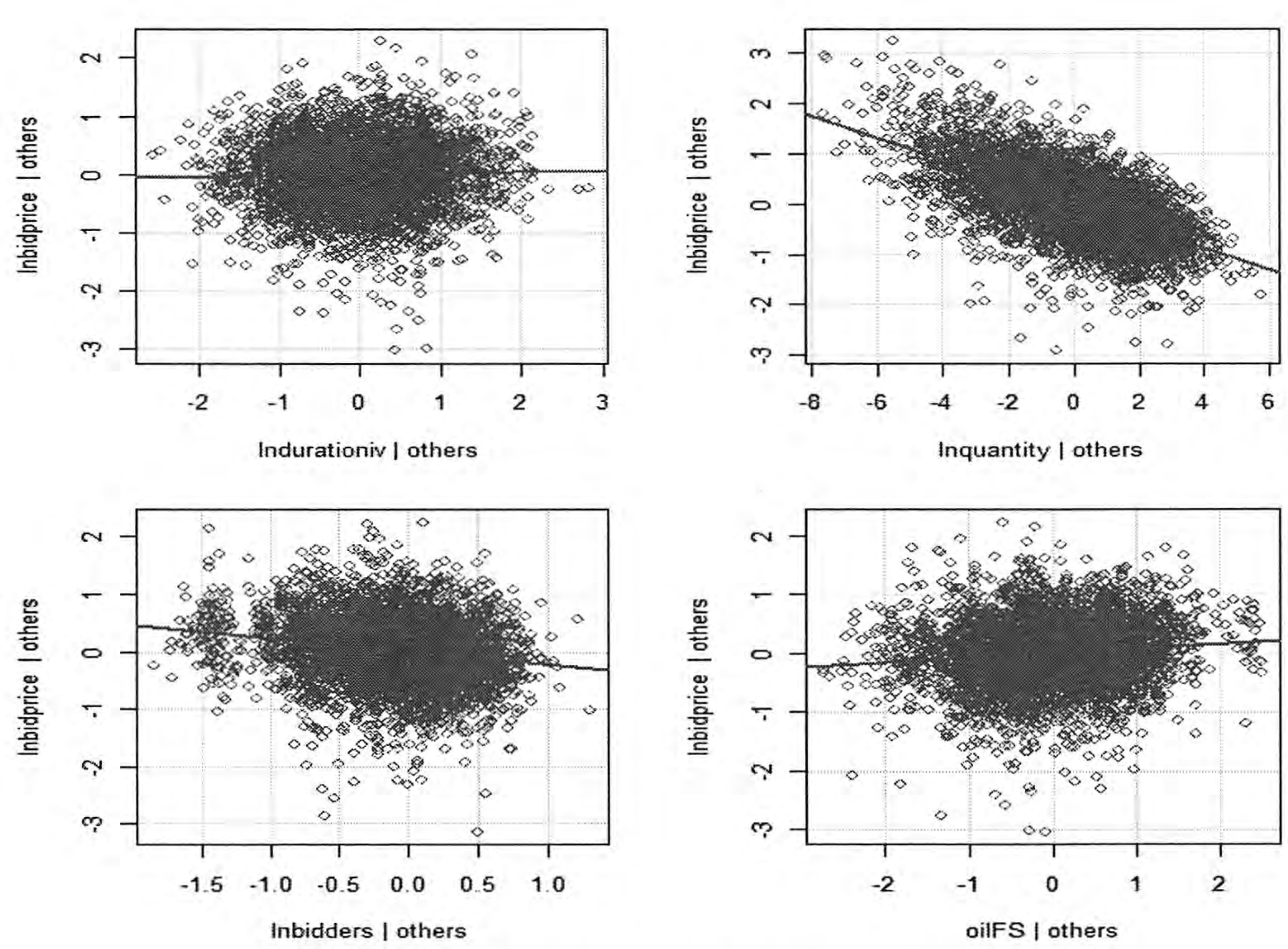


Figure C-10 Added-variable plots

The output from R below gives the correlations between all the independent variables in model 1. Notice none of them is greater than 0.7, thus there is no problem for highly correlated independent variables.

<u>Output from R: Correlations between the predictors</u>

	lndurationiv	lnquantity	lnbidders	oilFS	quarter1	quarter2
lndurationiv	1.000	0.526	0.057	0.077	-0.036	-0.013
lnquantity	0.526	1.000	0.028	-0.008	-0.029	-0.018
lnbidders	0.057	0.028	1.000	-0.138	0.068	-0.011
oilFS	0.077	-0.008	-0.138	1.000	-0.124	0.085
quarter1	-0.036	-0.029	0.068	-0.124	1.000	-0.294
quarter2	-0.013	-0.018	-0.011	0.085	-0.294	1.000
d1	-0.019	-0.029	-0.066	-0.011	-0.020	-0.011
d3	-0.032	-0.031	0.005	-0.009	0.004	0.001
d4	-0.130	-0.008	0.098	0.003	-0.018	0.019
d5	-0.009	-0.028	-0.038	-0.001	0.036	-0.009
d7	-0.063	-0.060	0.003	-0.037	-0.010	-0.022
d9	0.014	0.000	-0.005	-0.008	0.002	0.006
d10	0.039	0.031	0.143	0.033	0.010	-0.051
d11	0.050	0.032	-0.070	-0.001	-0.006	0.020
d12	-0.024	-0.050	0.121	-0.016	-0.002	-0.010
d16	-0.022	-0.071	0.004	0.002	-0.003	0.008
d18	0.042	-0.060	-0.005	0.010	-0.022	-0.008
d19	0.045	0.127	-0.155	-0.031	0.039	0.002
d21	0.081	0.073	0.117	0.014	0.004	-0.011
d22	-0.012	-0.017	0.035	0.006	0.002	0.041

	d1	d3	d4	d5	d7	d9	d10	d11	d12
lndurationiv	-0.019	-0.032	-0.130	-0.009	-0.063	0.014	0.039	0.050	-0.024
lnquantity	-0.029	-0.031	-0.008	-0.028	-0.060	0.000	0.031	0.032	-0.050
lnbidders	-0.066	0.005	0.098	-0.038	0.003	-0.005	0.143	-0.070	0.121
oilFS	-0.011	-0.009	0.003	-0.001	-0.037	-0.008	0.033	-0.001	-0.016
quarter1	-0.020	0.004	-0.018	0.036	-0.010	0.002	0.010	-0.006	-0.002
quarter2	-0.011	0.001	0.019	-0.009	-0.022	0.006	-0.051	0.020	-0.010
d1	1.000	-0.032	-0.052	-0.032	-0.039	-0.034	-0.056	-0.028	-0.054
d3	-0.032	1.000	-0.050	-0.031	-0.037	-0.033	-0.053	-0.027	-0.051
d4	-0.052	-0.050	1.000	-0.050	-0.061	-0.054	-0.088	-0.044	-0.085
d5	-0.032	-0.031	-0.050	1.000	-0.037	-0.033	-0.054	-0.027	-0.052
d7	-0.039	-0.037	-0.061	-0.037	1.000	-0.040	-0.065	-0.033	-0.063
d9	-0.034	-0.033	-0.054	-0.033	-0.040	1.000	-0.057	-0.029	-0.055
d10	-0.056	-0.053	-0.088	-0.054	-0.065	-0.057	1.000	-0.047	-0.091
d11	-0.028	-0.027	-0.044	-0.027	-0.033	-0.029	-0.047	1.000	-0.045
d12	-0.054	-0.051	-0.085	-0.052	-0.063	-0.055	-0.091	-0.045	1.000
d16	-0.038	-0.036	-0.060	-0.036	-0.044	-0.039	-0.064	-0.032	-0.061
d18	-0.039	-0.038	-0.062	-0.038	-0.046	-0.040	-0.066	-0.033	-0.064
d19	-0.044	-0.042	-0.070	-0.043	-0.052	-0.046	-0.075	-0.037	-0.072
d21	-0.048	-0.046	-0.076	-0.046	-0.056	-0.049	-0.081	-0.041	-0.078
d22	-0.035	-0.033	-0.055	-0.034	-0.041	-0.036	-0.059	-0.029	-0.057

	d16	d18	d19	d21	d22
lndurationiv	-0.022	0.042	0.045	0.081	-0.012
lnquantity	-0.071	-0.060	0.127	0.073	-0.017
lnbidders	0.004	-0.005	-0.155	0.117	0.035
oilFS	0.002	0.010	-0.031	0.014	0.006
quarter1	-0.003	-0.022	0.039	0.004	0.002

quarter2	0.008 -0.008 0.002 -0.011 0.041
d1	-0.038 -0.039 -0.044 -0.048 -0.035
d3	-0.036 -0.038 -0.042 -0.046 -0.033
d4	-0.060 -0.062 -0.070 -0.076 -0.055
d5	-0.036 -0.038 -0.043 -0.046 -0.034
d7	-0.044 -0.046 -0.052 -0.056 -0.041
d9	-0.039 -0.040 -0.046 -0.049 -0.036
d10	-0.064 -0.066 -0.075 -0.081 -0.059
d11	-0.032 -0.033 -0.037 -0.041 -0.029
d12	-0.061 -0.064 -0.072 -0.078 -0.057
d16	1.000 -0.045 -0.051 -0.055 -0.040
d18	-0.045 1.000 -0.053 -0.057 -0.041
d19	-0.051 -0.053 1.000 -0.064 -0.047
d21	-0.055 -0.057 -0.064 1.000 -0.047
d22	-0.040 -0.041 -0.047 -0.047 1.000

Table C-1 shows that there is no VIF greater than 5, thus there is no evidence of multicollinearity and thus the associated regression coefficients are well estimated.

Table C-1 VIF test

Variable symbol	*Qn*	*N*	*VD*	*oilFS*	T_1	T_2
VIF	1.44	1.17	1.46	1.06	1.12	1.11

Regression output 1 from R – Item 1 roadway excavation

lm(formula = lnbidprice ~ lndurationiv + lnquantity + lnbidders + oilFS + quarter1 + quarter2 + d1 + d3 + d4 + d5 + d7 + d9 + d10 + d11 + d12 + d16 + d18 + d19 + d21 + d22, weights = lnmoreone)

Coefficients:

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.443017   0.047437  72.580  < 2e-16 ***
lndurationiv 0.021663   0.010878   1.991 0.046485 *
lnquantity  -0.217740   0.003719 -58.547  < 2e-16 ***
lnbidders   -0.226191   0.017922 -12.621  < 2e-16 ***
oilFS        0.084194   0.009698   8.682  < 2e-16 ***
quarter1    -0.057363   0.019712  -2.910 0.003629 **
quarter2    -0.093401   0.018281  -5.109 3.35e-07 ***
d1          -0.137831   0.046270  -2.979 0.002907 **
d3          -0.225298   0.046091  -4.888 1.05e-06 ***
d4           0.238731   0.030212   7.902 3.34e-15 ***
d5           0.137290   0.046829   2.932 0.003386 **
d7          -0.256516   0.038925  -6.590 4.84e-11 ***
d9          -0.266986   0.043497  -6.138 8.98e-10 ***
d10          0.111703   0.028591   3.907 9.47e-05 ***
d11          0.208297   0.054606   3.815 0.000138 ***
d12          0.130083   0.029273   4.444 9.02e-06 ***
d16         -0.235662   0.039641  -5.945 2.95e-09 ***
d18         -0.240427   0.038654  -6.220 5.37e-10 ***
```

```
d19         -0.549740   0.037855 -14.522  < 2e-16 ***
d21          0.230702   0.031643   7.291 3.55e-13 ***
d22         -0.167120   0.041658  -4.012 6.11e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6922 on 5157 degrees of freedom
Multiple R-Squared: 0.5236,     Adjusted R-squared: 0.5217
F-statistic: 283.4 on 20 and 5157 DF,  p-value: < 2.2e-16
```

Regression output 2 from R – Item 2 roadway embankment

```
lm(formula = lnbidprice ~ lndurationiv + lnquantity + lnbidders + oilFS + d4 + d5 + d10 + d11 +
d19 + d21 + d24, weights = lnmoreone)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.398554   0.054036  62.894  < 2e-16 ***
lndurationiv 0.030186   0.013345   2.262   0.0237 *
lnquantity  -0.221569   0.004603 -48.139  < 2e-16 ***
lnbidders   -0.179334   0.020827  -8.611  < 2e-16 ***
oilFS        0.026496   0.005866   4.517 6.44e-06 ***
d4           0.159101   0.034077   4.669 3.11e-06 ***
d5           0.273954   0.052102   5.258 1.52e-07 ***
d10          0.170406   0.030915   5.512 3.74e-08 ***
d11          0.261466   0.063348   4.127 3.73e-05 ***
d19         -0.649760   0.045623 -14.242  < 2e-16 ***
d21         -0.145095   0.038982  -3.722   0.0002 ***
d24          0.263631   0.058008   4.545 5.64e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7969 on 4723 degrees of freedom
Multiple R-squared: 0.4504,     Adjusted R-squared: 0.4491
F-statistic: 351.9 on 11 and 4723 DF,  p-value: < 2.2e-16
```

Regression output 3 from R – Item 3 flexible base

```
lm(formula = lnbidprice ~ lnquantity + lnbidders + oilFS +  quarter1 + quarter2 + d5 + d7 + d9 +
d10 + d13 + d16 + d18 + d19 + d20 + d22, weights = lnmoreone)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.006707   0.039729 100.852  < 2e-16 ***
lnquantity  -0.131916   0.003542 -37.243  < 2e-16 ***
lnbidders   -0.122044   0.016483  -7.404 1.64e-13 ***
oilFS        0.007963   0.004619   1.724  0.08481 .
quarter1    -0.085556   0.018217  -4.696 2.75e-06 ***
quarter2    -0.048778   0.017213  -2.834  0.00463 **
```

```
d5          0.523574  0.052220  10.026  < 2e-16 ***
d7          0.143547  0.024829   5.781 8.05e-09 ***
d9          0.222671  0.035699   6.237 4.97e-10 ***
d10         0.222995  0.038431   5.802 7.11e-09 ***
d13         0.249893  0.042021   5.947 3.00e-09 ***
d16         0.181547  0.032923   5.514 3.75e-08 ***
d18         0.153311  0.030984   4.948 7.85e-07 ***
d19        -0.452601  0.033127 -13.663  < 2e-16 ***
d20        -0.261683  0.056640  -4.620 3.97e-06 ***
d22         0.213198  0.039146   5.446 5.50e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5377 on 3533 degrees of freedom
Multiple R-squared: 0.4006,     Adjusted R-squared: 0.3981
F-statistic: 157.4 on 15 and 3533 DF,  p-value: < 2.2e-16
```

Regression output 4 from R – Item 4 HMAC

```
lm(formula = lnbidprice ~ lnquantity + lnbidders + oilFS + quarter1 + d2 + d4 + d5 + d6 + d8 +
d11 + d14 + d15 + d16 + d19 + d22, weights = lnmoreone)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.439502  0.017967 247.099  < 2e-16 ***
lnquantity  -0.125377  0.001677 -74.743  < 2e-16 ***
lnbidders   -0.125336  0.007369 -17.008  < 2e-16 ***
oilFS        0.015740  0.002116   7.437 1.13e-13 ***
quarter1    -0.032372  0.007696  -4.206 2.62e-05 ***
d2           0.137415  0.020089   6.840 8.45e-12 ***
d4          -0.109398  0.010744 -10.182  < 2e-16 ***
d5           0.140972  0.020906   6.743 1.65e-11 ***
d6           0.104600  0.027243   3.840 0.000124 ***
d8           0.217247  0.027779   7.821 5.89e-15 ***
d11          0.100755  0.024535   4.107 4.05e-05 ***
d14          0.101270  0.027232   3.719 0.000201 ***
d15          0.265093  0.031920   8.305  < 2e-16 ***
d16          0.145954  0.016359   8.922  < 2e-16 ***
d19         -0.225863  0.018493 -12.213  < 2e-16 ***
d22          0.060857  0.014770   4.120 3.82e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3819 on 8428 degrees of freedom
Multiple R-squared: 0.4419,     Adjusted R-squared: 0.4409
F-statistic: 444.8 on 15 and 8428 DF,  p-value: < 2.2e-16
```

Regression output 5 from R – Item 5 regular beams

```
lm(formula = lnbidprice ~ lndurationsteelhv + lnquantity + lnbidders + steelFS + d2 + d15 + d24,
weights = lnmoreone)

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        4.423065   0.047972  92.201  < 2e-16 ***
lndurationsteelhv  0.114010   0.009822  11.607  < 2e-16 ***
lnquantity        -0.099652   0.006374 -15.634  < 2e-16 ***
lnbidders         -0.126318   0.017582  -7.184 1.01e-12 ***
steelFS            0.097699   0.014918   6.549 7.67e-11 ***
d2                 0.171923   0.056570   3.039  0.00241 **
d15                0.382221   0.124235   3.077  0.00213 **
d24                0.155129   0.048748   3.182  0.00149 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4088 on 1675 degrees of freedom
Multiple R-squared: 0.2347,     Adjusted R-squared: 0.2315
F-statistic: 73.38 on 7 and 1675 DF,  p-value: < 2.2e-16
```

Regression output 6 from R – Item 6 CRCP

```
lm(formula = lnbidprice ~ lnduration + cementFS + lnquantity + lnbidders + d5 + d10 + d11 +
d12 + d13 + d19, weights = lnmoreone)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.962821   0.093860  63.529  < 2e-16 ***
lnduration  -0.057216   0.016945  -3.377 0.000766 ***
cementFS     0.147692   0.031878   4.633 4.14e-06 ***
lnquantity  -0.134569   0.006082 -22.127  < 2e-16 ***
lnbidders   -0.097646   0.027684  -3.527 0.000441 ***
d5          -0.184005   0.066445  -2.769 0.005735 **
d10         -0.116521   0.034214  -3.406 0.000690 ***
d11         -0.207599   0.054071  -3.839 0.000132 ***
d12         -0.155797   0.036525  -4.265 2.21e-05 ***
d13         -0.375442   0.030702 -12.229  < 2e-16 ***
d19         -0.262584   0.079353  -3.309 0.000974 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4322 on 890 degrees of freedom
Multiple R-squared: 0.5627,     Adjusted R-squared: 0.5578
F-statistic: 114.5 on 10 and 890 DF,  p-value: < 2.2e-16
```

Regression output 7 from R – Item 7 retaining wall

```
lm(formula = lnbidprice ~ lndurationcementhv + lnquantity + lnbidders + cementFS + d1 + d8 +
d19, weights = lnmoreone)

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         3.60153    0.08421  42.767  < 2e-16 ***
lndurationcementhv  0.13921    0.01916   7.266 6.87e-13 ***
lnquantity         -0.10460    0.00897 -11.661  < 2e-16 ***
lnbidders          -0.08518    0.03150  -2.704  0.00696 **
cementFS            0.02623    0.04274   0.614  0.53956
d1                 -0.26597    0.09816  -2.709  0.00684 **
d8                 -0.28460    0.10173  -2.798  0.00524 **
d19                -0.21716    0.08120  -2.674  0.00759 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6362 on 1138 degrees of freedom
Multiple R-squared: 0.1371,     Adjusted R-squared: 0.1318
F-statistic: 25.83 on 7 and 1138 DF,  p-value: < 2.2e-16
```

APPENDIX D

AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODEL

Table D-1 Tentative Order Selection by ESACF Option - cement price

ARMA(p+d,q) Tentative Order Selection Tests

--ESACF--

p+d	q
1	0

(5% Significance Level)

Table D-2 Tentative Order Selection by SCAN and MINIC Option - steel price

Minimum Table Value: BIC(1,0) = -1.53415

ARMA(p+d,q) Tentative Order Selection Tests

---------SCAN--------

p+d	q	BIC
1	0	-1.53415

(5% Significance Level)

Table D-3 Tentative Order Selection by SCAN Option - oil price

ARMA(p+d,q) Tentative Order Selection Tests

---------SCAN--------

p+d	q	BIC
1	1	2.227354

(5% Significance Level)

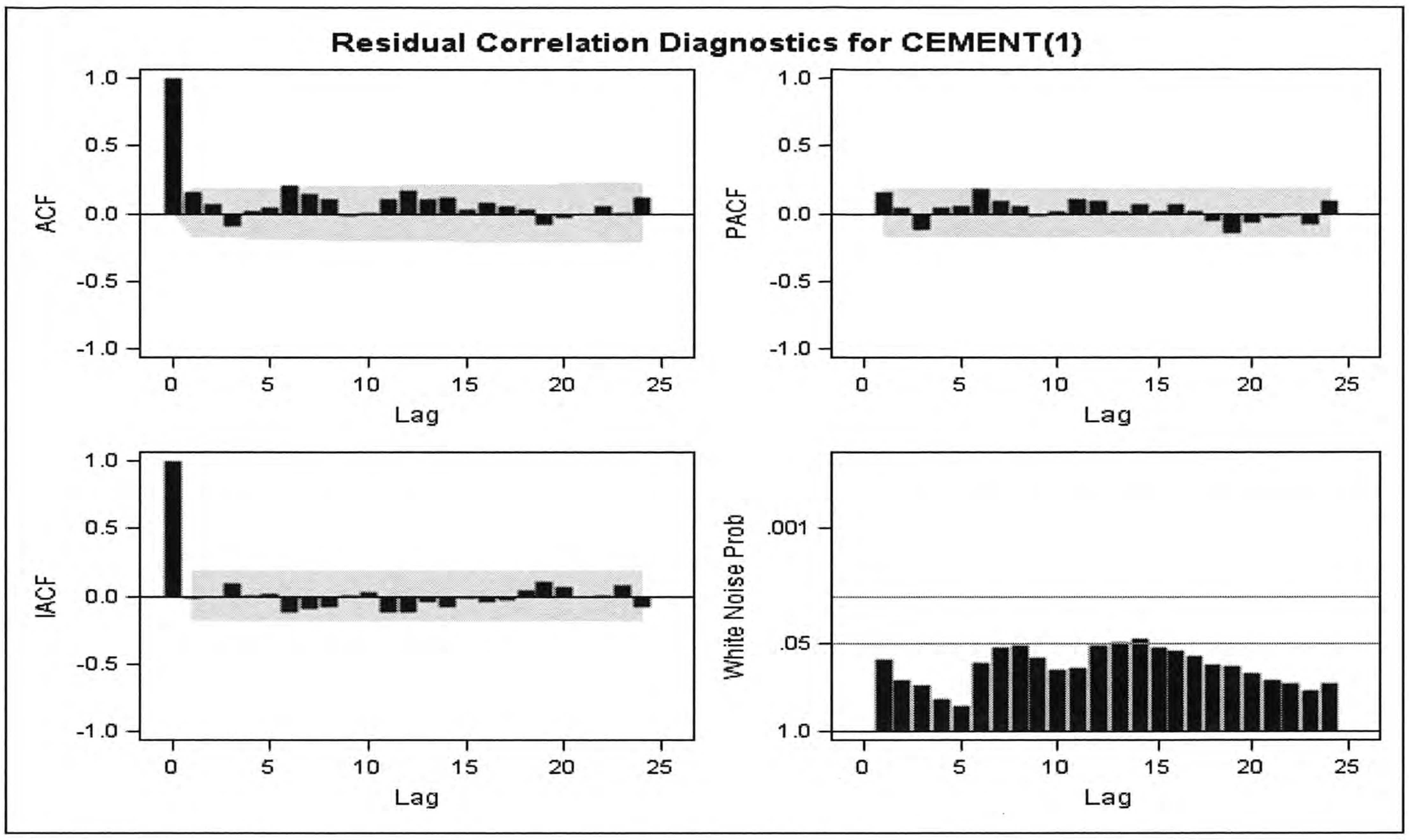


Figure D-1 White Noise Check of Residuals for the model of cement price

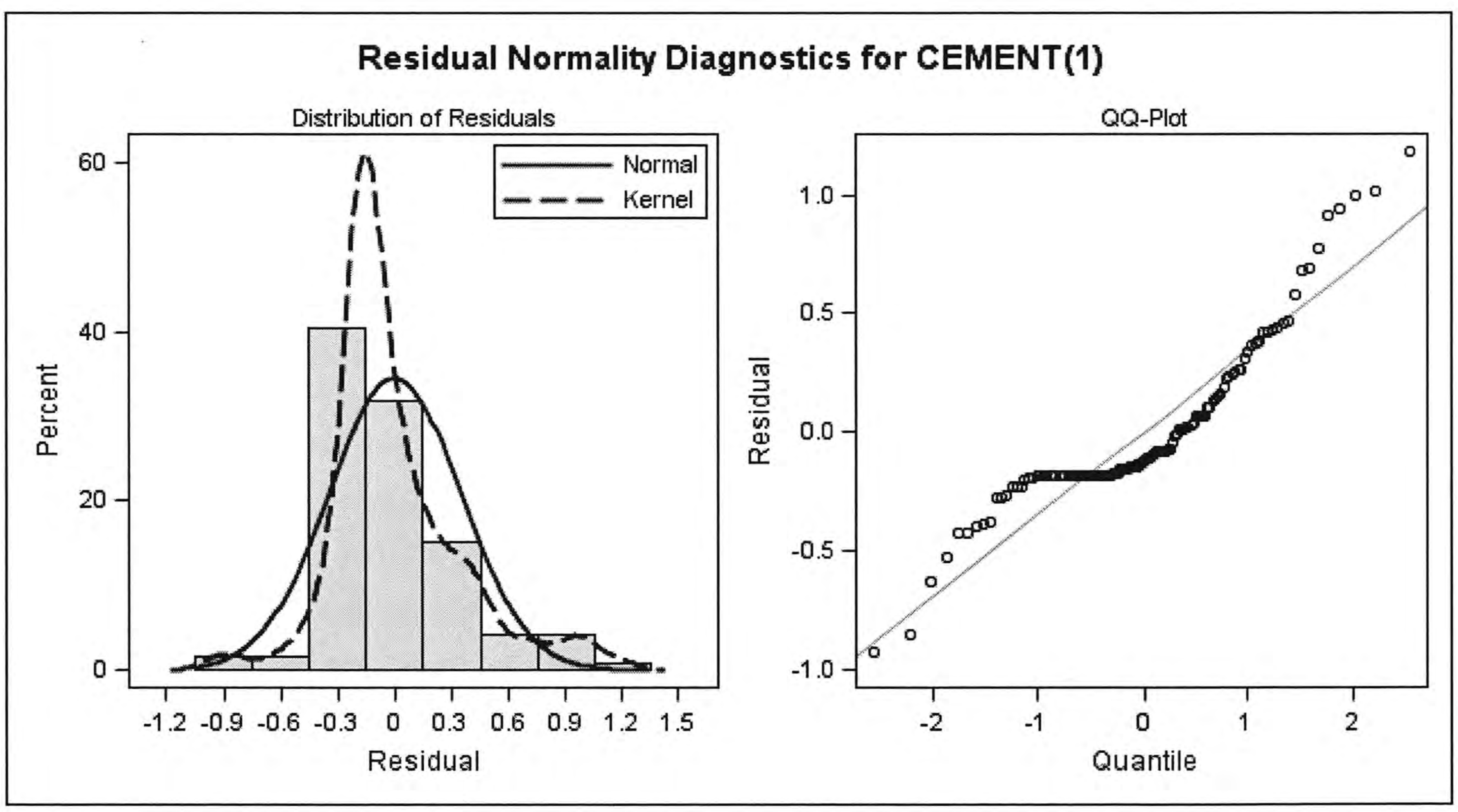


Figure D-2 Normality Check of Residuals for the model of cement price

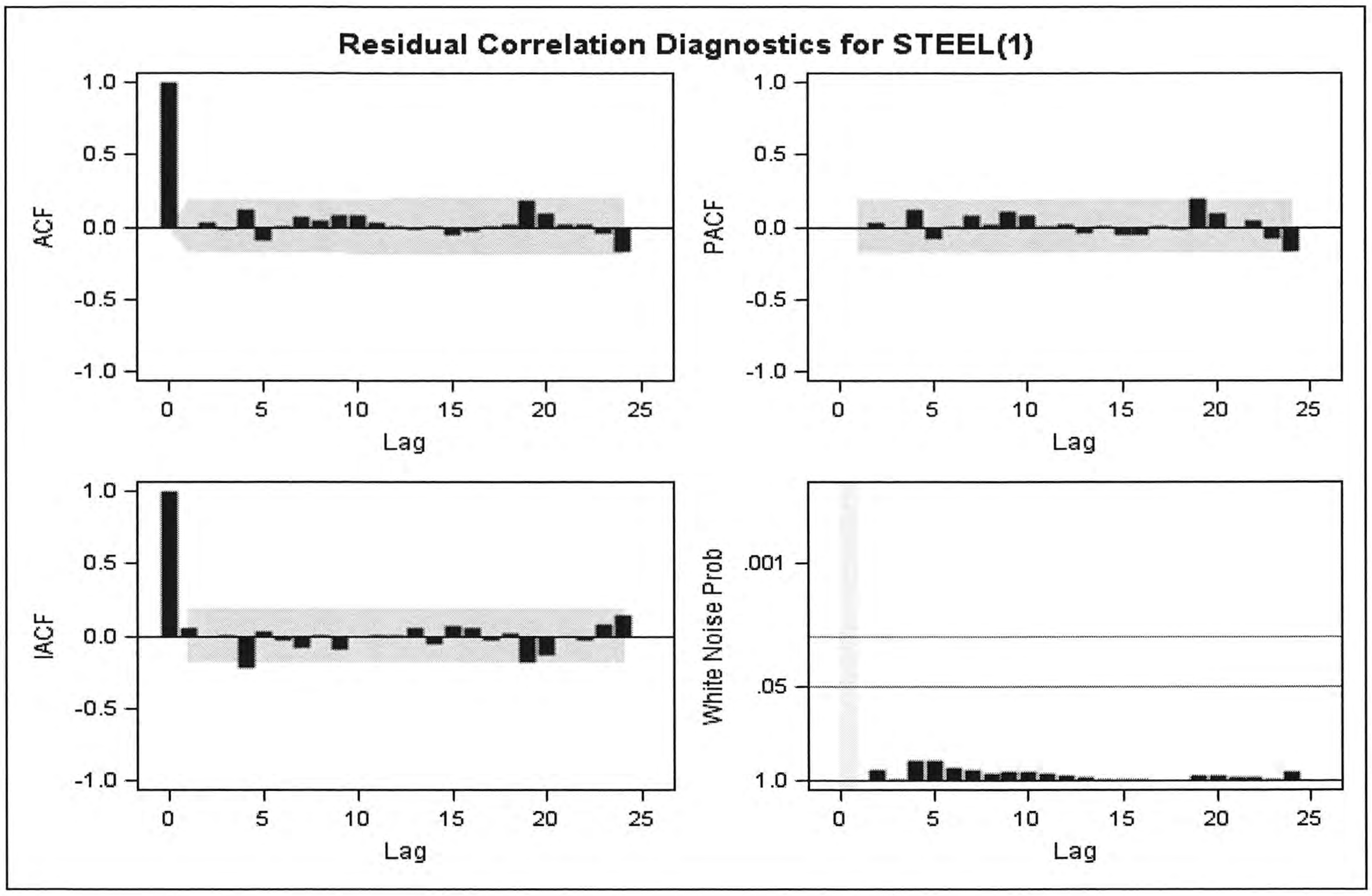


Figure D-3 White Noise Check of Residuals for the model of steel price

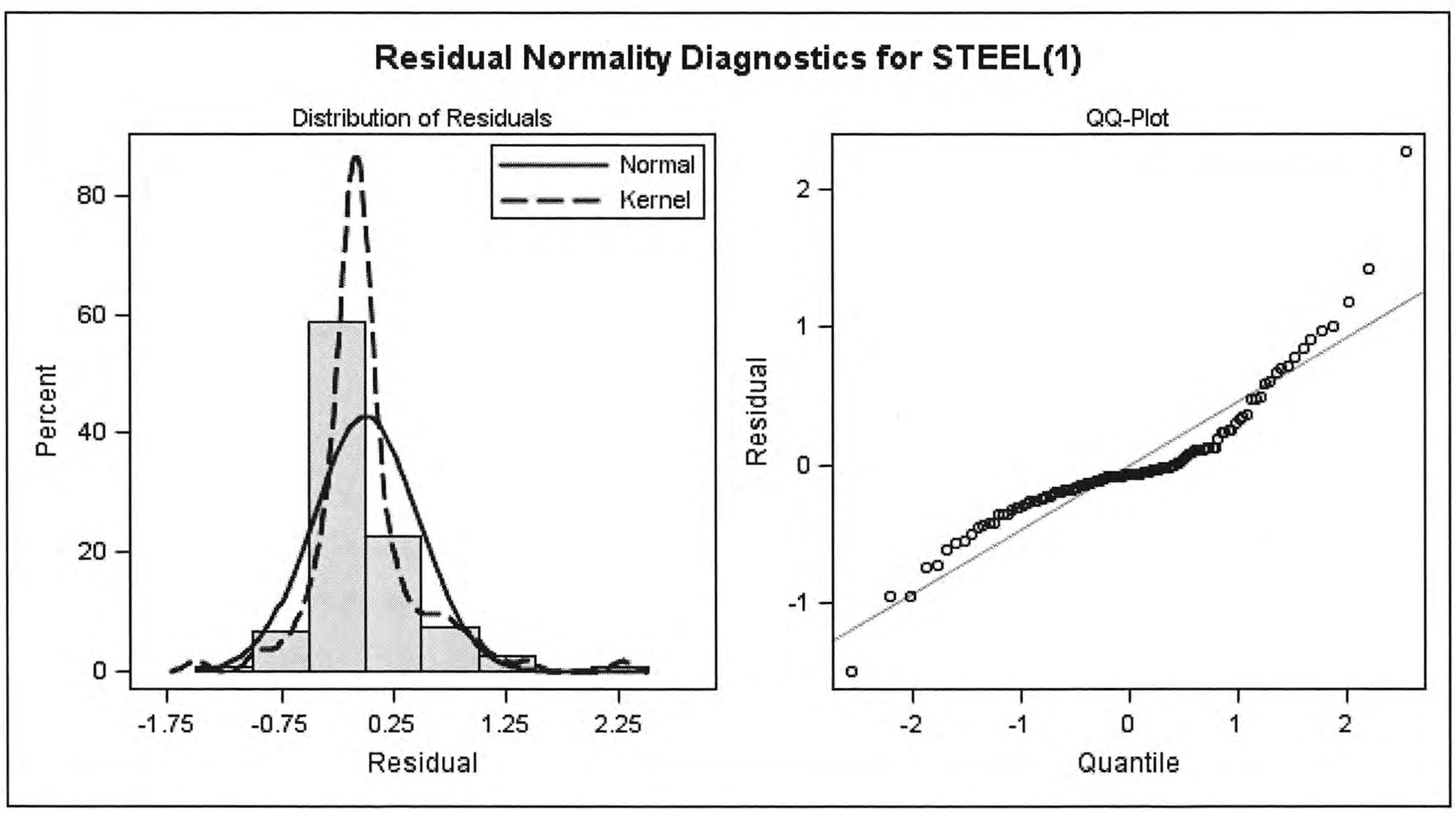


Figure D-4 Normality Check of Residuals for the model of steel price

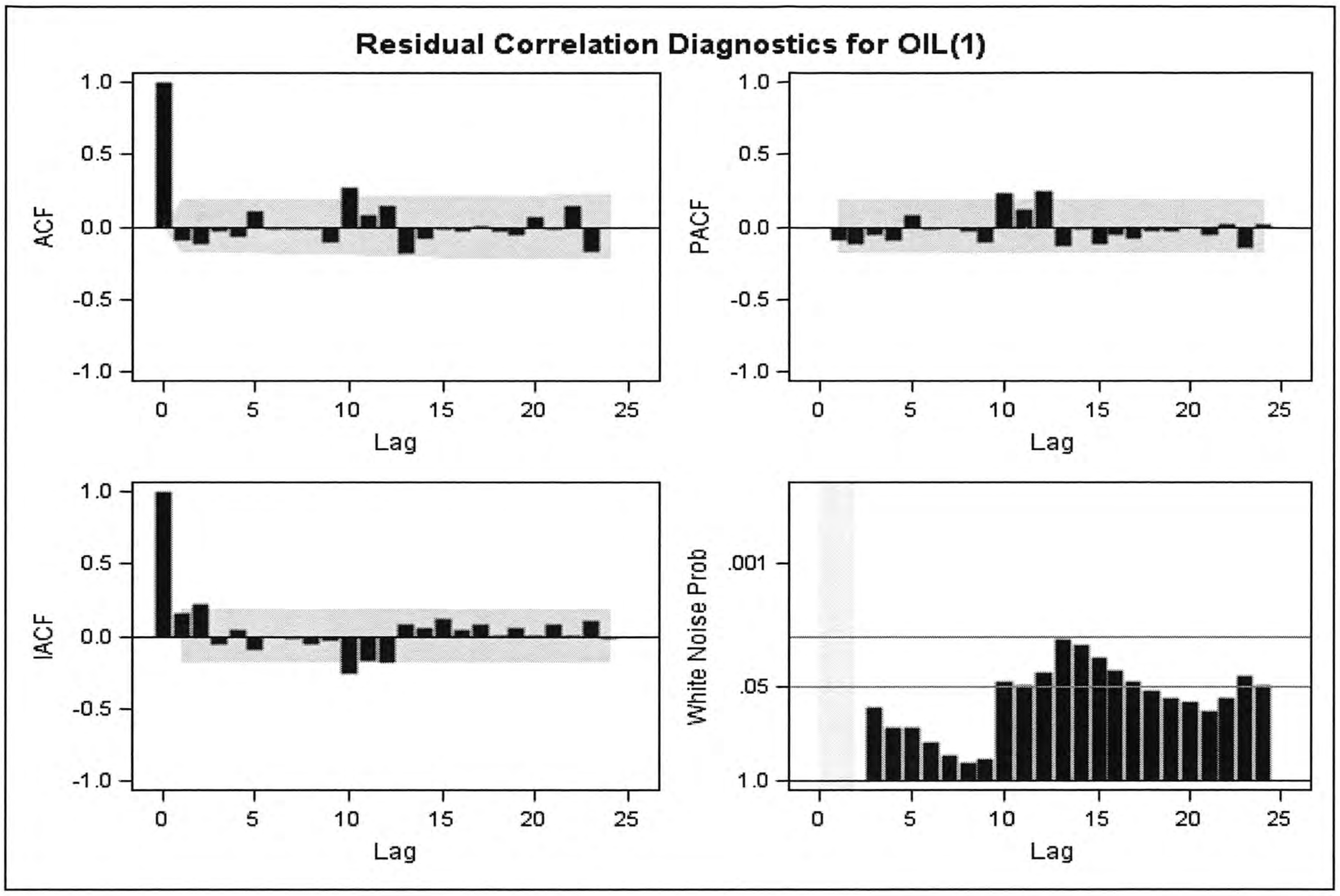


Figure D-5 White Noise Check of Residuals for the model of oil price

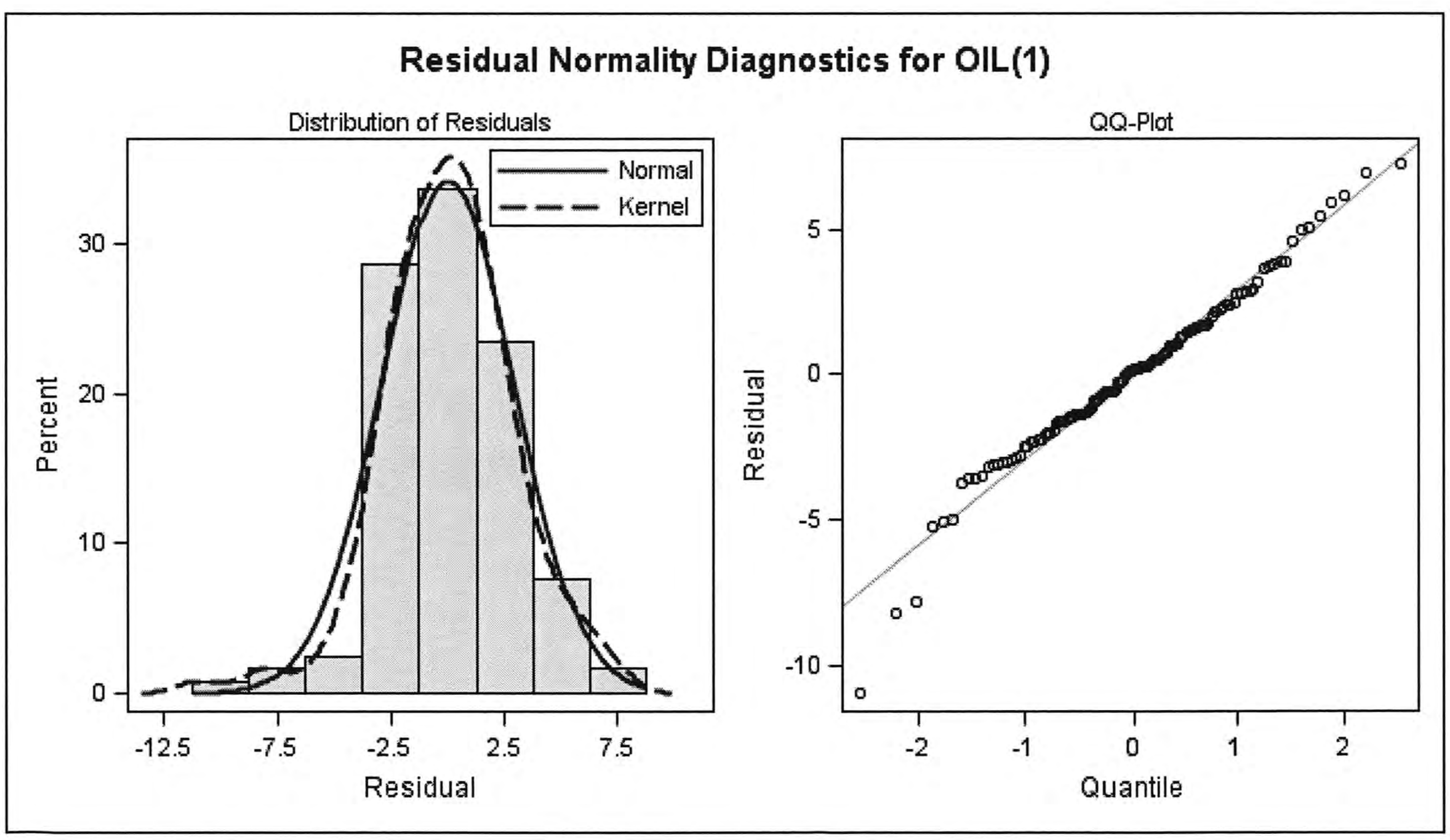


Figure D-6 Normality Check of Residuals for the model of oil price

APPENDIX E

VECTOR AUTOREGRESSIVE MOVING AVERAGE (VARMA) MODEL

This appendix provides supplement computed materials supporting the developed VARMA model in Section 5.

Step 1: Tentative order selection

Figure E-1 gives the result of tentative order selection suggesting the VAR model with AR order 2 and no MA term according to the smallest value of AICC.

Minimum Information Criterion Based on AICC

Lag	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	8.1986197	7.6733172	7.586941	7.4491416	7.4044994	7.3762644
AR 1	-1.535598	-1.502097	-1.524127	-1.475135	-1.397069	-1.284298
AR 2	-1.560388	-1.551874	-1.42439	-1.420295	-1.346016	-1.239576
AR 3	-1.479959	-1.461403	-1.355284	-1.328222	-1.198507	-1.136581
AR 4	-1.433103	-1.442941	-1.290802	-1.168553	-1.001633	-1.026524
AR 5	-1.313427	-1.380836	-1.210052	-1.033199	-0.846832	-0.862312

Figure E-1 Result of MIMIC method

Figure E-2 shows again that the model can be obtained by an AR order m=2 since partial autoregression matrices are insignificant after lag 2 with respect to two standard errors.

```
Schematic Representation of Partial Autoregression
Variable/
Lag          1      2      3      4      5      6

CEMENT       +..    ...    ...    ...    ..-
STEEL        .+.    .-+    -..    ...    ..-
OIL          ..+    ...    -..    ...    ...

             + is > 2*std error,  - is <
             -2*std error,  . is between
```

Figure E-2 Partial Autoregression Matrices

The partial cross-correlation matrices in Figure E-3 are insignificant after lag 2 with respect to two standard errors. This indicates again that an AR order of m=2 can be an appropriate choice.

Partial Cross Correlations by Variable

Variable	Lag	CEMENT	STEEL	OIL
CEMENT	1	0.84643	-0.10121	0.06048
	2	-0.03811	-0.10462	0.03682
	3	0.01664	-0.18347	-0.15591
	4	0.04475	0.06515	0.03064
	5	0.05447	0.06002	0.02778
	6	0.01730	-0.14354	0.08643
STEEL	1	-0.04895	0.91512	0.01581
	2	0.00289	-0.27550	-0.15767
	3	0.11428	-0.01940	0.05928
	4	-0.01433	0.03595	-0.07773
	5	0.02863	0.06480	-0.02765
	6	0.01457	0.13353	-0.00392
OIL	1	0.05680	0.06316	0.93964
	2	0.03860	0.27196	-0.00834
	3	-0.07130	0.06062	0.09738
	4	-0.00898	-0.02800	-0.05860
	5	-0.15005	-0.24998	-0.06832
	6	0.00948	0.05669	-0.12928

Schematic Representation of
Partial Cross Correlations

Variable/ Lag	1	2	3	4	5	6
CEMENT	+..	...	...	...	...	
STEEL	.+.	.-.	...	...	...	
OIL	..+	.+.	...	...	.-.	

+ is > 2*std error, - is < -2*std error, . is between

Figure E-3 Partial Cross Correlation

Figure E-4 shows that after lag m=2, the partial canonical correlations are insignificant with respect to the 0.05 significance level, indicating again that an AR order of m=2 can be an appropriate choice.

Partial Canonical Correlations

Lag	Correlation1	Correlation2	Correlation3	DF	Chi-Square	Pr > ChiSq
1	0.96950	0.96350	0.76961	9	292.81	<.0001
2	0.41403	0.13594	0.01185	9	22.43	0.0076
3	0.26888	0.13622	0.05082	9	10.93	0.2804
4	0.10922	0.08183	0.01999	9	2.21	0.9878
5	0.31716	0.04530	0.01252	9	11.82	0.2235
6	0.23448	0.10425	0.02445	9	7.57	0.5775

Figure E-4 Partial Canonical Correlations

Step 3: Model Estimation and Model Diagnostic Check

The VARMAX Procedure

Type of Model	VECM(2)
Estimation Method	Maximum Likelihood Estimation
Cointegrated Rank	1

Long-Run Parameter Beta

Estimates When RANK=1

Variable	1
CEMENT	1.00000
STEEL	-3.06727
OIL	-0.26142

Adjustment Coefficient
Alpha Estimates
When RANK=1

Variable	1
CEMENT	-0.01133
STEEL	-0.00382
OIL	0.00164

Figure E-5 Parameter Estimates

Model Parameter Estimates

Equation	Parameter	Estimate	Standard Error	t Value	Pr > |t|	Variable
D_CEMENT	AR1_1_1	-0.01133	0.00188			CEMENT(t-1)
	AR1_1_2	0.03474	0.00578			STEEL(t-1)
	AR1_1_3	0.00296	0.00049			OIL(t-1)
	AR2_1_1	0.03876	0.09163	0.42	0.6731	D_CEMENT(t-1)
	AR2_1_2	-0.06105	0.06490	-0.94	0.3489	D_STEEL(t-1)
	AR2_1_3	-0.00700	0.01009	-0.69	0.4894	D_OIL(t-1)
D_STEEL	AR1_2_1	-0.00382	0.00264			CEMENT(t-1)
	AR1_2_2	0.01173	0.00810			STEEL(t-1)
	AR1_2_3	0.00100	0.00069			OIL(t-1)
	AR2_2_1	0.16181	0.12843	1.26	0.2103	D_CEMENT(t-1)
	AR2_2_2	0.21958	0.09097	2.41	0.0174	D_STEEL(t-1)
	AR2_2_3	-0.03096	0.01414	-2.19	0.0306	D_OIL(t-1)
D_OIL	AR1_3_1	0.00164	0.01726			CEMENT(t-1)
	AR1_3_2	-0.00502	0.05293			STEEL(t-1)
	AR1_3_3	-0.00043	0.00451			OIL(t-1)
	AR2_3_1	0.81131	0.83908	0.97	0.3357	D_CEMENT(t-1)
	AR2_3_2	0.94565	0.59433	1.59	0.1144	D_STEEL(t-1)
	AR2_3_3	0.04093	0.09236	0.44	0.6585	D_OIL(t-1)

Figure E-6 Parameter Estimates Continued

Figure E-7 shows the innovation covariance matrix estimates, the various information criteria results, and the tests for white noise residuals. The residuals do not have very significant correlations at lag 2 or lag 3. These results show that a VECM(2) model fits well with the data.

The VARMAX Procedure

Covariances of Innovations

Variable	CEMENT	STEEL	OIL
CEMENT	0.10275	0.00882	-0.11830
STEEL	0.00882	0.20185	-0.10131
OIL	-0.11830	-0.10131	8.61578

Information
Criteria

AICC	-1.53614

Schematic Representation of Cross
Correlations of Residuals

Variable/ Lag	0	1	2	3	4	5	6
CEMENT	+..	...	...	-..	...	...	...
STEEL	.+.	...	...	...	...	...	...
OIL	..+	...	..-	...	.+.	...	...

+ is > 2*std error, - is < -2*std error, . is between

Portmanteau Test for Cross Correlations of Residuals

Up To Lag	DF	Chi-Square	Pr > ChiSq
3	9	26.39	0.0018
4	18	40.21	0.0020
5	27	50.88	0.0036
6	36	58.90	0.0094

Figure E-7 Diagnostic Checks

Figure E-8 describes how well each univariate equation fits the data. The residuals for cement and steel are off from the normality. There are no AR effects on other residuals. Except the residuals for oil, there are no ARCH effects on other residuals.

The VARMAX Procedure

Univariate Model ANOVA Diagnostics

Variable	R-Square	Standard Deviation	F Value	Pr > F
CEMENT	0.1443	0.32055	3.78	0.0034
STEEL	0.1317	0.44927	3.40	0.0068
OIL	0.0275	2.93527	0.63	0.6743

Univariate Model White Noise Diagnostics

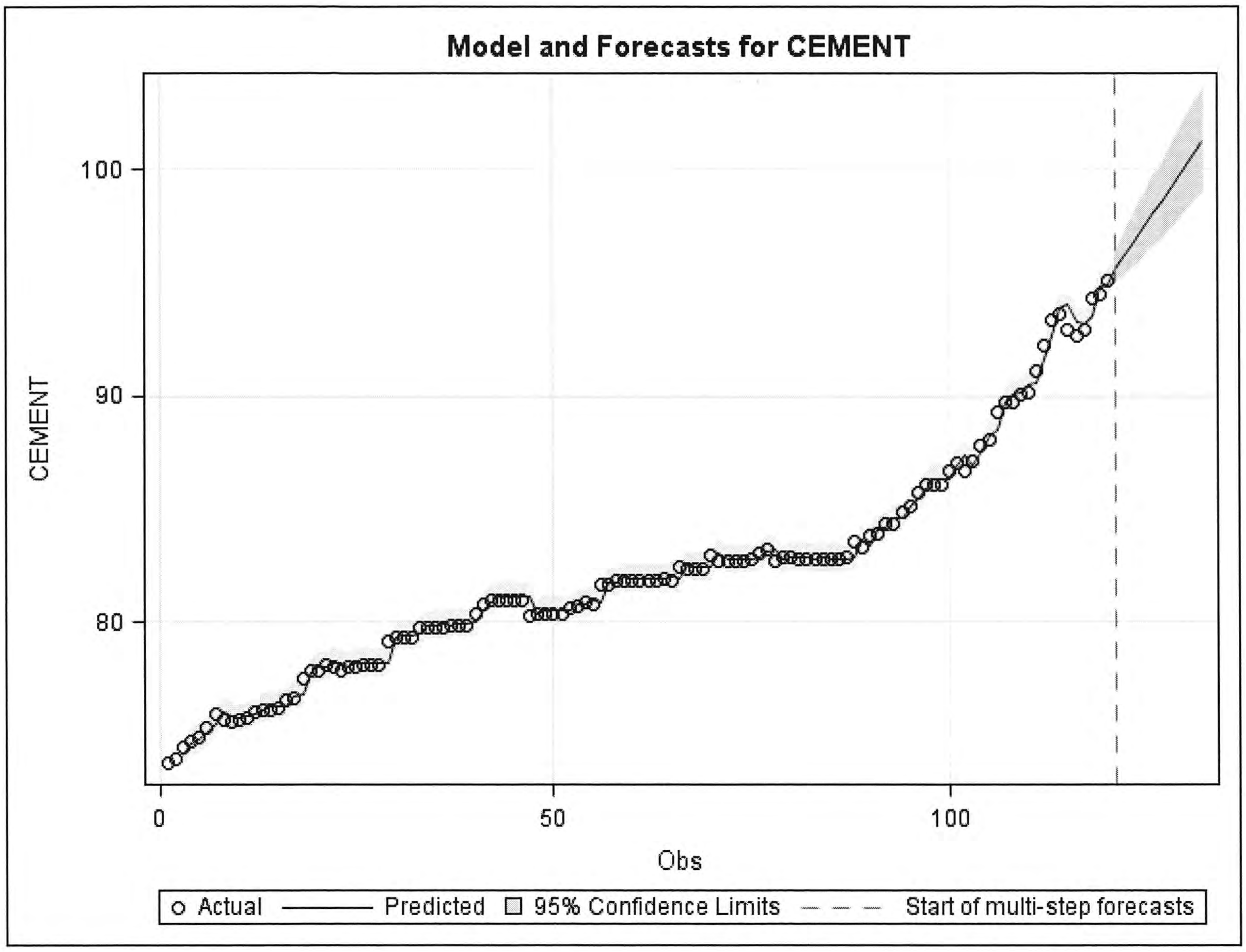
Variable	Durbin Watson	Normality Chi-Square	Normality Pr > ChiSq	ARCH F Value	ARCH Pr > F
CEMENT	1.99249	22.04	<.0001	0.18	0.6731
STEEL	1.96865	116.60	<.0001	26.09	<.0001
OIL	2.00441	11.37	0.0034	0.02	0.8756

Univariate Model AR Diagnostics

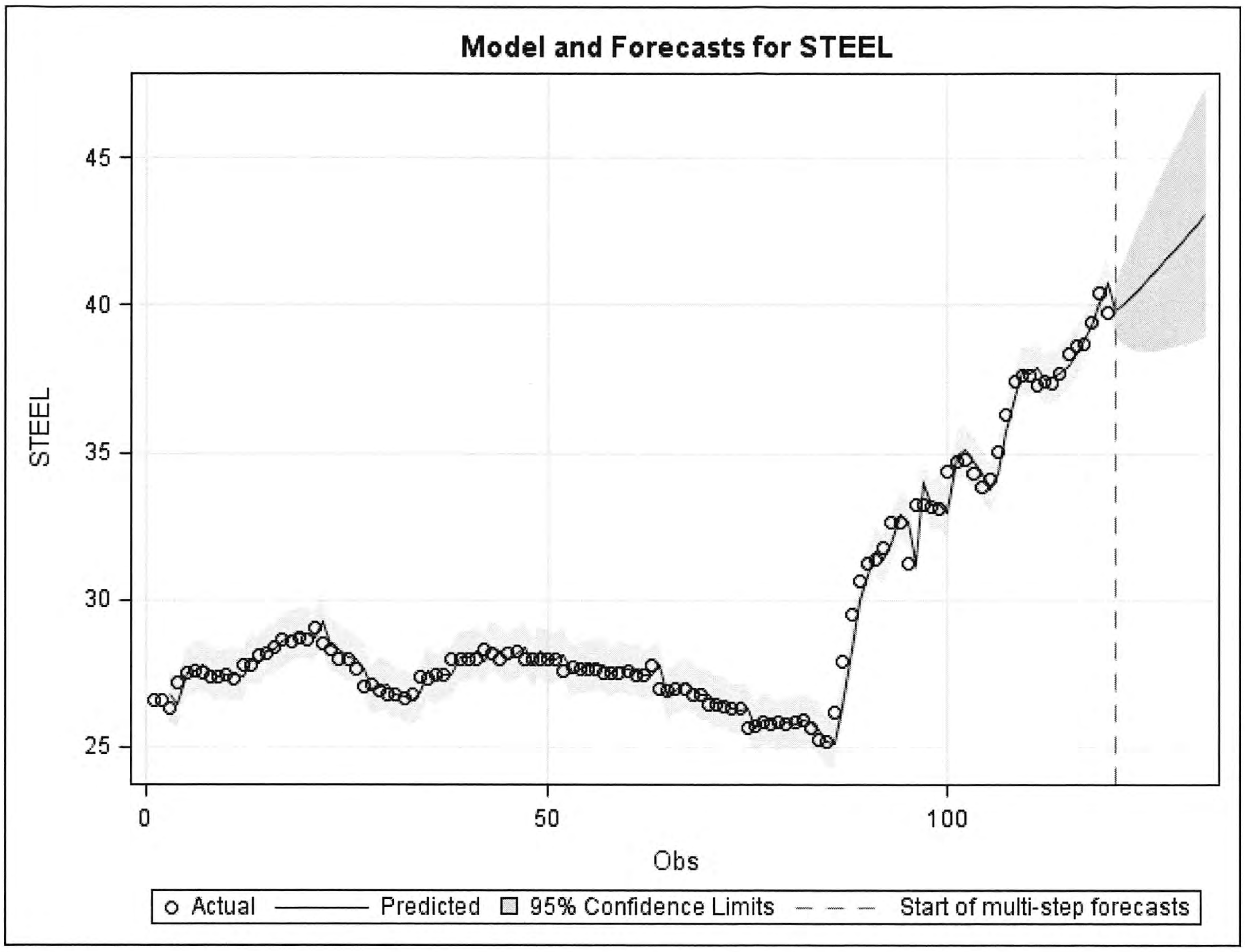
AR1 AR2 AR3 AR4

Variable	F Value	Pr > F	F Value	Pr > F	F Value	Pr > F	F Value	Pr > F
CEMENT	0.00	0.9748	0.11	0.8980	1.75	0.1607	1.69	0.1570
STEEL	0.01	0.9155	0.45	0.6418	0.30	0.8285	0.48	0.7469
OIL	0.01	0.9255	2.80	0.0651	1.83	0.1462	2.53	0.0449

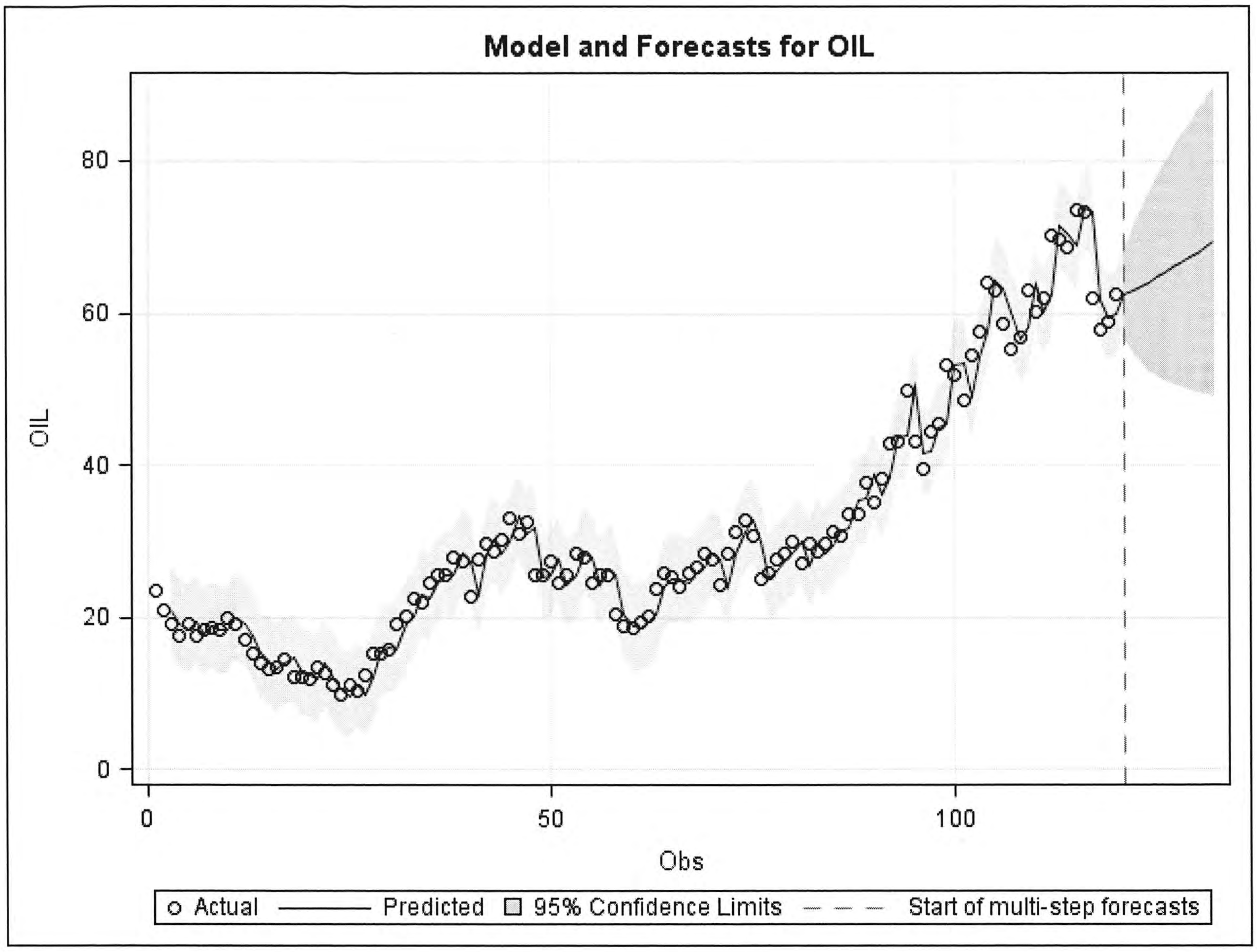
Figure E-8 Diagnostic Checks Continued



(a)


Model and Forecasts for STEEL
45
40
35
30
25
STEEL
0
50
100
Obs
Actual
Predicted
95% Confidence Limits
Start of multi-step forecasts

(b)



(c)

Figure E-9 The fitted time series and the forecast trend plots

APPENDIX F

LOSS SIMULATIONS BASED ON UNIVARIATE TIME SERIES MODEL AND VECTOR TIME SERIES MODEL

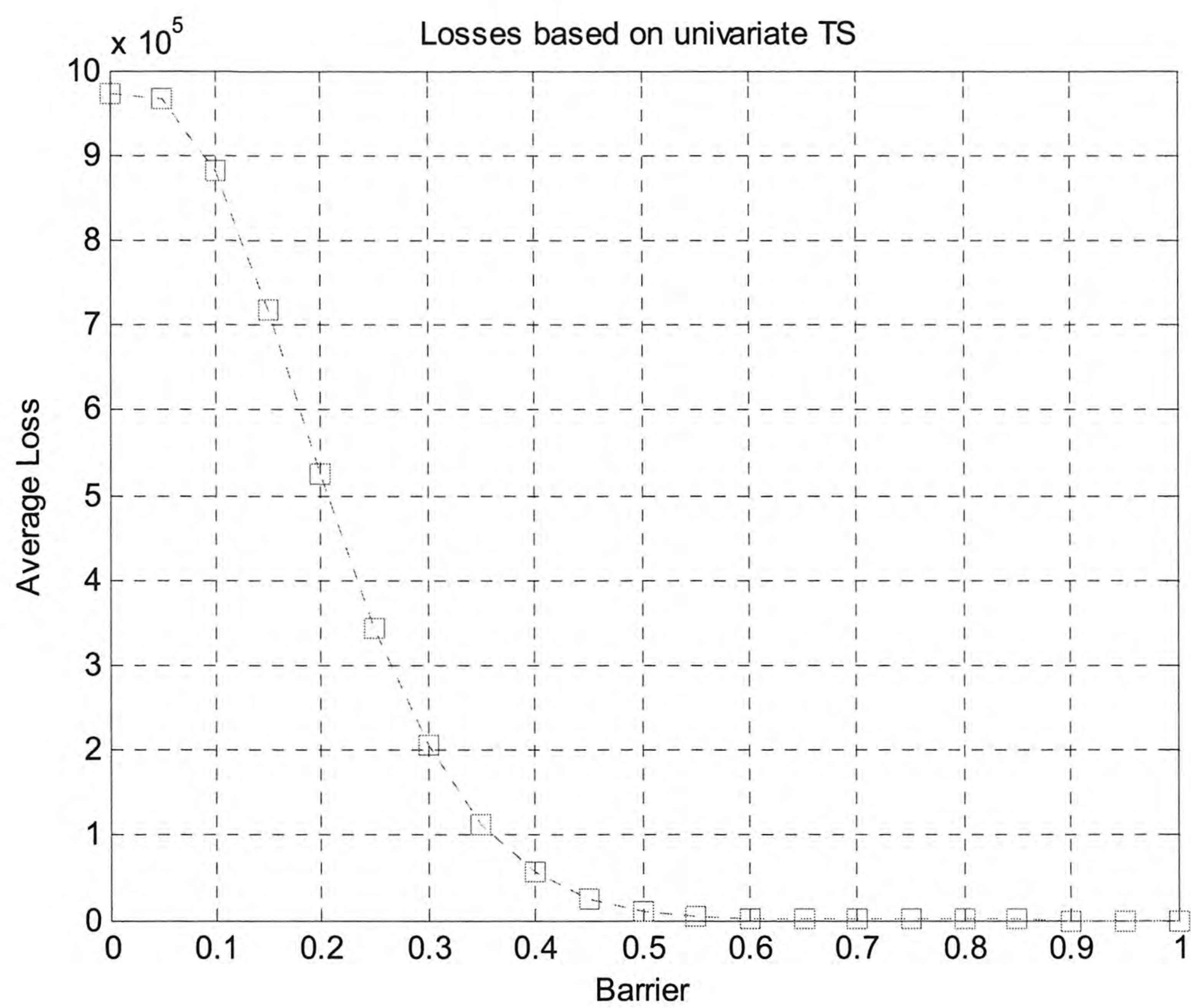


Figure F-1 "Expected losses" for embankment

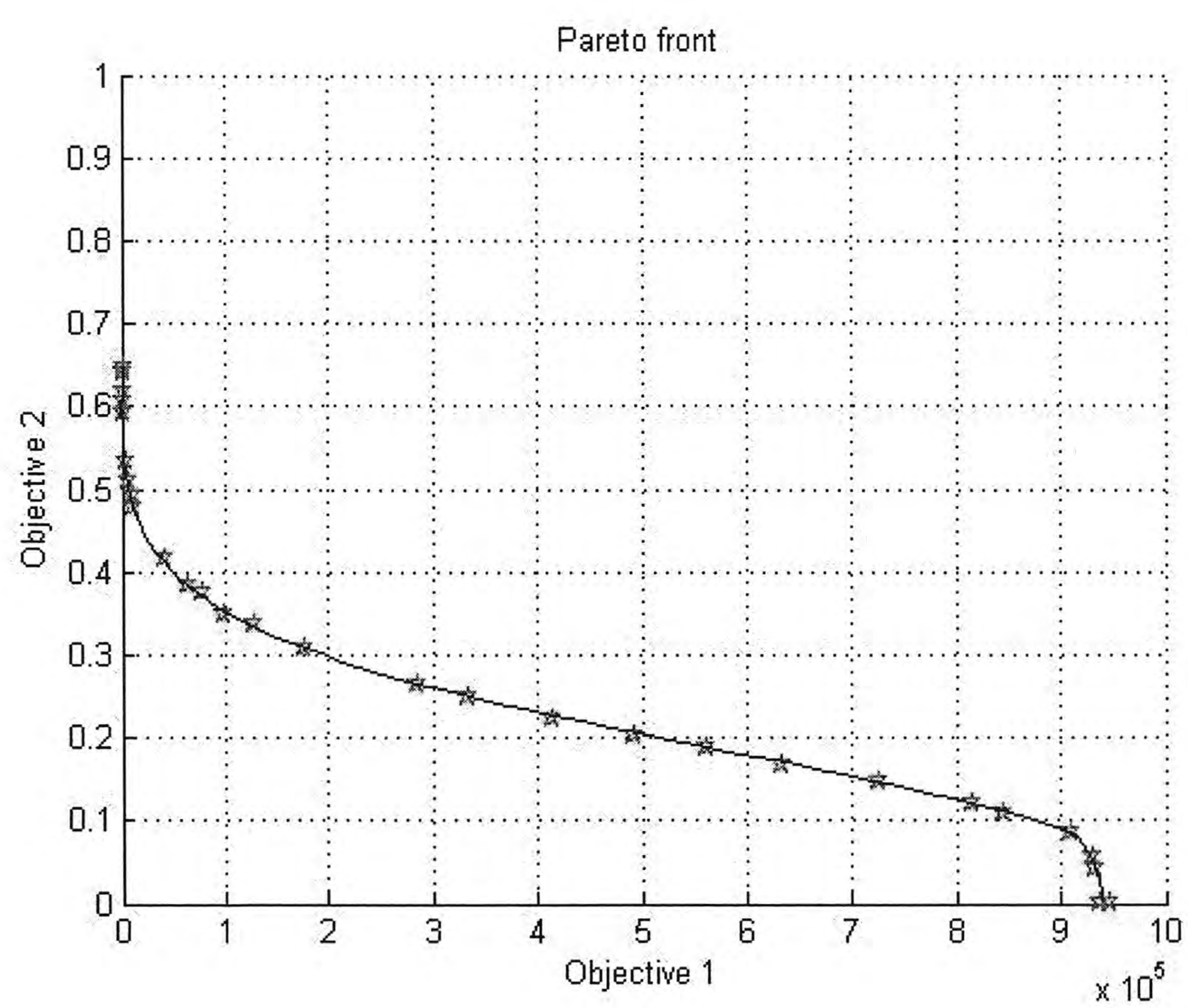


Figure F-2 Pareto front for embankment

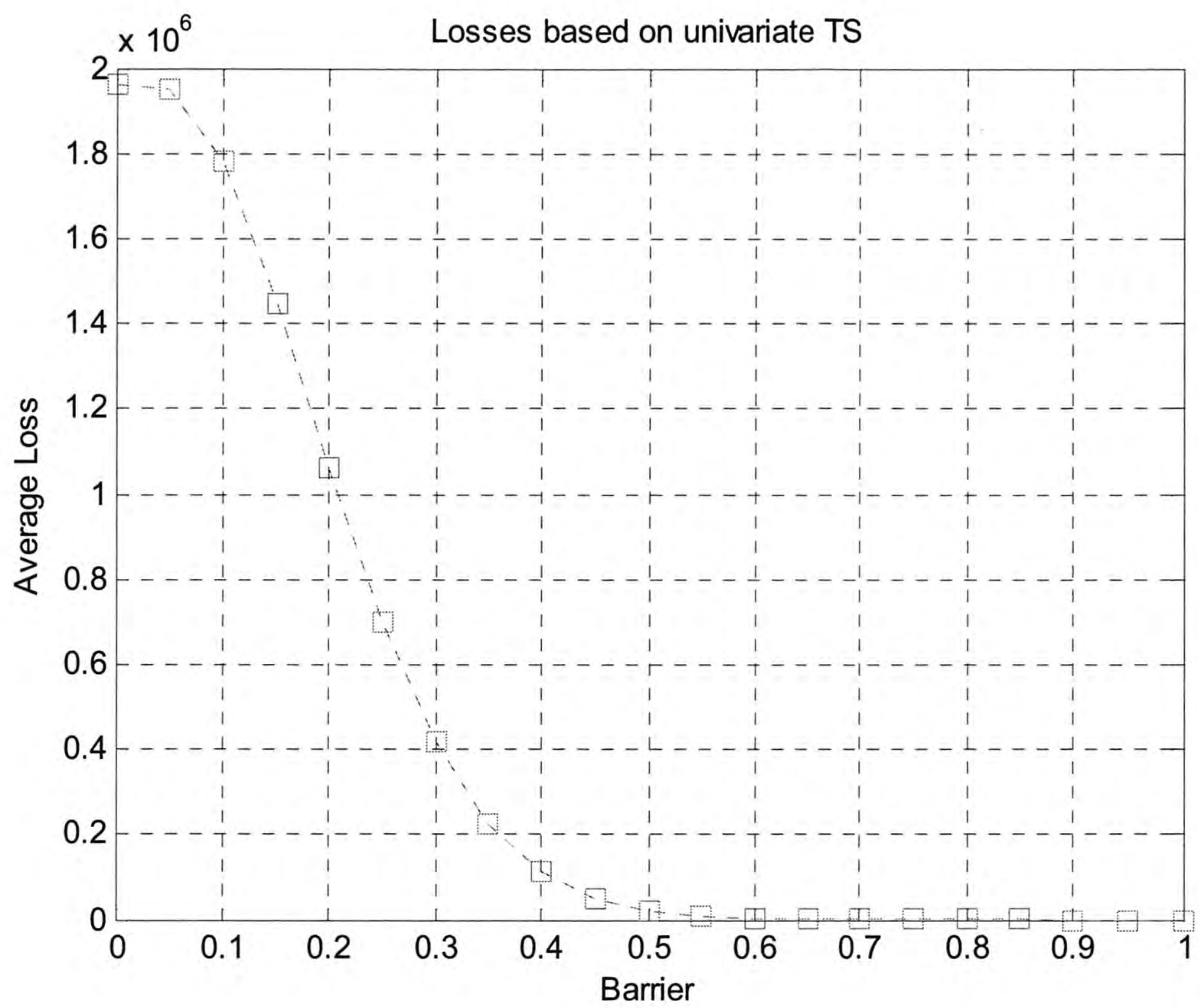


Figure F-3 "Expected losses" for HMAC

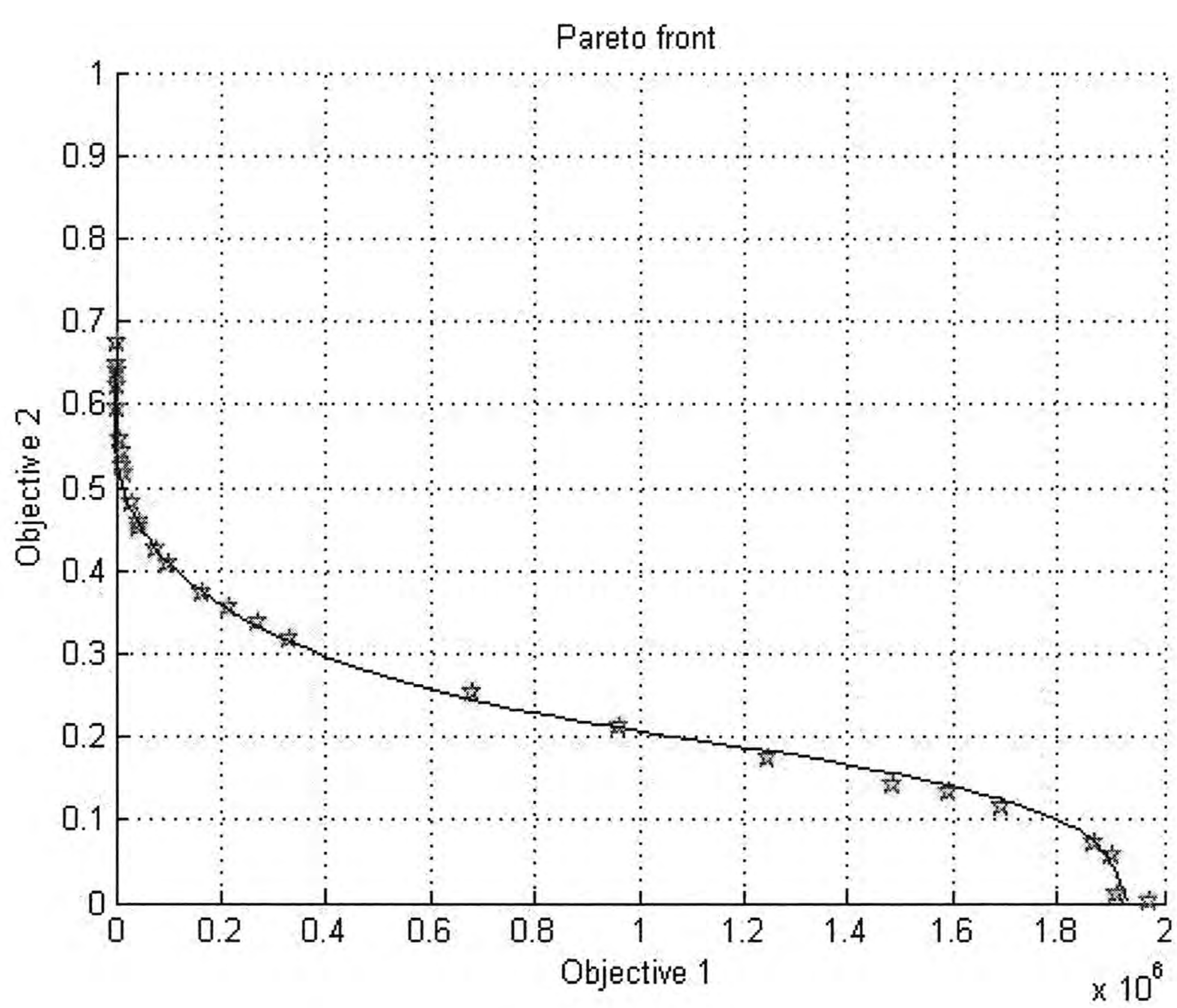


Figure F-4 Pareto front for HMAC

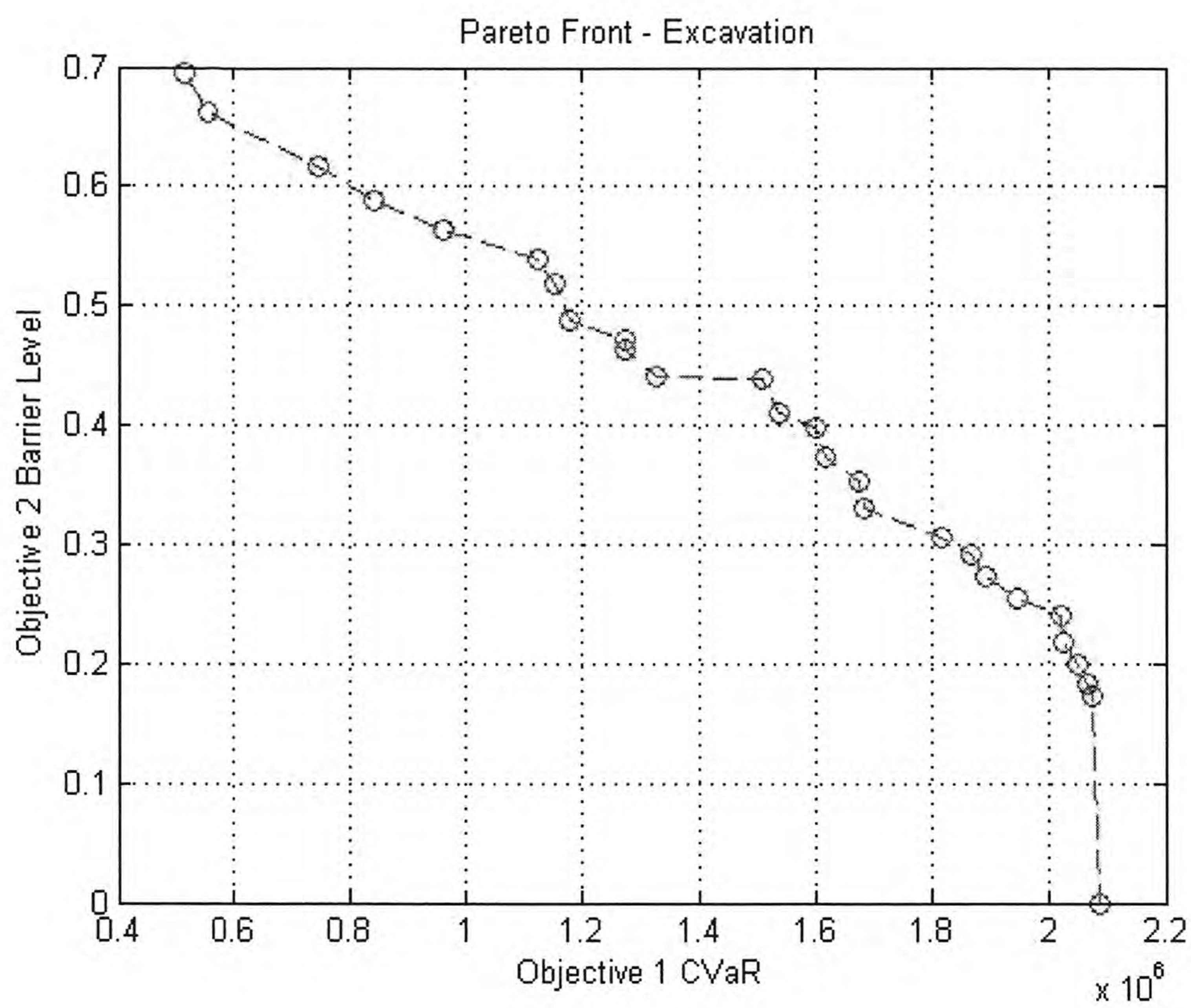


Figure F-5 Pareto front for excavation (associated with CVaR)

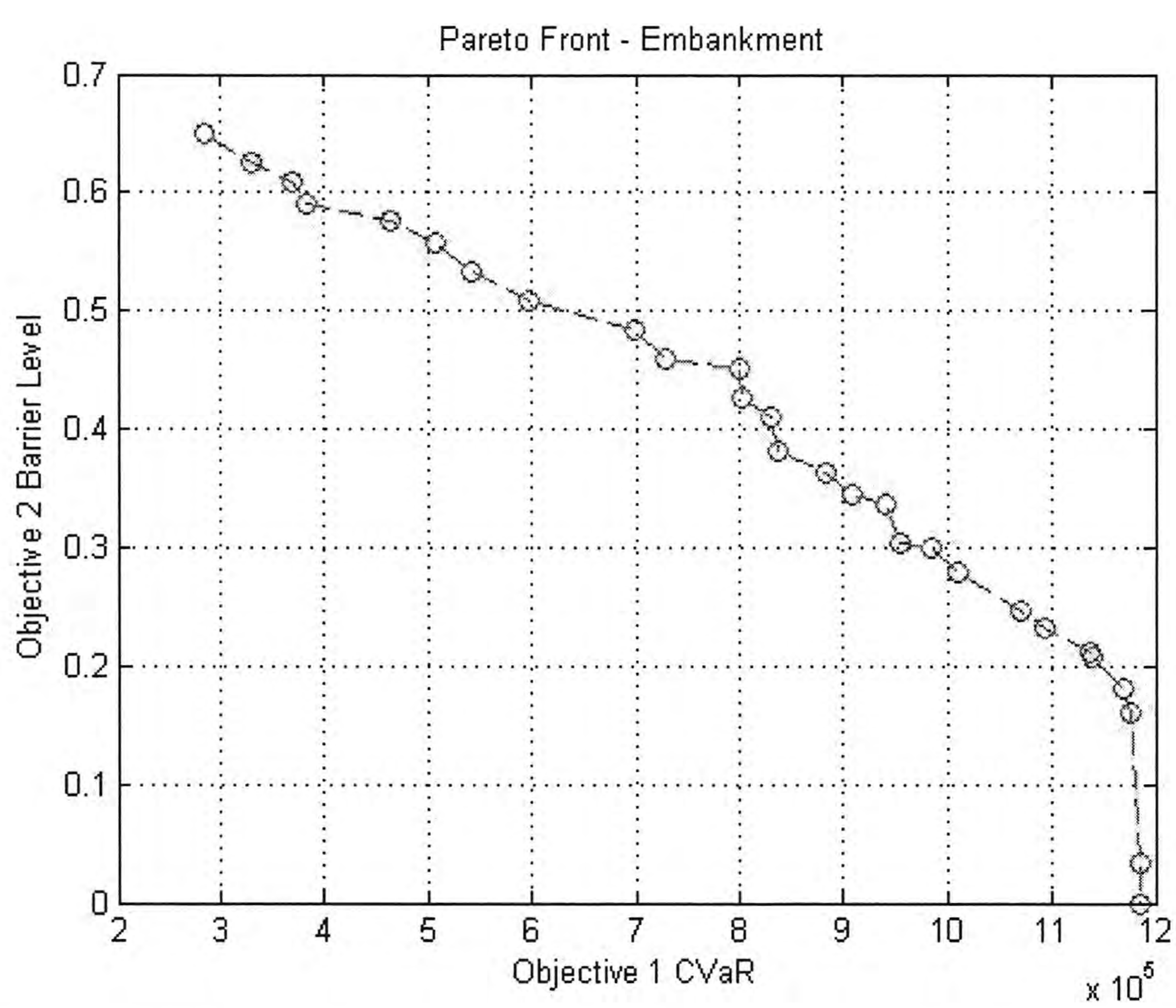


Figure F-6 Pareto front for embankment (associated with CVaR)

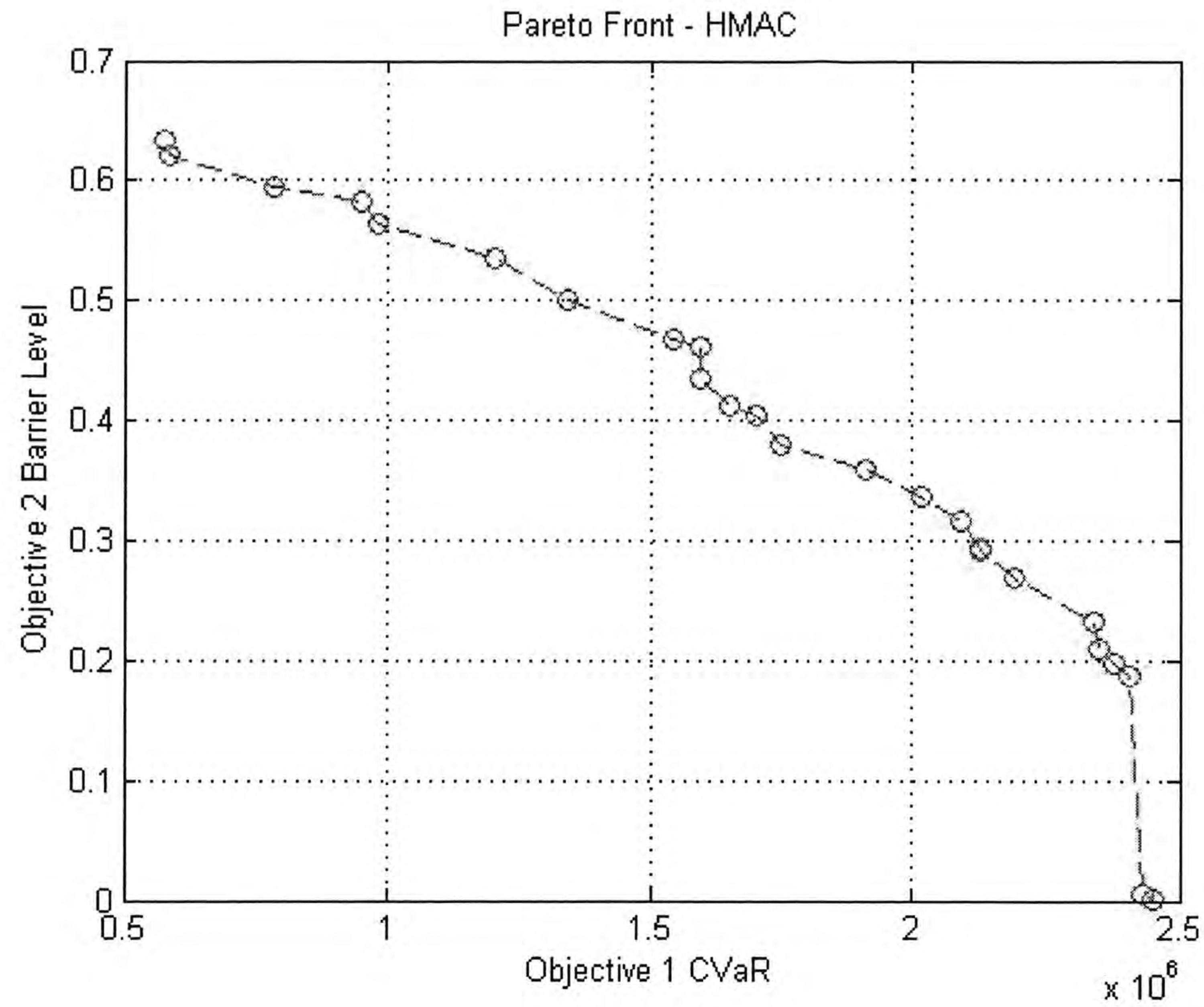


Figure F-7 Pareto front for HMAC (associated with CVaR)

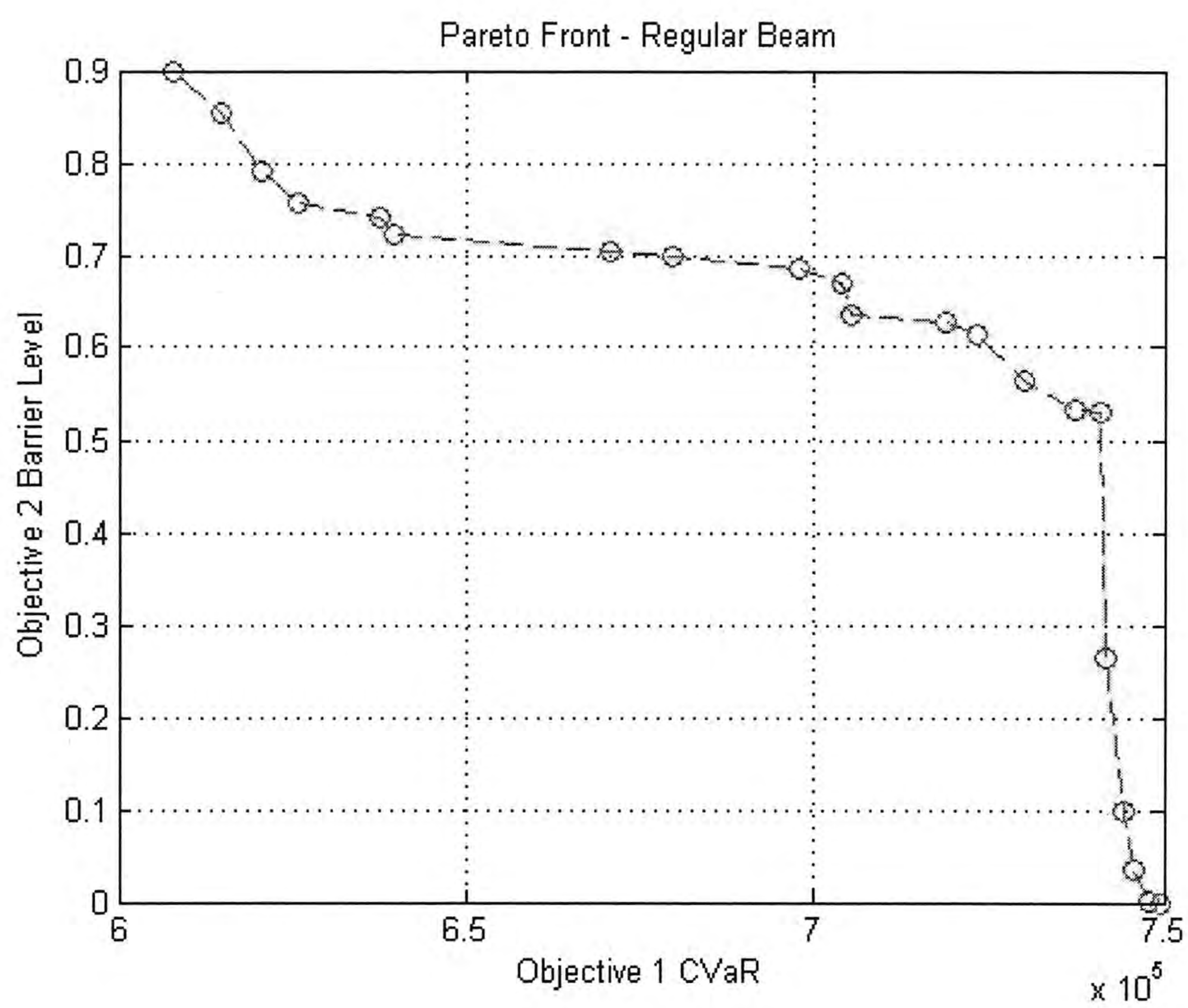


Figure F-8 Pareto front for regular beam (associated with CVaR)

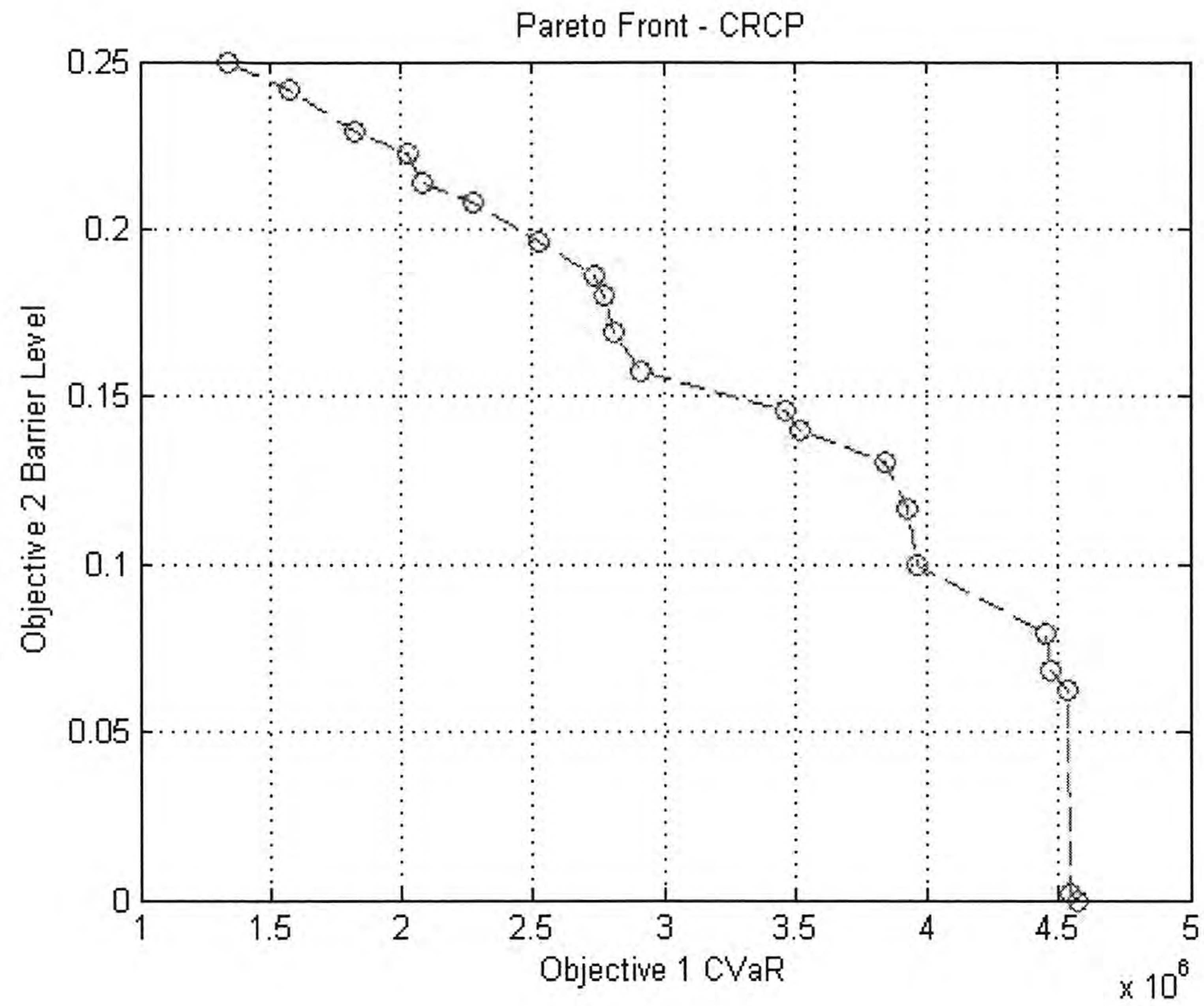


Figure F-9 Pareto front for CRCP (associated with CVaR)

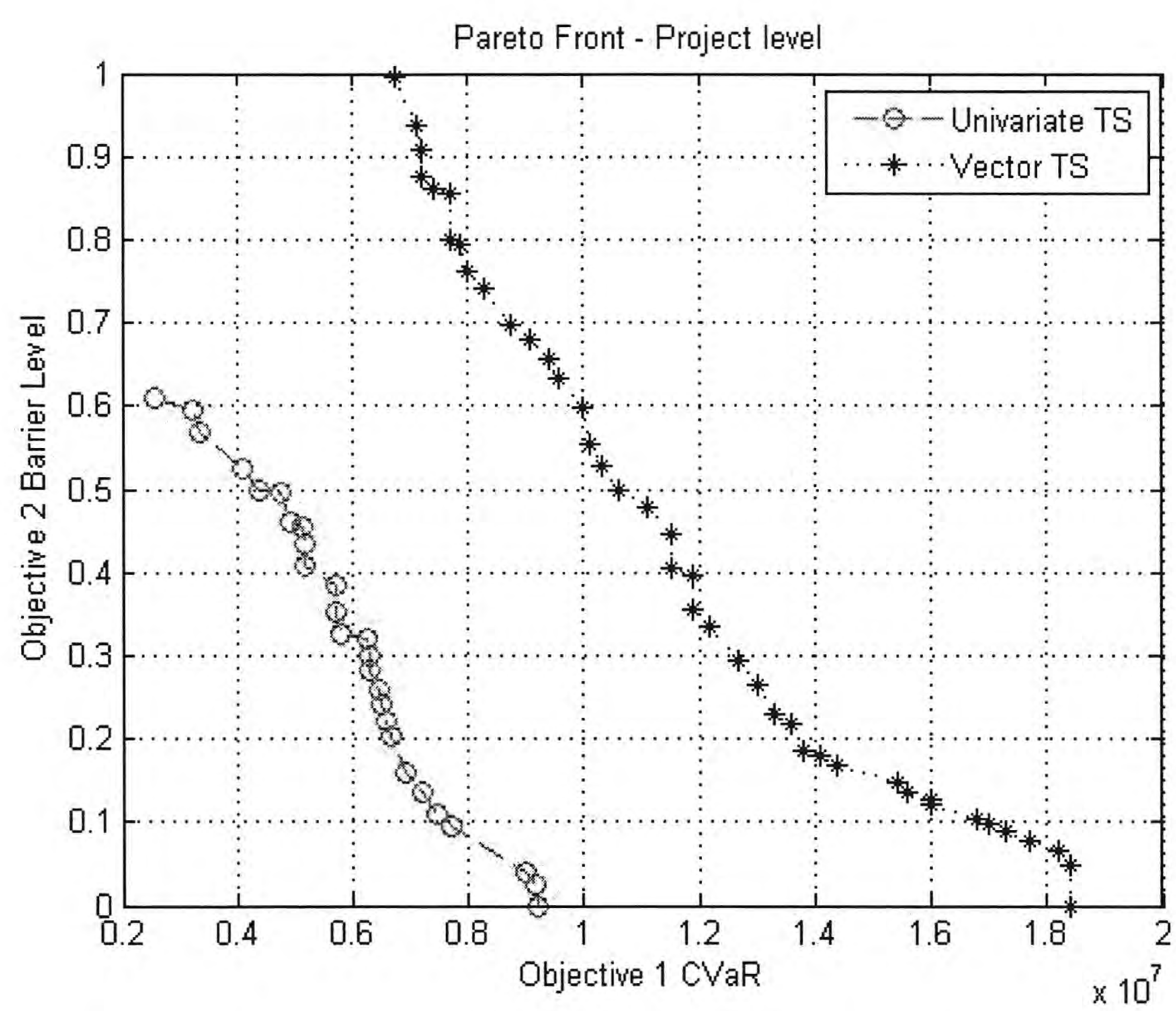


Figure F-10 Pareto front on project level (associated with CVaR) - single barrier

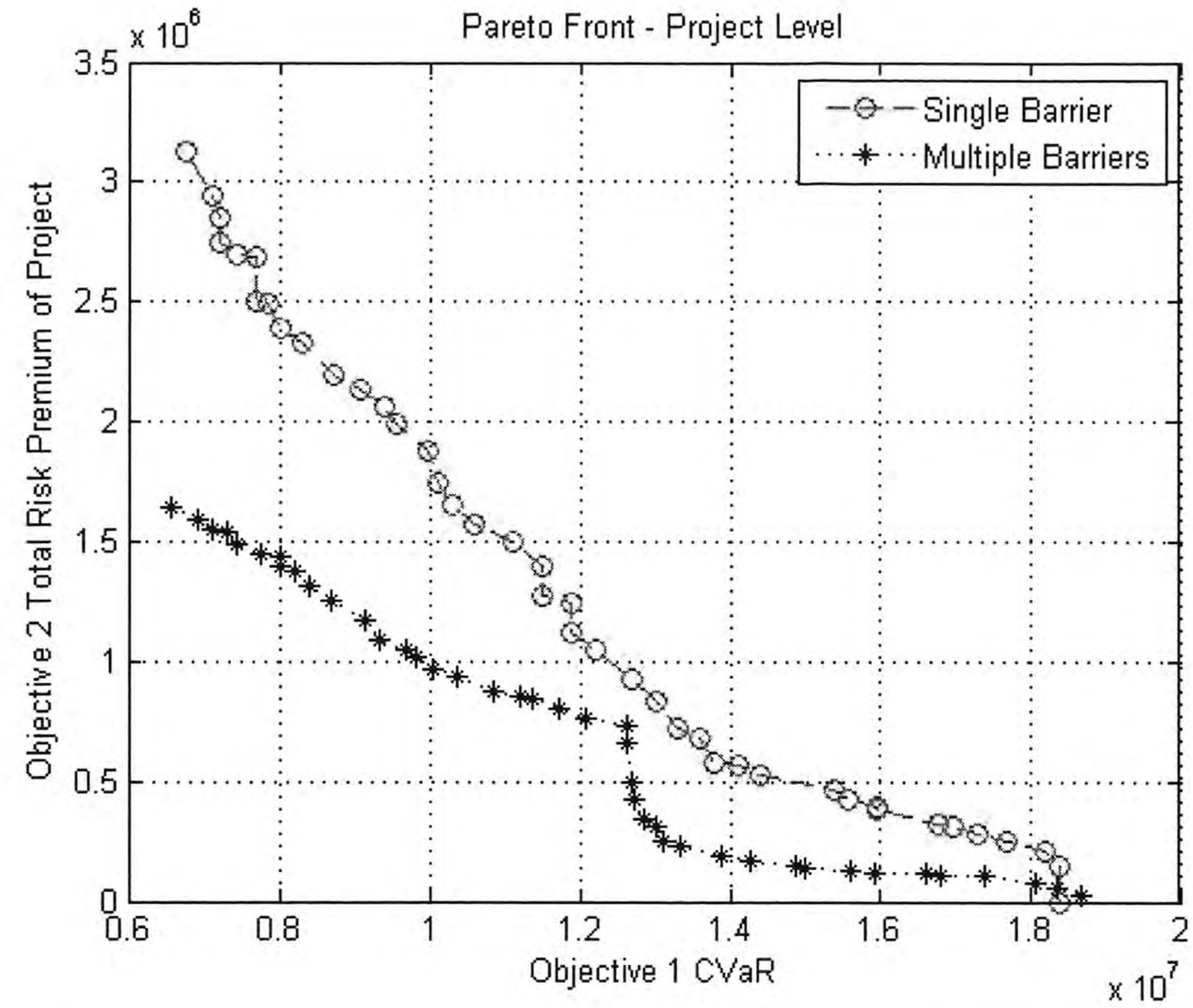


Figure F-11 Pareto front on project level (associated with CVaR) - multiple barrier