PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

ROAD MAINTENANCE UNIT COSTS ESTIMATION THROUGH AN EXPONENTIAL ROBUST REGRESSION MODEL

CRISTÓBAL IGNACIO MOENA MADRID

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Advisor:
ALFREDO SERPELL BLEY

Santiago, Chile, July 2012

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

ROAD MAINTENANCE UNIT COSTS ESTIMATION THROUGH AN EXPONENTIAL ROBUST REGRESSION MODEL

CRISTÓBAL IGNACIO MOENA MADRID

Members of the Committee:

ALFREDO SERPELL BLEY

MARCELA ALONDRA CHAMORRO GINÉ

PRISCILA HIDALGO SANDOVAL

BONIFACIO FERNÁNDEZ LARRAÑAGA

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the degree of Master of Science in Engineering

Santiago, Chile, July 2012

A mi familia.

ACKNOWLEDGEMENTS

This research was possible through the collaboration of different people, from private sector and government institutions.

First I would like to remark the efforts carried out at *Pontificia Universidad Católica de Chile* to increase the research activity at all fields, particularly in the engineering sciences. In order to become a developed country, Chile needs to create more in house knowledge and be able to transfer it to its economy and industry. This research is part of these efforts, and I am very grateful for the opportunity I had to develop it.

I am also very grateful to the *Dirección de Vialidad* (Roads Directorate, dependent of the Ministry of Public Works of Chile) and *APSA Gestión de Infraestructura Vial*. Both supported, promoted and sponsored this research, providing information, resources and knowledge. Both were benefited from this research. Thanks to Priscila Hidalgo and Bárbara Rozas (*APSA*) with whom I worked during this period, and to Miguel Valdés (*Dirección de Vialidad*) who provided useful information.

At the same time I am grateful to people at the university. Thanks to my advisor, Professor Alfredo Serpell, who showed me opportunities for the research development and guided me during these months. Thanks to the other members of the research committee, Professor Bonifacio Fernández and Professor Alondra Chamorro, who provided very useful advice in refining this thesis. I am also very grateful to my research classmates, with whom I was able to share different ideas and mutually support each other. Thanks to secretaries Lidia Soto and Edith Garrido, without their help it would have been much harder. And in general I am very grateful to everyone at the Engineering and Construction Management Department.

Finally, thanks to my family, especially to my parents for their constant support and motivation.

I sincerely hope this research to be a real contribution to engineering sciences and to Chilean road planning.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Good highway preservation schemes allow reducing costs and extending service life of roads. There are several methods to plan these project maintenances, but all of them require input information about maintenance costs. These costs can be very different depending on the geographical zone and the contracted volume due to economies of scale.

The main objective of this research is to propose a structured methodology to estimate the unit prices of road maintenance activities by modeling the exponential decay nature due to economies of scale as contracted volume increases. Using a combination of an exponential and a robust regression, curves were adjusted and parameters for price estimation were generated.

In addition, a geographical division of available data was developed allowing decreasing price variability. Also, a cost contingency analysis was performed in order to provide cost ranges associated to specific contracted volumes. Finally a procedure to update the estimated prices according to macro-economic variables was presented.

The validation of the methodology was carried out through its use in the estimation of real unit prices of historic road maintenance projects in Chile. It was found that the proposed methodology successfully allows modeling the unit prices for most of the activities considered, and for the typical kind of data records available. It was also found that for most of the studied activities the variability of unit prices decreases as contracted volume increases, which justifies that the methodology uses a differentiated cost contingency analysis for each volume range. Finally, it was found that for most activities the average unit price significantly differs depending on the geographical location in Chile where they were performed. Factors for price adjusting between zones were calculated.

This procedure may be used by road planners as well as contractors looking for a more confident approach in unit price estimation before participating in a bid. Furthermore, this methodology is not limited to road maintenance, but also to any other field where economies of scale and exponential fitting is observed.

Keywords: cost estimation; road maintenance; unit prices; exponential robust regression; cost contingency.

RESUMEN

Un buen programa de conservación de carreteras permite reducir costos y extender la vida útil de los caminos. Existen variados métodos para planificar las mantenciones, pero todos ellos requieren información de entrada sobre los costos de las actividades de mantención asociadas. Estos costos pueden variar considerablemente dependiendo de la zona geográfica y el volumen de contrato debido a las economías de escala.

El principal objetivo de esta investigación es proponer una metodología estructurada para la estimación de los precios unitarios de las acciones de conservación de caminos y carreteras, modelando el decaimiento exponencial que presentan producto de las economías de escala. Las curvas y parámetros para la estimación de precios fueron generados por medio de la combinación de una regresión exponencial y una regresión robusta.

Adicionalmente se introdujo una división geográfica de datos que permite disminuir la variabilidad en los precios. Así mismo, se muestra cómo es posible realizar un análisis de contingencia de costos para proveer rangos de precios asociados a volúmenes de contrato específicos. Finalmente, se presenta un procedimiento para la actualización de precios por medio de variables macro-económicas.

La validación de la metodología fue realizada por medio de la estimación de precios unitarios reales de proyectos históricos de mantención de carreteras en Chile. Se determinó que, considerando la información disponible, la metodología propuesta permite modelar los precios unitarios para la mayoría de las actividades estudiadas. También se obtuvo que la variabilidad de los precios unitarios disminuye a medida que el volumen de contrato aumenta, lo cual justifica utilizar un análisis de contingencia diferenciado para cada rango de volumen. Finalmente, se verificó que el precio unitario promedio varía significativamente de acuerdo a la zona geográfica de Chile considerada. Entonces, factores de ajuste entre zonas fueron calculados.

El procedimiento puede ser utilizado tanto por planificadores de caminos como por contratistas que buscan una estimación más confiable antes de participar en una licitación. Además, esta metodología no está limitada a la mantención de caminos, sino que puede ser aplicada en cualquier otro ámbito en que las economías de escala y el ajuste exponencial sean requeridos.

Palabras claves: estimación de costos; mantención de carreteras; precios unitarios; regresión robusta exponencial; contingencia de costos.

1.- INTRODUCTION

1.1.- Background

1.1.1.- Economic analysis

Highways and roads are an important part of every country public asset, allowing land occupation and communication among communities. As well as other types of infrastructure, road assets require a proper planning, which involves feasibility studies, design, construction, operation, preservation and rehabilitation. An economic analysis allows highway agencies to identify, quantify, and value the economic benefits and costs of highway projects and programs over a multiyear timeframe, being able to generate a cost-effective design and construction (Geiger, 2003).

Road management starts from the premise that the road network is an asset which needs to be maintained and improved so as to secure the best performance and value-for-money and the maximum service life (Robinson et al., 1998). The management of works and activities undertaken as part of road management is best viewed in terms of four main functions. Planning: defining road standards which minimize cost; determining the budget required to support defined standards. Programming: determining the work program that can be undertaken within the budgetary period. Preparation: design of works; preparation and issue of contract or work instruction. Operations: undertaking tasks as part of works activity (Robinson et al., 1998).

Within the planning function, there are several methods to assess life cycle costs of roads and compare highway investments. While basic methods for life cycle costs analysis (LCCA) have been described extensively in the literature (AASHTO, 2007; Abelson & Flowerdew, 1975; Anderson et al., 2009; Gransberg & Scheepbouwer, 2010; Salem, 1999; Schexnayder et al, 2003; Villacres, 2005), present worth and cost-effectiveness methods have been used almost exclusively in the pavement field (Haas et al., 2006). LCCA uses economic principles to compare different investment strategies, and it has always been an important tool in supporting decisions on the most cost-effective structure or rehabilitation treatment.

LCCA is used appropriately only to select from among design alternatives that would yield the same level of performance or benefits to the project’s users during normal operations (Geiger, 2003). If benefits vary among the design alternatives, then the alternatives cannot be compared solely on the basis of cost. Rather, the analyst would need to employ benefit-cost analysis (BCA), which measures the monetary value of life-cycle benefits as well as costs. Accordingly, LCCA should be viewed as a distinct, cost-only subset of BCA. Even with these restrictions, however, LCCA has many useful applications (Geiger, 2003).

LCCA is also important in determining the affordability of a project, including both the initial construction costs and any future costs that may arise (Haas et al., 2006). However, most LCCA parameters are estimates based on past knowledge and anticipated trends. Therefore, there is an element of uncertainty with most LCCA parameters, which is not accounted in deterministic analysis (Whiteley et al., 2005). Incorporating these uncertainties in the analysis becomes more important with long analysis periods (Osman, 2005).

Cost estimate is considered one of the most important and critical phases of a construction project (Jrade & Alkass, 2007; Anderson et al., 2009). Studies report that inaccuracy in cost estimation ranges from 20.4 to 44.7% (Liu et al., 2010), and despite decades of efforts to reduce project cost overruns, large infrastructure projects still continues to be plagued by delays and large cost overruns (Liu et al., 2010).

However, in road and highway projects not only the initial construction costs are important, but also operating and maintenance costs during its life: it is important to identify, quantify, and value the economic benefits and costs of highway projects and programs over a multiyear timeframe (Geiger, 2003). Well planned road preservation schemes allow significantly reducing costs involved during road life and extending its lifetime. Low resources assignment and poor maintenance planning carry important road assets loses (Acevedo & Muñoz, 2010).

Traditionally, project alternatives are characterized by the initial investment followed by proper maintenance in subsequent years (Archondo-Callao, 2011). Good highway preservation policies reduce costs and extend the service life of roads. There are several methods to plan these project maintenances (Dakin et al., 2006), but all of them require input information about maintenance costs. These costs can be very different depending on the geographical zone and the contracted volume (economies of scale) (Tighe, 2001).

1.1.2.- Pavement preservation

There are differences about how pavement preservation terminology is interpreted among local and State transportation agencies in the U.S. (Geiger, 2005), and with more reason in the rest of the world. This can cause inconsistency relating to how the preservation programs are applied and their effectiveness measured (Geiger, 2005).

Pavement Preservation is the overall program employing a network level, long-term strategy that enhances pavement performance by using an integrated, cost-effective set of practices that extend pavement life, improve safety and meet vehicle expectations.

An effective pavement preservation program will address pavements while they are still in good condition and before serious damage. By applying a cost-effective treatment at the right time, the pavement is restored almost to its original condition. The cumulative effect of systematic, successive preservation treatments is to postpone costly rehabilitation and reconstruction. During the life of a pavement, the cumulative discount value of the series of pavement preservation treatments is substantially less than the discounted value of the more extensive, higher cost of reconstruction and generally more economical than the cost of major rehabilitation. Additionally, performing a series of successive pavement preservation treatments during the life of a pavement is less disruptive to uniform traffic flow than the long closures normally associated with reconstruction projects.

Figure 1-1 shows the different preservation terminology that is currently used in the U.S Federal Highway Administration, as described by Geiger (Geiger, 2005).

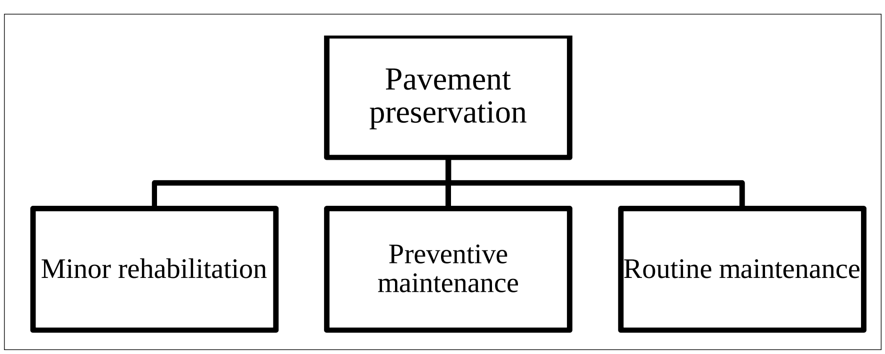

Figure 1-1 – Components of pavement preservation

Preventive Maintenance is a planned strategy of cost-effective treatments to an existing roadway system and its accessories that preserves the system, retards future deterioration, and maintains or improves the functional condition of the system (without significantly increasing the structural capacity).

Preventive maintenance is typically applied to pavements in good condition having significant remaining service life. As a major component of pavement preservation, preventive maintenance is a strategy of extending the service life by applying cost-effective treatments to the surface or near-surface of structurally sound pavements. Examples of preventive treatments include asphalt crack sealing, chip sealing, slurry or micro-surfacing, thin and ultra-thin hot-mix asphalt overlay, concrete joint sealing, diamond grinding, dowel-bar retrofit, and isolated, partial and/or fulldepth concrete repairs to restore functionality of the slab.

Pavement Rehabilitation consists of structural enhancements that extend the service life of an existing pavement and/or improve its load carrying capacity. Rehabilitation techniques include restoration treatments and structural overlays.

Rehabilitation projects extend the life of existing pavement structures either by restoring existing structural capacity through the elimination of age-related, environmental cracking of weakened pavement surface or by increasing pavement thickness to strengthen existing pavement sections to accommodate existing or projected traffic loading conditions. Two sub-categories exists: *minor rehabilitation* consists of non-structural enhancements made to the existing pavement sections to eliminate age-related, top-down surface cracking that develop in flexible pavements due to environmental exposure (included in preservation), and *major rehabilitation* consists of structural enhancements that both extend the service life of an existing pavement and/or improve its load-carrying capability.

Routine Maintenance consists of work that is planned and performed on a routine basis to maintain and preserve the condition of the highway system or to respond to specific conditions and events that restore the highway system to an adequate level of service.

Routine maintenance consists of day-to-day activities that are scheduled by maintenance personnel to maintain and preserve the condition of the highway system at a satisfactory level of service. Examples of pavement-related routine maintenance activities include cleaning of roadside ditches and structures, maintenance of pavement markings and crack filling, pothole patching and isolated overlays. Crack filling is another routine maintenance activity which consists of placing a generally, bituminous material into “non-working” cracks to substantially reduce water infiltration and reinforce adjacent top-down cracks.

Typical road maintenance activities include operations on the road strip, soil transportation, drainage, asphaltic pavements, concrete pavements, gravel and natural soils, bridges and structures, security, and snow control (Lorenzen, 2001). This research considers activities referred to pavement preservation (pavement rehabilitation, preventive maintenance, and routine maintenance) for asphalt, concrete and natural or gravel roads, and eventually some of major rehabilitation.

1.1.3.- Pavement performance

Road condition or state is generally characterized by 5 categories (Schliessler & Bull, 1994): very good (new roads), good (flawless roads requiring only routine maintenance), regular (roads with flaws and reduced structural capacity), bad (very reduced structural capacity, requiring immediate rehabilitation), very bad (needs reconstruction prior demolition).

The state of the road depends on many different factors, but mainly in its traffic, climate, rain, superficial waters, solar radiation and changes in temperature (Schliessler & Bull, 1994). The normal road cycle consists of four phases starting with construction and followed by a slow and hidden damage. After some years of use, road starts a faster deterioration diminishing its supporting capacity, which ends in a total decomposition if no preservation is performed (Schliessler & Bull, 1994).

Figure 1-2 shows an example of a normal road cycle in combination with agency costs needed for reinforcement, and shows the optimal maintenance policy. It is observed that there is a critical phase for reinforcement where road condition starts to deteriorate faster. This moment coincides with a faster increase in user’s vehicle operating costs generated by a deficient road.

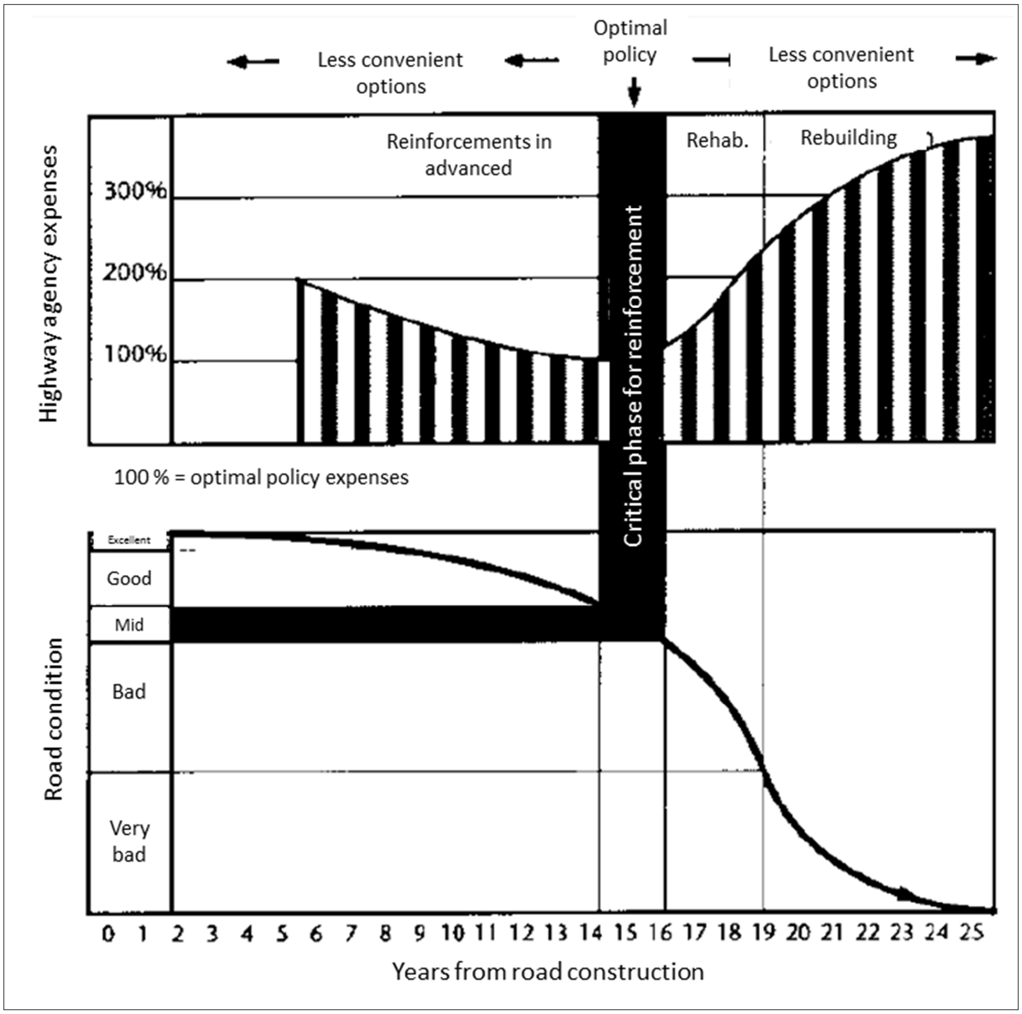

Figure 1-2 – Road deterioration cycle and reinforcement costs
Source: Schliessler & Bull, 1994

Figure 1-2 shows the importance of determining the appropriate moment for road intervention and remarks the need of creating a road preservation program for existing roads, by preserving at reasonable cost, keeping road security at acceptable levels, optimizing system cost-benefits, and reducing environmental adverse effects (Schliessler & Bull, 1994).

Experts urge to include different views in road maintenance and management (Lin & Lin, 2010; Pan et al, 2011), mainly in financing and in organizational aspects (Schliessler & Bull, 1994). For example, it has been proposed to use similar methods from private companies, such as a periodical company asset assessment. Also, when comparing dissimilar alternative maintenance strategies over a finite evaluation period, it is important to take account of the condition of the road pavement at the end of the analysis period because this may vary depending on whether a treatment would be made just within or beyond the analyzed period (Heggie, 1994). Others go even further, suggesting that recycling is a much better alternative than only focusing in maintenance. Transportation department of Florida (USA) compared traditional resurface against asphalt recycling: 35% annual savings were obtained by recycling (Hesham, 2011). This means that if $300 million dollars are invested in road resurfacing, $100 million dollars would have been saved by recycling.

1.1.4.- Unit cost estimation

There is no single approach to developing construction unit costs. Typically agencies have developed their own process for preparing their project estimates made to suit their requirements. As a result, highway construction projects employ a number of estimating procedures (Anderson et al, 2009).

The most common estimating technique reported is the historical bid-based estimation. Historical bid-based estimation is a method of developing estimates using data from the unit cost database (AASHTO, 2007). The unit cost database is a repository of the costs associated with all standard items of work taken from the previously awarded contracts or bids. This database stores information in a suitable format to aid the estimator in preparing cost estimates for highway projects. The unit price from this database is adjusted to reflect the specific project geographic and market conditions (Anderson et al, 2009).

There are some common factors that estimators need to consider when determining the unit prices (Anderson et al., 2009): project location, project size, quantity of materials, time of year, current market conditions, constructability, price-volatile materials, sequence of construction, contractor's familiarity of process, risks to contractors, and inflation.

Economies of scale are related to quantity of materials and project size, and it has been suggested that unit prices of road maintenance activities present an exponential decay due to economies of scale (Tighe, 2001). There are other examples where exponential decay needs to be modeled. Hernandez-Sancho et al. (2010) proposed a methodology for a better understanding of the cost structure of wastewater treatment processes, where a logarithmic transformation of curves was used. They concluded that the water volume plays a very important role in the determination of the operating and maintenance costs, i.e. these costs are affected by the economies of scale (Hernandez-Sancho et al., 2010). Similarly, by using an exponential approach, Maurer et al. (2010) introduces a generic model that represents the combined sewer infrastructure of a settlement quantitatively. The simulation results confirm that an economy of scale exists for combined sewer systems, and that it can be modeled with a logarithmic transformation.

At the same time, availability of historical unit cost data is an important factor in developing accurate project cost estimates. It is well-known that the ordinary least squares (OLS) regression estimators are highly sensitive to outlier presence in the datasets, and as a result robust regression methods have been proposed as alternatives (Arslan, 2011). Bai et al. (2012) used a robust method which utilizes a bisquare function, and demonstrated that it works well and it is much more efficient than OLS when there are outliers. In addition, when there is no outlier data the procedure works equally fine than OLS (Bai et al., 2012).

1.1.5.- Chilean background

During last decades in Chile more than 1500 km of asphalt roads were paved, which means that production jumped from 80,000 asphalt tons in 1990s, to 210,000 asphalt in 2000s (Orellana, 2006). By 2009 Chilean road network had 78,424.88 kilometers, and it was mainly composed by asphalt paved roads, gravel roads and natural roads. Road composition is shown in figure 1-3 (Acevedo & Muñoz, 2010).

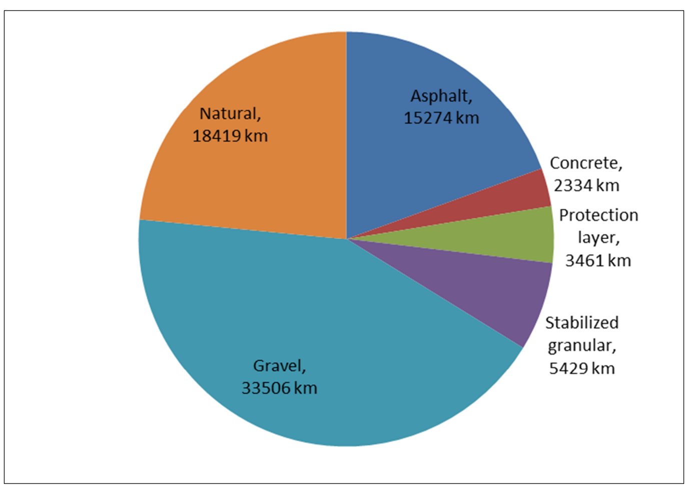

Figure 1-3 – Type of road distribution in Chile (in km).
Source: *Dirección de Vialidad, 2010.*

The national road asset is defined as the sum of every road value, considering current deterioration state (Acevedo & Muñoz, 2010). The value of a non-new road is calculated from the value it would have if it was brand new minus costs involved in repairing it and leaving it as it was new. Subtracted value corresponds to resources needed to be invested to fix any road failure. By 2009 Chilean road asset value was US$29,472 million (Acevedo & Muñoz, 2010). This includes both paved and unpaved roads, highway concessions, bridges and tunnels. This result is below the recommended value for Chile (Schliessler & Bull, 1994), which means that additional user costs are being generated due to operation and travelling time, and will require also greater maintenance investments in the future.

In order to achieve higher goals in terms of paved kilometers, it is necessary to introduce the "maintenance concept" (Orellana, 2006). The current paradigm consists of "*construct* → *zero maintenance* → *wait for degradation* → *reconstruct at higher costs*"; a new paradigm should include "*construct* → *preservation/rehabilitation at low costs* → *recovering road assets and recycling*". If only 5% of Chilean road assets were destined to an efficient maintenance program, which means being able to keep serviceability and functional road characteristics at high level, it would produce savings for about US$109 million dollars a year (Orellana, 2006).

For benefit of the Chilean national economy and its people, it is important increasing road budget and keeping maintenance at an adequate level. It is recommended to develop a road management system in charge of optimizing assignment of resources based on technical and economic principles that maximize global welfare (Acevedo & Muñoz, 2010). This important challenge is responsibility of the Ministry of Public Works, in cooperation with many other entities.

1.2.- Problem definition

Within an economic analysis there are many factors involved in the technical evaluation of roads life and preservation: initial construction costs, initial pavement performance, discount rate considered, cost of maintenance activities, and performance of maintenance activities. These factors have their own variability and produce variability in the output, as shown in figure 1-4.

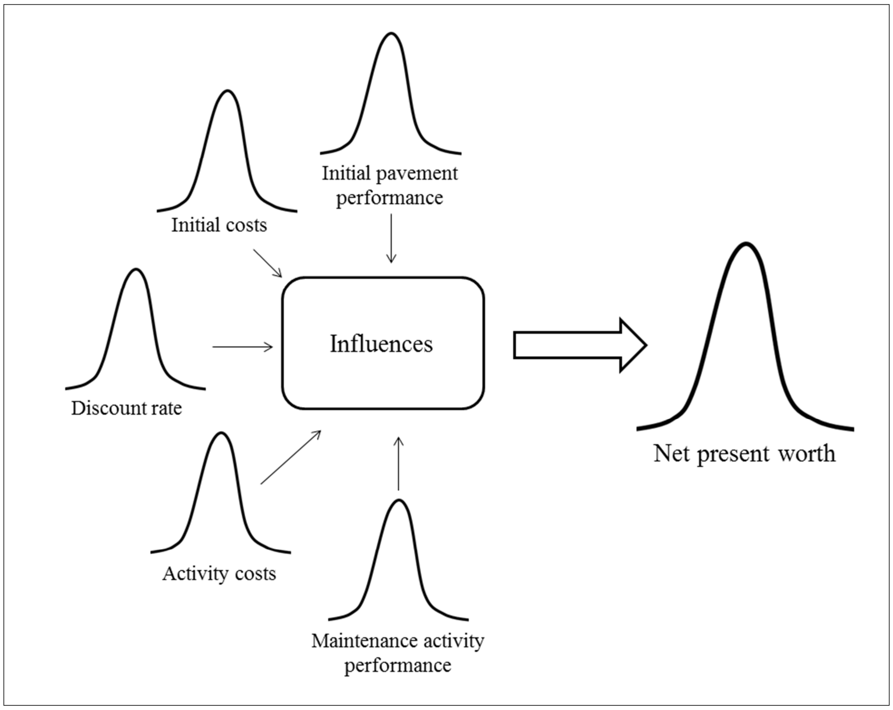

Figure 1-4 – Factors producing variability in road assessment

The road management level which this research is part of involves the estimation of initial costs and maintenance activity costs. Nevertheless, pavement performance and discount rates are other essential aspects for road assessment which are not accounted in this research.

According to the Federal Highway Administration from USA, state highway agencies have different approaches for developing unit costs. Though some state agencies have a systematic process in place for preparing project estimates, they have no written documentation on the entire process, including documentation for developing unit costs. Agencies do not have a formal process or method to adjust unit prices for project characteristics, current market condition, or current day prices. Adjustments to unit costs are based mainly on experience and engineering judgment of the estimators in all the state agencies (Anderson et al., 2009).

If such a systematic approach is not defined, estimators will spend considerable time searching databases for unit costs (Anderson et al., 2009).

Similarly in Chile there is not a group of unit prices to be used in road projects assessment, neither any formal procedure for estimating them. However, costs associated to road maintenance activities are essential to perform economic-technical evaluations (Valdés &

Guerra, 2011). Those prices definitely influence the final project portfolio that maximizes social benefits. Chile is currently progressing in creating an effective road maintenance program. In order to carry out these programs, unit prices of road maintenance activities are required for every maintenance program intended to be applied.

How can road maintenance unit prices be estimated? What variables affect unit prices estimation? Does the geographical region affect road maintenance unit prices? Is there a way to update those prices in time? How can economies of scale be modeled? These are some of the questions this research pretends to answer.

In this research a structured methodology to estimate unit prices of road maintenance activities is proposed, modeling the exponential decay nature of economies of scale. A geographical division of data records is presented allowing decreasing price variability. Using an exponential robust fit regression procedure curves are adjusted and parameters are generated. A cost contingency analysis may be performed in order to provide cost ranges associated to specific contracted volumes. Finally a procedure to update prices according to macro-economic variables is presented.

The methodology is programmed and applied in MATLAB language. A group of parameters result from this stage, and they are combined in a Microsoft Excel sheet to estimate unit prices.

The methodology validation was carried out through its use in the estimation of real unit prices of historic road maintenance projects in Chile. This procedure may be used by road planners within an economic analysis as well as contractors looking for a more confident approach before participating in a bid.

Furthermore, this methodology is not limited to road maintenance, but also to any other field where economies of scale and exponential fitting shall be needed.

1.3.- Research

1.3.1.- Hypothesis

This research tries to answer the following main questions. How can road maintenance unit prices be estimated? What variables should be included in a model?

The stated hypothesis proposes that road maintenance unit prices can be estimated using a mathematical methodology based on exponential robust fitting, considering the geographical zone and the contracted volume of each activity.

In addition it is stated that this methodology allows performing a cost contingency analysis in price estimation, and price updating from macro-economic variables of inflation, dollar value, and crude oil price.

1.3.2.- Objectives

The main objective of this research is to develop a methodology to estimate unit prices of road maintenance activities. This procedure must be able to be reproduced in any place and time if correct information and data is provided.

Specific objectives are to:

- Develop a procedure to model economies of scale in unit prices, including the possible presence of outliers in data records.
- Determine influence of geographical region in unit price variability.
- Verify random behavior of unit prices.
- Develop a procedure to analyze cost contingency in price estimation.
- Develop a procedure to update prices from macro-economic variables.
- Generate model parameters for selected road maintenance activities.
- Determine the incidence of variables in the estimation of unit prices.
- Generate a prototype example for the implementation of the proposed model.

1.3.3.- Research methodology

The research methodology consists of basically three stages:

- Literature review and object of study: prior to defining the specific object of study (unit price estimation of road maintenance activities) a literature review about road preservation, cost estimating, cost contingency, statistical analysis and other subjects needs to be performed. Worldwide literature as well as Chilean information is considered. This stage is not time-limited and accompanies during all research.
- Model definition and data gathering: consists of the mathematical definition of the model intended to be applied for the estimation of unit prices. This needs to be done

considering the historical data gathering and the type of information available to develop the model.

- Model application and analysis of results: model is applied, methodology is validated, and results are obtained. Then results are analyzed and conclusions are drawn. This may be an iterative stage where after analyzing results, some adjustments to the model may be required.

The scope of this research is to show that the proposed methodology is a good tool for the price estimation of road maintenance activities, provided that good data records are available. With the required input information, this procedure may be easily used to create price estimation models.

The aim of this research is not determining current road maintenance activities parameters in Chile, but to demonstrate that the procedure is robust enough to estimate unit prices.

1.4.- Thesis structure

This thesis is structured into five chapters. The first chapter presents an overall look at the thesis background, problem definition, hypothesis, objectives, methodology and main results to be obtained.

Chapter two includes an explanation of the experimental design of this research with its questions, road maintenance activities considered, type of data required, databases description, independent and dependent variables, and the factorial design.

Chapter three provides a description and analysis of variables in which this methodology relies on. It includes economies of scale modeling through exponential robust regression, geographical classification of data, cost contingency analysis, and price updating explanation.

Chapter four consists of the implementation of the methodology. It explains the validation procedure and provides some case studies with detailed results.

Chapter five contains the main conclusions, recommendations and suggestions for future research.

2.- EXPERIMENTAL DESIGN

2.1.- Introduction

In this chapter, an explanation of the experimental design of this research with its questions, road maintenance activities considered, type of data required, databases description, independent and dependent variables, and the factorial design is presented.

2.2.- Methodology description

There are several variables affecting and producing variability to the unit price of specific road maintenance activities: type of maintenance, geographical zone of maintenance, market conditions, contractor risks, constructability, contract negotiation ability, and contracted volume are some of the main factors that can affect the value of unit prices. This research focuses mainly in contracted volumes due to its proven influence in prices (Tighe, 2001; Valdés & Guerra, 2011) and its ease of quantification, and incorporates a cost contingency analysis to deal with other variables that are more difficult to measure. Neglecting the importance of cost contingency analysis may cause project costs overrun (Lawrence, 2007).

In order to estimate unit prices, this research proposes a methodology that consists of a combination of an exponential regression and a robust regression model. Exponential regression is used when input data (unit prices) tends to exponentially decay as the independent variable (contracted volume) increases. Robust regression is used when input data presents outliers that would distort the results: big databases inevitably contain mistakes that can affect regression performance. Then, with a combination of both procedures this problem can be addressed.

Historical input data is then processed in MATLAB, which generates parameters for each road maintenance activity studied. These parameters are handled in an Excel sheet, where the unit price estimation and cost contingency analysis is performed. In this stage, price updating is included in order to include the effects produced by changes in macro-economic variables such us inflation, dollar price, and crude oil price, since data was gathered. This is performed by finding relationships between these macro variables and different supply price indexes (asphalt, cement, diesel, gravel, and steel price indexes). Then, prices are updated using these indexes and considering the internal composition of the unit prices.

Figure 2-1 shows a flowchart of the proposed methodology for unit price estimation.

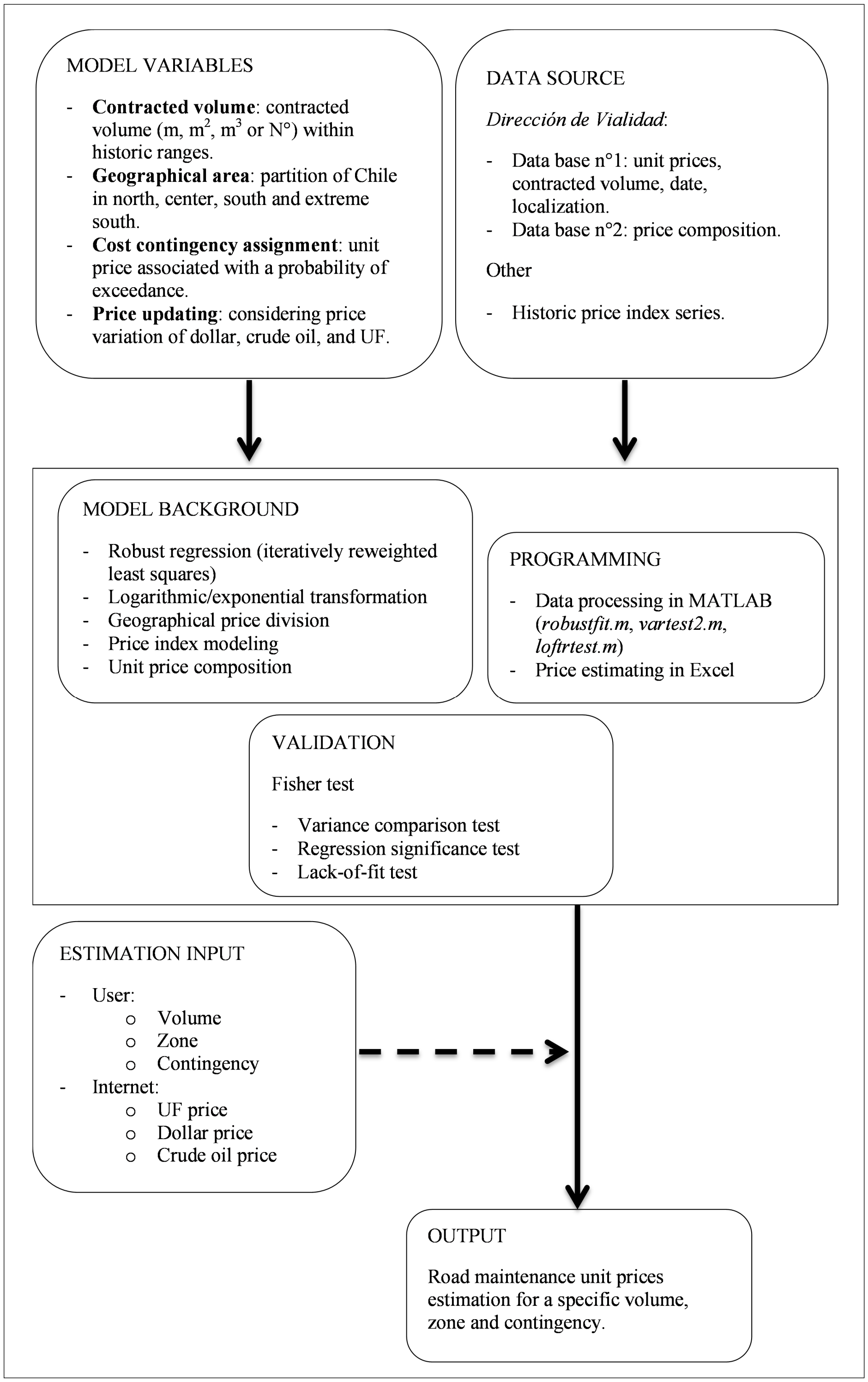

Figure 2-1 – Methodology flowchart

Hence, in order to evaluate the proposed methodology for unit price estimation, this research uses a non-experimental design in multiple variables.

2.3.- Research questions

As mentioned earlier, there are some common factors that estimators need to consider when determining the unit prices. This research focuses on mainly on the factor of quantity of materials (economies of scale), and also addresses variability due to project location and current market conditions.

The following main questions arise: How can the unit prices of road maintenance activities be estimated? What variables do affect the estimation of unit prices? Does the geographical region affect road maintenance unit prices? Is there a way to update those prices in time? How can economies of scale be modeled?

2.4.- Experiment objectives

The main objective of these experiments is to determine whether exponential curves can be adjusted to model the price variation due to economies of scale using a robust regression, and if these curves present goodness of fit.

Secondary objectives of these experiments are to determine if there are significant differences in unit prices among the different zones of the geographical division of data presented and the different contracted volume ranges defined.

2.5.- Population description

In this research, the population corresponds to the unit prices of road maintenance activities that have already taken place in Chile during the last years. The experimental design samples correspond to a subset of the available data from activities related to asphalt, concrete, and natural and gravel roads. Hence, it is very important to describe the type of historical information available to develop the estimation model.

This section includes a brief explanation of road maintenance activities considered in this research, a description of databases used for the methodology testing, and the type of data required.

2.5.1.- Road maintenance activities

This section presents and describes road maintenance activities considered for the methodology testing. These activities were selected from *Manual de Carreteras Volumen 7* (Lorenzen et al., 2001).

In Chile, *Dirección de Vialidad* is the institution in charge of improving road connectivity to society by planning, building, and maintaining Chilean road assets[1]. It is dependent on Ministry of Public Works of Chile.

Manual de Carreteras is a normative document created by *Dirección de Vialidad* used as a guide to many different road actions. It establishes policies, criteria, procedures and methods that road projects should comply in relation with planning, assessment, design, construction, security, maintenance, quality and environmental impacts (Lorenzen et al., 2001). Volume 7 corresponds to road maintenances, where all road maintenance activities and their technical description are included.

In *Manual de Carreteras* road maintenance activities are divided into nine groups: road strip, soil transportation, drainage, asphaltic pavements, concrete pavements, gravel and natural soils, bridges and structures, security, and snow control. These activities are described in detail, and are similar but more specific to those proposed in HDM-4 software (Dakin et al., 2006).

For the methodology validation 68 pavement maintenance activities were considered due to their direct relation with pavement performance. They correspond to activities of asphalt road maintenance (codes starting with 7.304), concrete road maintenance (codes starting with 7.305), and gravel and natural roads maintenance (codes starting with 7.306). A list of these activities with their respective units of measurement is presented in Appendix A.

2.5.2.- Databases

This section describes data records used for methodology testing. Some statistics regarding data composition are also discussed.

The information consists of a compilation of awarded contracts by *Dirección de Vialidad* for road maintenance projects during the last decade (since 1998). Two databases are considered, and they are described next.

---
[1] http://www.vialidad.cl

2.5.2.1.- Database 1: bid awarding information

Principal database (D1) contains extensive historical information regarding bid awarding information, date of contract, people in charge, road information, contracted volume, unit prices, and type of work. A typical record in this database contains the following information:

- Region
- Contract
  - Internal code
  - Starting date
  - Ending date
  - Inspector
  - Contractor
  - Project total cost
- Road
  - Name
  - Road code
- Activity
  - Name
  - Activity code
  - Contracted volume
  - Unit price (before taxes, in Chilean pesos)

Data information is stored in an Excel sheet. It contains approximately 100,000 records, but, for purpose of this research, some of them provided null or incorrect data (e.g. negative values), and were dismissed from database. Moreover, many of the activity codes are misspelled or written in a slightly different way (e.g. 7-304-1 b instead of 7.304.1b). This produces that an important amount of information is useless, unless it is corrected.

The following charts describe the type of information contained in databases.

Figure 2-2 shows the geographical distribution of data. There is more information in central regions and it is probably due to higher maintenance frequencies in populated areas. Due to an unknown reason, there is a high amount of information in XIV region.

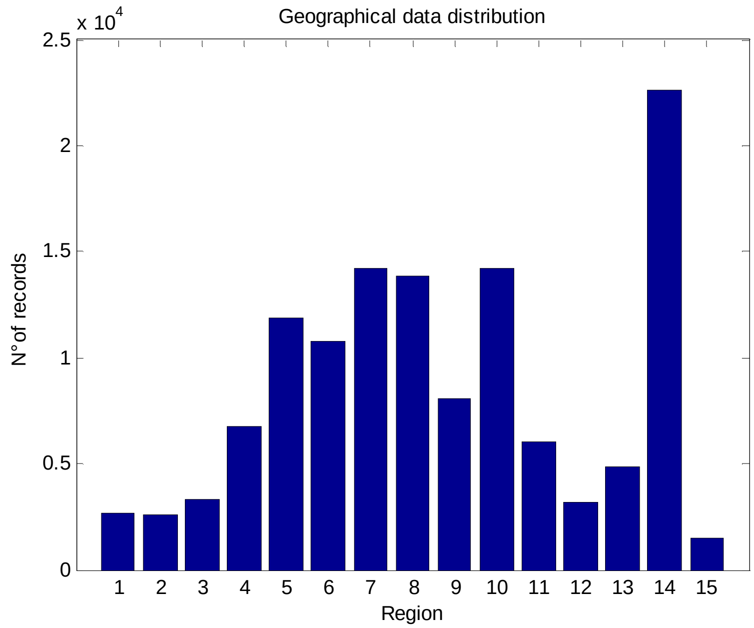

Figure 2-2 – Data distribution by region in database D1

Figure 2-3 presents the temporary distribution of data. Data starts in 1998. The highest concentration of records is produced between 2008 and 2011.

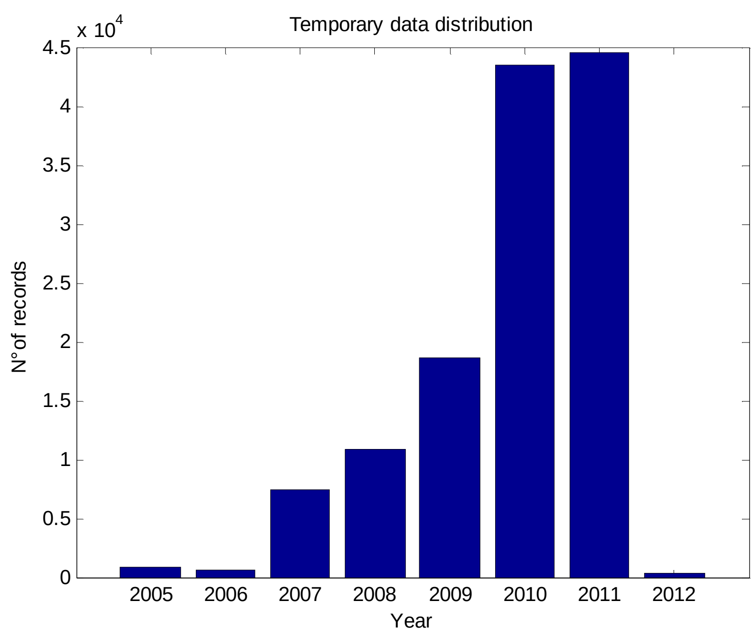

Figure 2-3 – Temporary distribution of data in database D1

Figure 2-4 shows data composition of road maintenance activities as described in *Manual de carreteras Volumen 7*. There is plenty of data from gravel and natural roads, and asphalt paved roads maintenance. On the other hand, there is little information about concrete paved road maintenances (mainly due to the lack of concrete roads in Chile), and, as it is verified in following chapters, this generates bad curve adjustments.

In order to have a common comparative basis, unit prices are read in MATLAB and converted into UF units (*unidad de fomento*, Chilean indexed monetary unit adjusted by inflation) considering the date of each record (Geiger, 2003). Dollars or pesos from one year can be converted into equivalent by using price indices to add or remove the effects of inflation. Dollars or pesos from which the inflation component has been removed are known as "real", and are able to buy the same amount of goods and services in a future year as in the base year of the analysis. Dollars or pesos that include the effects of inflation are known as "nominal" dollars. A nominal dollar will typically buy a different amount of goods and services in each year of the analysis period (Geiger, 2003).

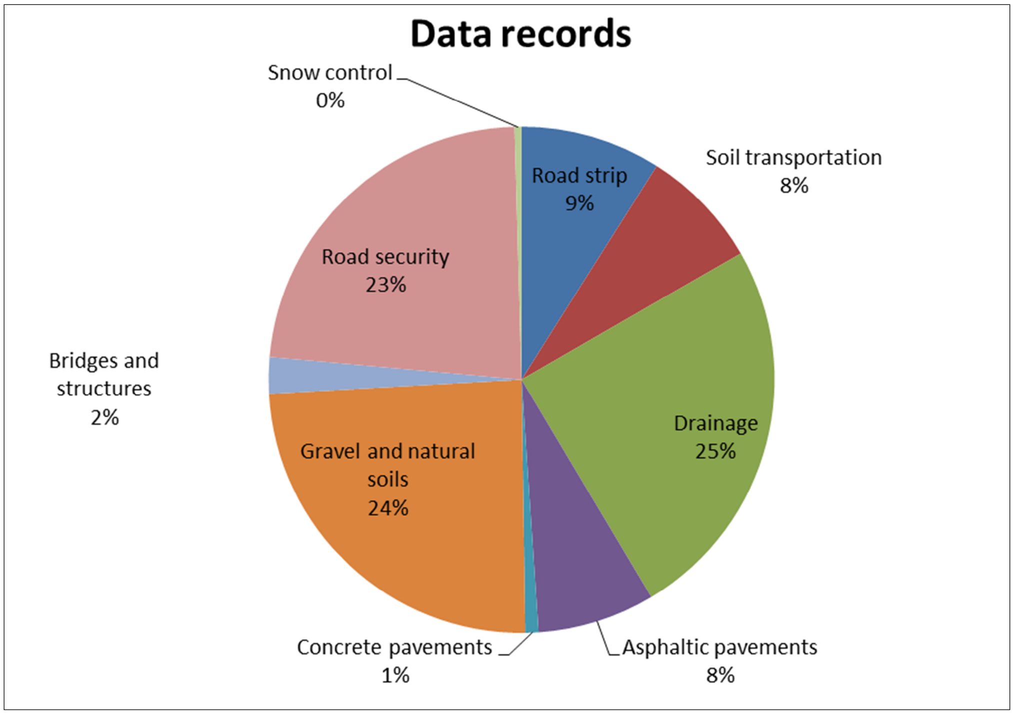

Figure 2-4 – Maintenance type composition of database D1

2.5.2.2.- Database 2: composition of unit prices

Second database (D2) contains fewer records than the first one, but provides useful information about internal composition of unit prices. A typical record in the database contains the following information:

- Region
- Contract
  - Date
- Activity
  - Name
  - Code
  - Contracted volume
  - Activity price (before taxes, in Chilean pesos)
    - Labor
    - Overhead
    - Equipment
    - Oil
    - Material

Data information is stored in an Excel sheet. Initially it had approximately 30,000 records, but, for purpose of this research, some of them provided null or incorrect data (e.g. negative values) that were deleted from database.

Figures 2-5 and 2-6 present two example cases of the cost composition from maintenance activities in *Manual de Carreteras*. These are illustrative examples and they were chosen randomly.

In figure 2-5 it is observed that 7.302.1b *Sand removal* makes heavy use of equipment. Moreover, cost component presenting the lowest variability also corresponds to equipment, which provides more accuracy in costs.

In figure 2-6 it is observed that 7.304.1a *Crack sealing up to 6 mm width* intensively uses labor, followed by material and overhead costs.

Using MATLAB, unit prices are computed dividing total cost by total contracted volume. In order to be comparable, unit prices are read in MATLAB and converted into UF units (*unidad de fomento*, Chilean indexed monetary unit adjusted by inflation) considering the date of each record (Geiger, 2003).

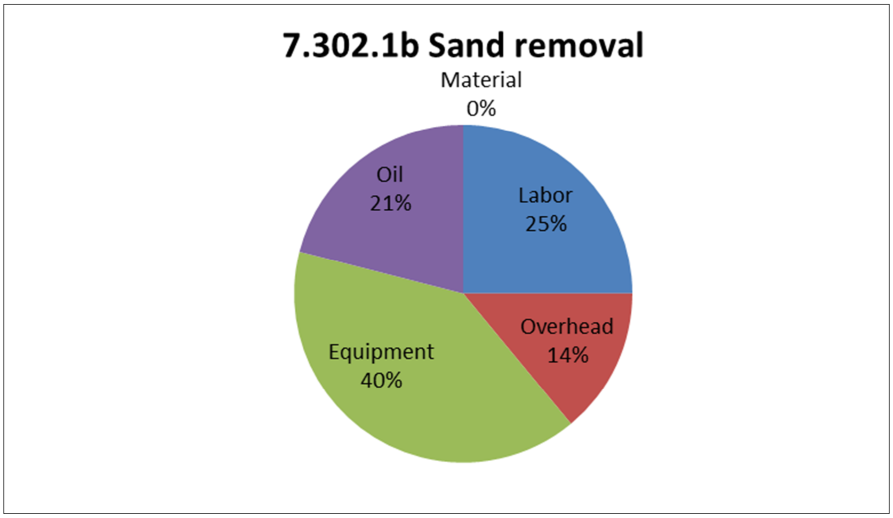

Figure 2-5 – Composition of unit price for 7.302.1b *Sand removal*

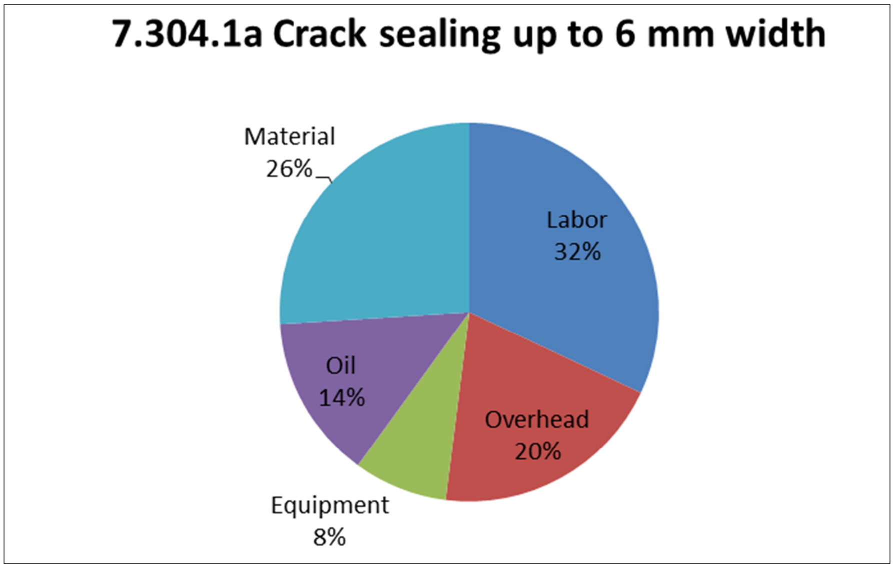

Figure 2-6 – Composition of unit price for 7.304.1a *Crack sealing of up to 6 mm width*

2.6.- Factorial design

The experimental sample units correspond to the unit price of each of the 68 road maintenance activities described earlier (activities 7.304, 7.305, and 7.306).

The primary variable corresponds to the contracted volume for each activity, generally represented by linear meters, squared meters, cubic meters, or quantity. This variable can be found in the databases presented, and also in every new project to be performed.

Background variables are the geographical location of the desired implementation of the activity, and current market conditions represented by UF price, dollar price, and crude oil price.

There are uncontrollable or hard-to-change variables that are difficult to measure and are included in the design. Some of them are constructability, sequence of construction, contractor’s familiarity of process, and risks to contractors. These variables do affect unit prices, but they are very hard to measure objectively, and there is no information within the available records. On the other hand, project size can be measured in different ways and is related to contracted volume, and thus it is not considered. All these uncontrollable variables produce variability in unit prices that may be addressed by a cost contingency analysis.

The treatment structure consists of the following levels of factors:

- Three levels of contracted volume: low, medium, and high, depending on historical data.
- Four different geographical zones: zone 1, zone 2, zone 3 and zone 4 (north, central, south, and extreme south of Chile respectively).
- Market fluctuations of macro-economic variables.
- Two cost contingency levels: high contingency (probability of exceedance of 15.9%), low contingency (probability of exceedance of 50.0%).

The factorial experimental design for one maintenance activity is shown in table 2-1.

Table 2-1 – Unit price estimation for one maintenance activity

ID	Contracted volume	Geographical zone	Cost contingency	UF price	Dollar price	Crude oil price	Unit price estimate
1	Low	Zone 1	Low				UP estimate
2	Low	Zone 1	High				UP estimate
3	Low	Zone 2	Low				UP estimate
4	Low	Zone 2	High				UP estimate
5	Low	Zone 3	Low				UP estimate
6	Low	Zone 3	High				UP estimate
7	Low	Zone 4	Low				UP estimate
8	Low	Zone 4	High				UP estimate
9	Medium	Zone 1	Low				UP estimate
10	Medium	Zone 1	High				UP estimate
11	Medium	Zone 2	Low				UP estimate
12	Medium	Zone 2	High		(Market value)		UP estimate
13	Medium	Zone 3	Low				UP estimate
14	Medium	Zone 3	High				UP estimate
15	Medium	Zone 4	Low				UP estimate
16	Medium	Zone 4	High				UP estimate
17	High	Zone 1	Low				UP estimate
18	High	Zone 1	High				UP estimate
19	High	Zone 2	Low				UP estimate
20	High	Zone 2	High				UP estimate
21	High	Zone 3	Low				UP estimate
22	High	Zone 3	High				UP estimate
23	High	Zone 4	Low				UP estimate
24	High	Zone 4	High				UP estimate

The unit prices of each activity are allocated randomly because the data selection used for the analysis considers only those prices that have correctly spelled their activity codes, and this misspelling is assumed to be produce randomly. At the same time, those unit prices consisting of negative values were not considered.

The data collection was performed by many people working for *Dirección de Vialidad* from 1998 to 2011. This suggests that no personal bias should have been produced when collecting the information, due to the multiplicity of different people involved. However, the drawback of this procedure is that data can be written in different formats, and more outliers can be easily found within the records.

2.7.- Inference space

The inference space of experimental results defines the range of conditions for which the subjects form a representative sample, and are assumed to be randomly selected from this well-defined population (SAS, 2005).

The final results are intended to be applied over the population of unit prices of road maintenance activities described earlier, considering the range of independent variables defined. This means the contracted volume within historical ranges of each activity; one of the geographical areas described; and market conditions similar to the period when the data was gathered.

Extrapolating results outside normal volume ranges may lead to results that cannot be evaluated by this methodology.

Considering other geographical area is not possible. However, a goal for future research should be testing this methodology with data from other regions of the world, assuming this data contains the same items that the data used for developing this methodology. For the Chilean case the specified zones are not necessarily homogenous. This implies that there may be price differences between specific areas in the defined zones, which are not considered in this research.

If market conditions considerably differ with conditions from years when the data was gathered, results may be inappropriate. Further details are explained in later chapters.

It is assumed that unit prices are the dependent variables influenced by independent variables that were accurately measured. It is also assumed that data collecting of unit prices was performed by different people using the same criteria for price determination, and that no repeated data is included in the records. Finally it is assumed that there is no large correlation between any of the independent variables.

2.8.- Summary

In this chapter, an explanation of the experimental design of this research was presented. The methodology for unit price estimation was described, research questions were presented, and experimental objectives were defined. A detailed description of the population used in this research was shown, including the activities considered for the study and a description of the databases used. The factorial design for the experiment was presented, including the ranges for the different independent variables considered. Finally, a description of the inference space of experimental results was provided.

3.- DATA ANALYSIS

3.1.- Introduction

In this chapter, a description and an analysis of the variables in which this methodology relies on are presented.

First an explanation of the economies of scale problem is shown, and how an exponential robust regression is able to handle it. Then it and proposes a geographical data division to reduce prices variability, and continues explaining the probabilistic nature of unit prices and how cost contingency allows modeling uncertainty. And finally, a price updating technique considering macro-economic variables is proposed.

3.2.- Economies of scale

3.2.1.- Modeling

This section describes the methodology used to model the economies of scale in road maintenance unit prices.

Ordinary least square method (OLS) is one of the simplest and most used methods to solve a multiple linear regression. It consists of minimizing the squared sum of the residuals and obtaining the regression coefficients (beta), which in its matrix form has the following expression.

$$\hat{\beta} = (X' \cdot X)^{-1} \cdot X' \cdot y = \left(\frac{1}{n} \sum x_i x_i' \right)^{-1} \cdot \left(\frac{1}{n} \sum x_i y_i \right)$$

Where:

- $\hat{\beta}$ is the vector of regression coefficients that defines the regression
- x_i are the explanatory or independent variables
- y_i is the response or dependent variable
- n is the length of the set of data

In presence of outliers OLS may provide wrong models (Arslan, 2011). An outlier is a piece of data that is suspected to be incorrect due to the remote probability that it is in fact correct

(Knight & Wang, 2009). In other words, outliers are defined as atypical data being outside normal ranges for common values, produced by an external variable (e.g. abnormal price negotiation, wrong typing, and urgent repair). If an outlier is considered in the regression, the output model will probably predict wrong results.

There are several robust regression methods to cope with this problem (Andersen, 2008; Björck, 1996). Some of them are the Danish Method by Krarup et al. 1980 (purely heuristic method with no rigorous statistical theory), Least Absolute Values method or L1-norm by Edgeworth 1887 (minimizes the sum of the absolute weighted residuals), Least Median Squares by Rousseeuw 1984 (minimizes the median of the weighted residuals squared), Least Trimmed Squares by Rousseeuw 1984 (excludes the largest weighted squared residuals from the minimization), R-estimators by Jaeckel 1972 (minimizes the sum of the scored ranked weighted residuals), M-estimators by Huber 1964 (iteratively minimizes a weighted function of the residuals), IGGIII estimators by Yang in 1999 (similar to M-estimators with a different weighting function), S-estimators by Yohai et al. 1984 (minimizes a robust measure of the scatter of the residuals), and MM-estimators by Yohai 1987 (combines the S-estimator and the M-estimator).

In 2009 Knight & Wang tested and compared many of these robust methods to identify which of these have the greatest ability to correctly exclude outliers. He found that no method correctly identifies 100% of outliers in all situations. From the results obtained, as the level of "contamination" increases, the robust methods of MM-estimators and the L1-norm achieved the highest rates of correct outlier exclusion. However they are more difficult to apply and sometimes can be more time consuming. On the other hand, the differences between the success rates are at most of the order of 10% (Knight & Wang, 2009).

One of the most widespread and easy to use methods corresponds to the *iteratively reweighted least squares method* (IRLS), which belongs to the M-estimators. It consists of iteratively adjusting some "weights" and minimizing the following expression

$$b^{t+1} = \min_{\beta} \sum_{i=1}^{n} w_i\left(\beta^t\right)\left|y_i - f_i\left(\beta\right)\right|^2$$

Where:

- β^t is the vector of regression coefficients at *t* step

- y_i is the response or dependent variable
- f_i is the response regression function
- w_i is the weighting factor
- n is the length of the set of data

While the parameter estimates cannot be obtained in a closed form, the main advantage of IRLS is that provides an easy way to compute the approximate solution. In this research, the MATLAB routine *robustfit.m* is used to perform calculations. Detailed explanation is shown in Appendix B.

Although robust regression allows easily detecting and removing unit prices that present an uncommon behavior, it cannot be immediately applied due to the exponential nature of economies of scale (Hernandez-Sancho et al., 2010). A common relationship in most of road maintenance activities is observed: as long as contracted volume increases, unit prices tend to decrease, and in many cases this relationship seems to decrease exponentially. Then a logarithmic transformation can be applied in both axis (unit prices and contracted volume), transforming the exponential point distribution into a linear point distribution (Maurer et al., 2010). In this new "space" robust regression can be used.

To illustrate the application of the model, an example analysis is performed and shown next. In figure 3-1 unit price against contracted volume of road maintenance activity 7.304.4c *Asphalt slurry seal* is plotted. Exponential decay while increasing contracted volume is easily observed and no possible linear regression can be fitted. Red line shows the final exponential robust fitting.

Then the data can be "linearized". The application of a logarithmic transformation (natural logarithm) in both axes is shown in figure 3-2.

It is clearly observed that data tends to group in a straight line in this space. Here, linear robust regression can be performed without problems. Also, four outliers are easily recognizable (top of the figure), and they were dismissed from the robust regression (red line).

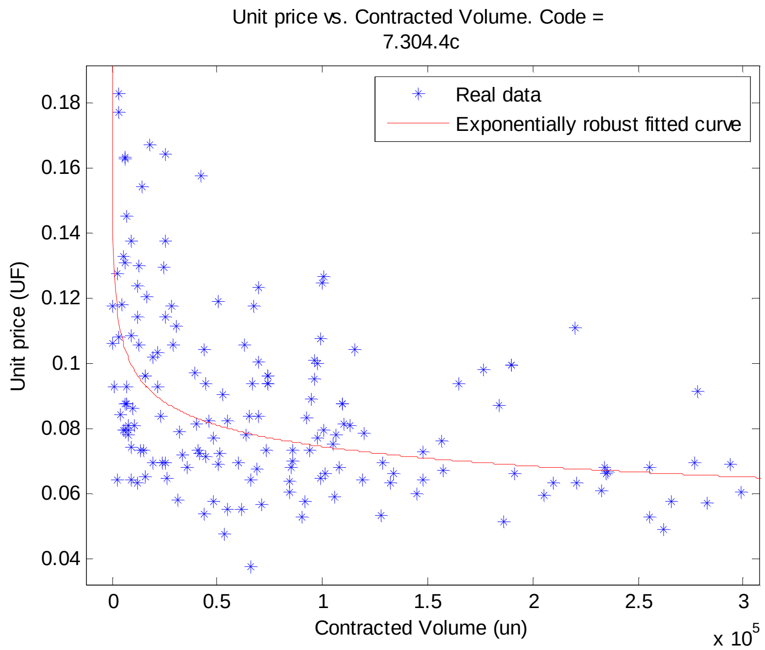

Figure 3-1 – Exponential decay of unit prices against contracted volume

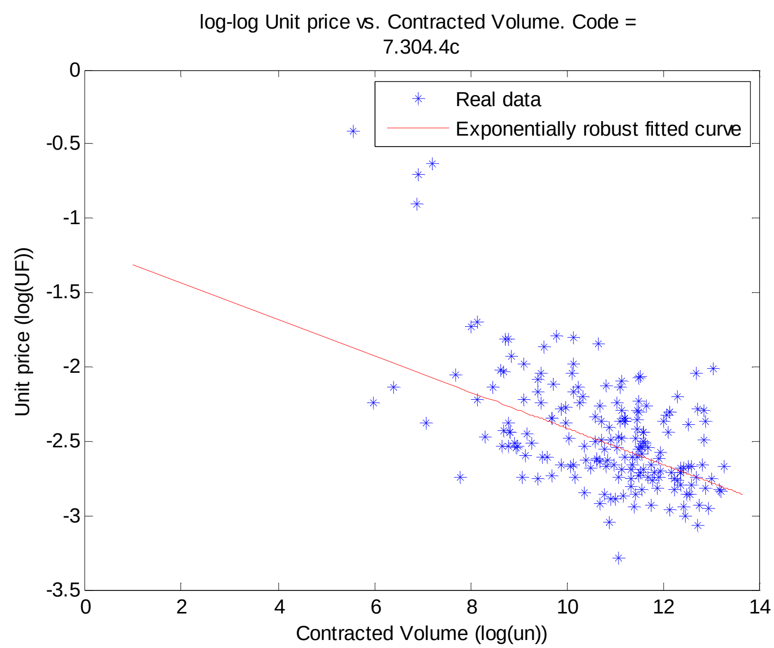

Figure 3-2 – Logarithmic space transformation

3.2.2.- Outlier influence

In this section, the influence of outliers in the regressions is described.

Figure 3-3 summarizes the presence of outliers in the databases used. A vector of weighting values results from each regression, containing factors ranging from 0 (record not considered) to 1 (record fully considered). A record is considered as an outlier if its resulting weighting value is less than 0.20. For each robust regression performed, outlier's ratio is computed (number of outliers divided by number of total records of a specific regression). Then, they are plotted in a normalized histogram, where *x axis* refers to outlier percentage presence in each regression (percentage of total records), and *y axis* refers to percentage of regressions that contains that amount of outliers.

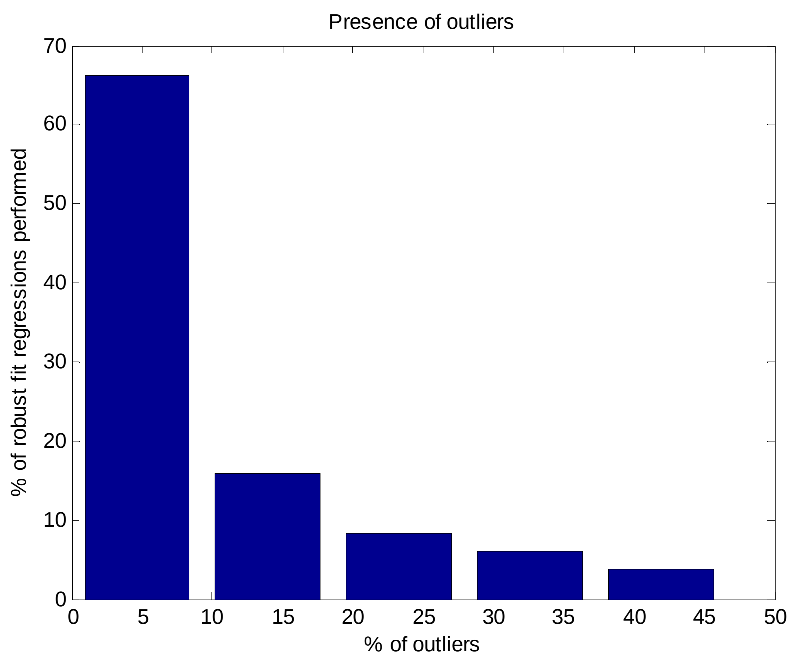

Figure 3-3 – Outlier presence distribution

As it is observed, presence of outliers in data ranges from 0% to 45%. While more than a half of performed robust regressions found less than 5% of outliers, in the rest of them outlier's records varies from 5% to 45% of the input data.

The following section shows examples of outlier's influence if they are included in the regression.

Figure 3-4 shows unit prices against contracted volume for activity 7.304.2b *Cold mix manual surface patching*. Robust regression as well as ordinary least square (OLS) regression method are performed and plotted.

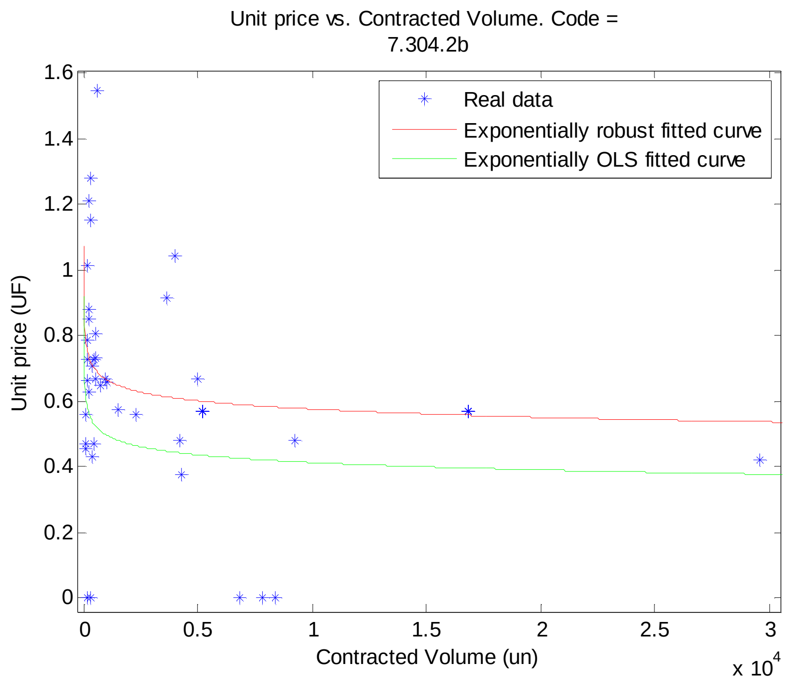

Figure 3-4 – 7.304.2b *Cold mix manual surface patching curve* fitting comparison

As observed from the figure, the green curve is shifted downwards due to low-value outliers. There is a noticeable difference in unit prices for a wide range of contracted volume (UF 0.20 approximately). For example, contracting 20,000 m^2 of cold mix manual surface patching could lead to a difference of UF 4,000 which is not a negligible amount (e.g. nearly 50% of contracts in this database have a total contract amount of UF 50,000 or less; therefore UF 4,000 would mean an 8% difference in total contract).

In another example, figure 3-5 shows unit prices against contracted volume for activity 7.305.3 *Total thickness fast repair*. Robust regression as well as ordinary least square (OLS) regression method are performed and plotted.

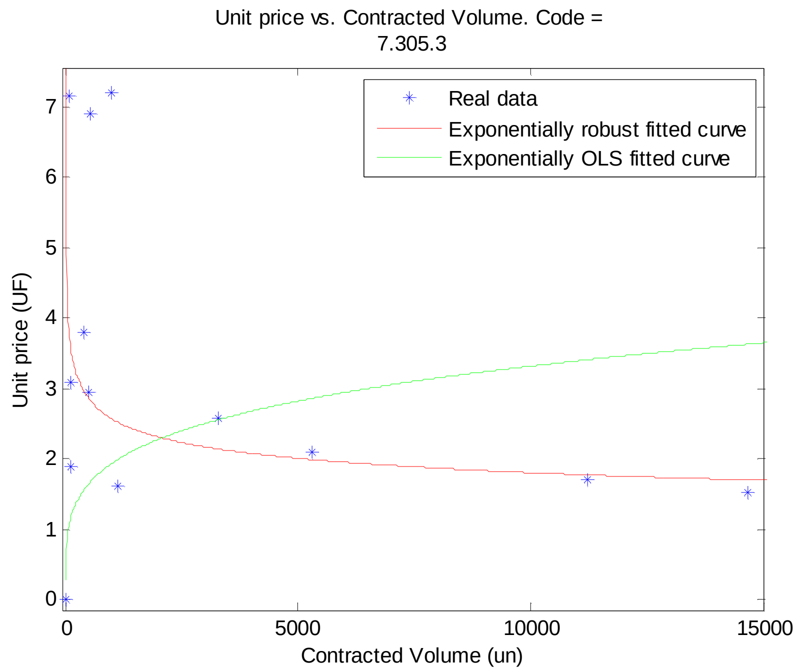

Figure 3-5 – 7.305.3 *Total thickness fast repair curve* fitting comparison

Green curve changes its convexity due to outlier's presence. At the left bottom corner of the graph there is one outlier that produces the curve changes. It obviously dramatically alters output results.

These are some real examples showing the distortions produced by existent outliers in data. By using a robust regression instead of ordinary least squares there is a greater chance of detecting outliers and avoiding distortions, with little more computational effort.

3.3.- Geographical classification of data

Experience suggests that price variability can be reduced by dividing data in different zones or areas. In this section, a Chilean geographical classification is presented and explained. Data verification is performed in next chapters.

While vehicle traffic and climate are two factors affecting pavement deterioration, road maintenance unit prices are influenced by proximity and connectivity to urban areas and contractor's availability. Population density relates to region activity and connectivity, and therefore there is a relationship between traffic volume and region population. The availability of supplies availability and supply facilities are related to the number of urban centers near the building site. These factors are considered and explained to propose a geographical data division.

Tables 3-1 to 3-4 present urban centers with more than 100,000 inhabitants.

Table 3-1 – Urban centers in northern Chile

Name	**Region**	**Population**
Antofagasta	Antofagasta	285,255
Iquique	Tarapacá	214,586
Arica	Arica y Parinacota	175,441
Calama	Antofagasta	136,600
Copiapó	Atacama	134,531

Table 3-2 – Urban centers in central Chile

Name	**Region**	**Population**
Santiago	Metropolitana de Santiago	5,631,839
Valparaíso	Región de Valparaíso	824,006
La Serena	Coquimbo	296,253
Rancagua	Libertador General O'Higgins	236,363
Talca	Maule	208,907
Quillota	Valparaíso	128,874
San Antonio	Valparaíso	106,101
Curicó	Maule	104,124

Table 3-3 – Urban centers in southern Chile

Name	**Region**	**Population**
Concepción	Biobío	848,023
Temuco	Araucanía	268,437
Puerto Montt	Los Lagos	175,140
Chillán	Biobío	165,528
Los Ángeles	Biobío	138,856
Osorno	Los Lagos	132,245
Valdivia	Los Ríos	127,750

Table 3-4 – Urban centers in extreme south of Chile

Name	Region	Population
Punta Arenas	Magallanes y la Antártica Chilena	116,005

There is no urban center with more than 100,000 inhabitants in Aysén region. Furthermore, Aysén and Magallanes are the only regions without land connectivity to the rest of the country (except for insular territories and Palena province). For this reason both are considered isolated regions.

On the other hand, different road surfaces are more intensively used in road building depending on the geographical region (Acevedo & Muñoz, 2010), and each surface needs different road maintenance activities. Whereas asphalt roads are uniformly distributed from northern Chile to X region, concrete roads are almost inexistent in the north and they are more frequent in XII region. Protection layer is used in central Chile (between V and VI region) and stabilized granular is used in north of Chile (mainly in III region). Gravel roads are typical from south of Chile and natural roads are found at extreme north, and regions VII, VIII and IX (Acevedo & Muñoz, 2010).

At the same time climate affects directly on road's deterioration. Sun, rain, wind, and snow are some of the variables that determine maintenance type and periodicity, and thus affecting maintenance costs. Table 3-5 presents the type of climates for each region.

Table 3-5 – Climates of Chilean regions

	Region														
Climate	**XV**	**I**	**II**	**III**	**IV**	**V**	**XIII**	**VI**	**VII**	**VIII**	**IX**	**XIV**	**X**	**XI**	**XII**
Desert climate with abundant cloudiness	x	x	x	x											
Normal desert climate	x	x	x												
Marginal high desert climate	x	x	x	x	x										
High steppe climate	x	x	x												
Low marginal desert climate				x	x										
Steppe climate with abundant cloudiness					x										
Steppe climate with very dry air					x	x									
Warm temperate climate with long dry season of 7 to 8 months					x	x		x							
Warm-temperate climate with winter rains, prolonged dry season (7 to 8 months) and high cloudiness						x		x							
Temperate climate with winter rains and long dry season (7 to 8 months)							x								
Warm temperate climate with dry season of 4 to 5 months						x	x	x	x	x					
Warm temperate climate with a short dry season (less than 4 months)										x	x				
Temperate rainy climate with mediterranean influence										x	x	x	x		
Cold temperate west coast with maximum winter rains												x	x	x	x
Trans-Andean continental climate with steppe regeneration														x	x
High freezing climate														x	x
Cold steppe climate															x
Tundra climate															x

Source: *Dirección General de Aeronáutica Civil. Dirección Meteorológica de Chile*

As it is shown in table 3-5, Chile roughly presents four different climates: desert, warm temperate, rainy temperate and cold climates. These climate regions all have similar number of urban centers and connectivity, except for cold climates (*Aysén* and *Magallanes* isolation).

In order to reduce variability in prices, and considering connectivity and climate factors presented, a geographical data record division is proposed in table 3-6. A statistical analysis of price variability reduction is performed in next chapters.

Table 3-6 – Geographical division of data

Zone	Region	Name
Zone 1	XV	Arica y Parinacota
	I	Tarapacá (incluye Arica y Parinacota)
	II	Antofagasta
	III	Atacama
Zone 2	IV	Coquimbo
	V	Valparaíso
	RM	Metropolitana de Santiago
	VI	Libertador General Bernardo O'Higgins
	VII	Maule
Zone 3	VIII	Biobío
	IX	Araucanía
	XIV	Los Ríos
	X	Los Lagos
Zone 4	XI	Aysén del General Carlos Ibáñez del Campo
	XII	Magallanes y de la Antártica Chilena

3.4.- Cost contingency

An economy of scale model gives a unit price against contracted volume curve. However, it provides a deterministic value, while in practice it behaves like a random variable. Instead of assigning a single value for the cost estimate, the probabilistic approach associates a probability distribution to the cost parameter. This random variable presents higher variability for smaller contracted values and tends to stabilize for higher ones.

Cost contingency is defined as the amount needed to add to an estimation in order to reduce the over exceeding risk to an acceptable level (Baccarini, 2004; Idrus et al., 2010). If we want to estimate unit prices, but we ignore material suppliers and contractors past performance, it is highly recommended to include a contingency item within the estimation. This is easily done with the regressed model described earlier.

The final purpose is to allow the user of this methodology to choose different probabilities that the unit price will not be exceeded, and thus determine the unit price associated to that level of confidence.

Figure 3-6 shows a typical example of how unit price variability decreases with higher contracted volumes for most road maintenance activities.

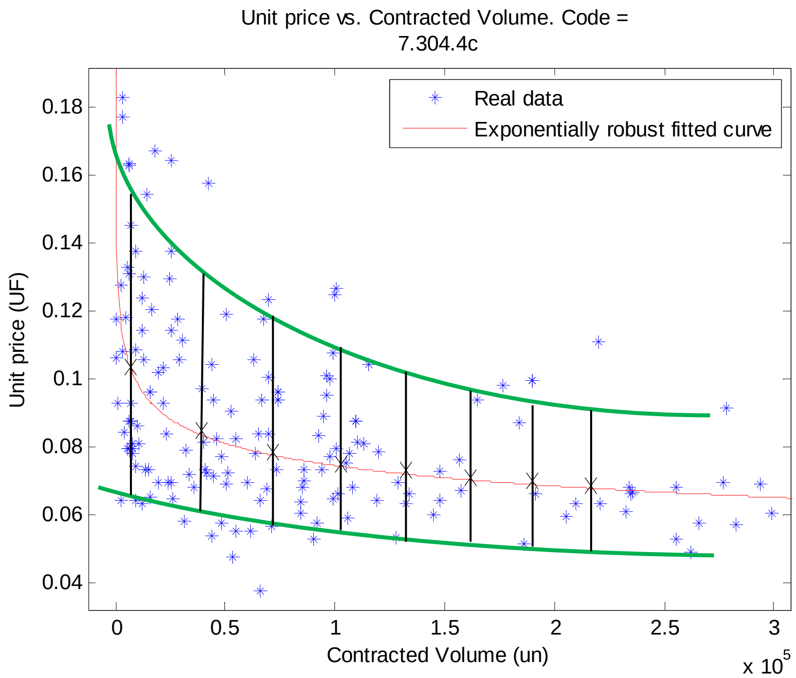

Figure 3-6 – Unit prices variability

The reason of this decrease in variability is that for small volume contracts most of providers use the sale price as they want, and contractors have no interest in negotiating this price. However, for big volumes there is an ability to take advantage of economies of scale, and cost estimators are pushed to estimate more accurately (Tighe, 2001).

To cope with this problem, the proposed methodology divides contracted volume range in three levels of contingency. It assumes that data is normally distributed around mean unit prices (Osman, 2005). For specific road maintenance activities, the proposed volume ranges vary according the total historic range and have the same length.

$$R_i = \max(V_i) - \min(V_i)$$

$$\Delta R_i = R_i / 3$$

Where:

- R_i is the total historic volume range for *i*-activity
- ΔR_i is the total historic volume range for *i*-activity divided by three

- V_i is the vector of volume data for *i*-activity

Lower and upper volume bounds for each interval are defined as following.

$$\lim_{i\,\inf\,j} = \min(V_i) + \Delta R_i \cdot (j-1) \qquad \forall j = 1,2,3$$
$$\lim_{i\,\sup\,j} = \min(V_i) + \Delta R_i \cdot j \qquad \forall j = 1,2,3$$

Where *j* corresponds to the contracted volume ranges: low (1), medium (2) or high (3).

Considering the unit prices according to each interval, average unit price $\mu_{UP\,ij}$ and its standard deviation $\sigma_{UP\,ij}$ are computed. At the same time, and average volume value $\mu_{V\,ij}$ for each interval is calculated as reference value for low, mid, and high contracted volume amount.

For each interval, a variance coefficient is defined as the ratio between standard deviation and mean value. It represents the ratio to consider as standard deviation for a specific unit price value.

$$f_{UP\,ij} = \frac{\sigma_{UP\,ij}}{\mu_{UP\,ij}}$$

Where:

- $\sigma_{UP\,j}$ is the unit price standard deviation of interval *j* for *i*-activity
- $\mu_{UP\,j}$ is the average unit price of interval *j* for *i*-activity
- $f_{UP\,j}$ is the variation coefficient of interval *j* for *i*-activity

After robust regression parameters are generated and mean unit prices calculated, a contingency analysis can be performed. Some variables are defined:

- UP: unit price random variable
- UP_{reg}: mean unit price obtained from regression
- UP_c: unit price considering a contingency level

Then, unit price associated to a contingency level is calculated as it follows. As an example, 15.9% probability of exceeding UP_{reg} is considered (i.e. adding one standard deviation to the mean).

$$P\left(UP \geq UP_c\right) = 84.1\% \quad \rightarrow \quad UP_c = UP_{reg} + 1 \cdot \sigma = UP_{reg} \cdot \left(1 + 1 \cdot f\right)$$

In general, for a desired X probability of not exceedance, the inverse cumulative normal function with zero mean and unitary standard deviation needs to be used to obtain the associated amount "z" of standard deviations of unit prices needed to be added to the average unit price.

$$\Phi^{-1}\left(\frac{z-\mu}{\sigma}\right) = \Phi^{-1}\left(\frac{z-0}{1}\right) = X \rightarrow z$$

Some common relations are the following

$$\begin{aligned} P &= 97.7\% \rightarrow +2 \cdot \sigma \\ P &= 84.1\% \rightarrow +1 \cdot \sigma \\ P &= 50.0\% \rightarrow +0 \cdot \sigma \\ P &= 15.9\% \rightarrow -1 \cdot \sigma \end{aligned}$$

It is worth noting that using a high contingency level does not ensure that real unit prices will be actually lower that UP_c calculated. In other words, if the "experiment" is repeated infinite times with different sample spaces, similar confidence levels would be obtained.

3.5.- Price updating

3.5.1.- Price updating introduction

A procedure for periodic unit price updating is proposed, considering macro-economic variables from the country such as inflation, dollar price and crude oil price.

Price inflation occurs when goods and services prices in an economy increase in time. Economists believe that inflation is more common when demand overcomes supply (Geiger, 2003). In the same way, road project costs tend to change in time due to inflation. The reason of this relies on project competition for same resources (labor, asphalt) that are required by others (Geiger, 2003).

The essential time to consider inflation is when the project budget is being prepared. If historical cost data are being used to develop base year cost estimates for a project, the historical cost data should be adjusted to base year dollars using an inflation index (Geiger, 2003).

The updating methodology is based on price indexes proposed by Chilean Institutes, Central Bank of Chile and *National Statistical Institute* (INE). Relationships are determined by using a robust multiple linear regressions between these indexes and macro-economic variables. Figure 3-7 shows a flowchart of macro-economic variables and price indexes relationships.

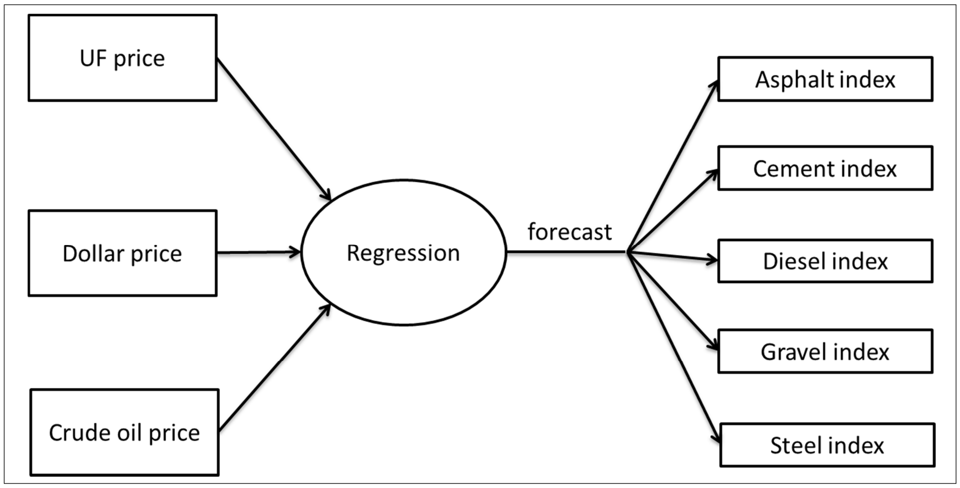

Figure 3-7 – Price indexes relationship flowchart

With these equations it is possible to update each unit price component. For each data record in database, respective prices indexes are calculated according to its date. Then, regressed curves references indexes are calculated by averaging record indexes. A general explanation of the price index estimation for each regressed curve is shown in figure 3-8.

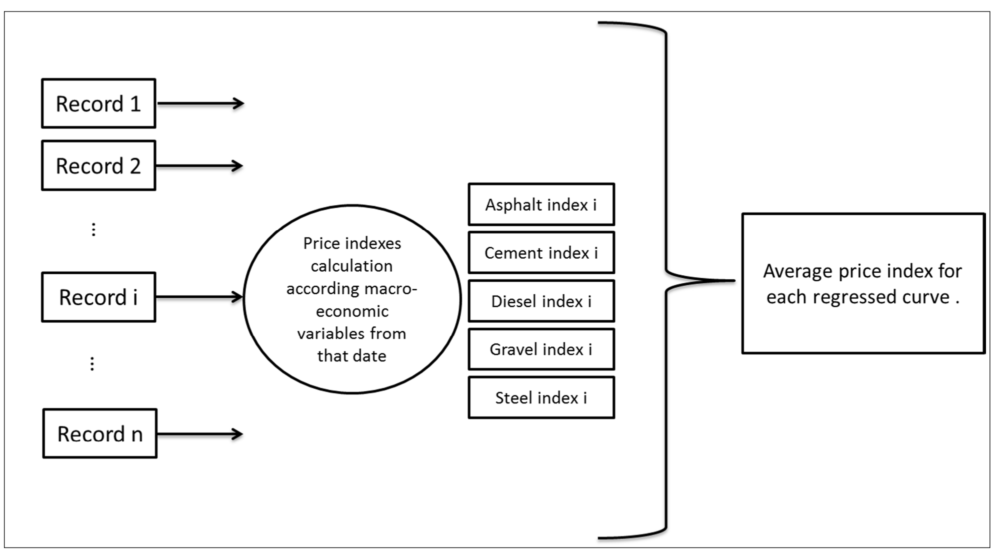

Figure 3-8 – Price index estimation for each regressed curve

Finally, considering current price indexes it is possible to calculate an adjustment factor for each index. These factors are intended to be applied to adjust unit prices according each activity composition. In general terms, asphalt, concrete, gravel and steel indexes will be applied to material item; diesel index will be applied to oil item; labor and overhead cost will be updated by inflation (Apanavičienė et al, 2009). In case of existence of more than one factor per composition item, mean value of factors is considered.

The studied road maintenance activities are divided into three groups: asphalt paved road activities, concrete paved road activities, and gravel and natural roads activities. Each of them uses different materials, equipment, and supplies; thus they must be considered separately. Considering descriptions given in *Manual de Carreteras Volumen 7* principal supplies can be summarized in Appendix A.

3.5.2.- Macro-economic variables

In this section, macro-economic variables of inflation, crude oil price and dollar price are presented. Different graphs showing their behavior during last decades are shown.

3.5.2.1.- Inflation

Consumer price index (IPC in Spanish) measures price variation from a basket of goods and services acquired by urban homes in Chile (INE, 2009). Items with higher weight are the following.

- Cars (new and used)
- Rental services
- Gasoline
- Lunch
- Transportation
- Travel
- Electricity
- Higher education
- Bread
- Beef
- Water
- Mobile and residential phone

- Beverages
- Maintenance labor

Many of these products are not directly associated with material item of maintenance activities. However, they do present a relationship between labor and overhead items. This is the reason why inflation is included either as IPC index or as UF (*unidad de fomento*, Chilean count unit adjusted by inflation).

Figure 3-9 shows UF monthly variation (in Chilean pesos) since 1987 until 2011.

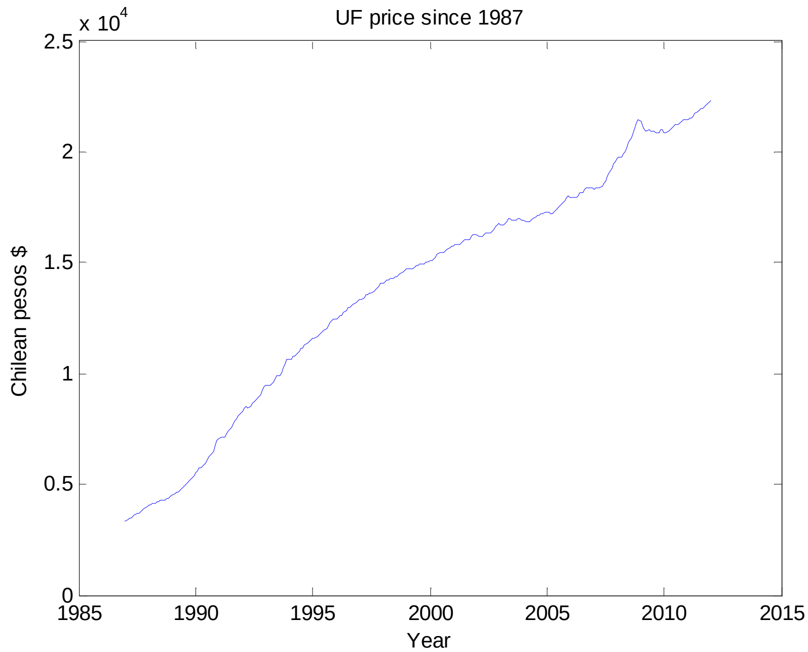

Figure 3-9 – UF price variation from 1987 to 2011

Source: *Central Bank of Chile*

3.5.2.2.- Crude oil price

Most of road maintenance activities, mainly those associated to asphalt paved roads, require asphalt binder to be performed. Figure 3-10 chart shows crude oil price monthly variation (in US$) since 1987 until 2011.

Between 2007 and 2009 there is an important price fluctuation in oil barrel price. This variation is important, and will affect price updating results as shown later.

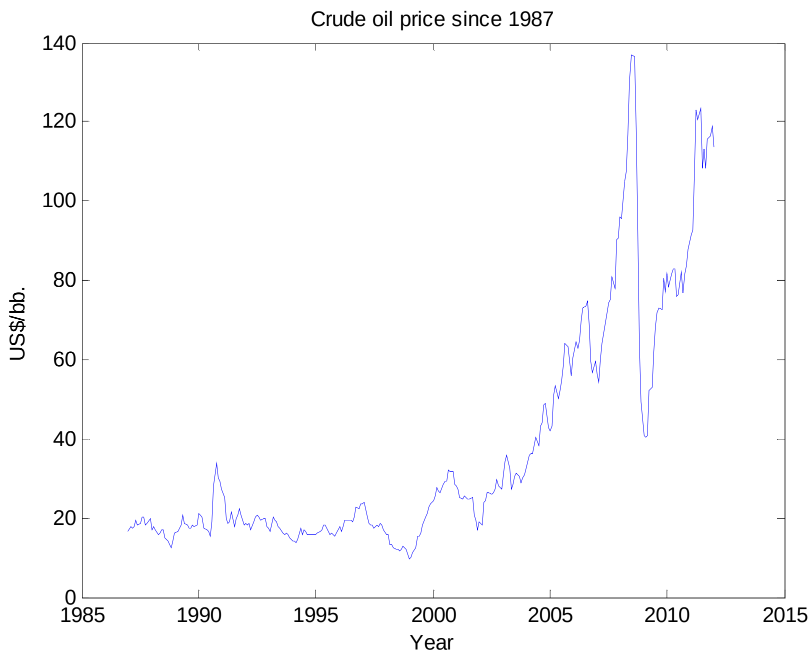

Figure 3-10 – Crude oil price variation from 1987 to 2011

Source: *Central Bank of Chile*

3.5.2.3.- Dollar price

Many road maintenance activities as well as equipment use imported supplies (materials, repairs and parts). Therefore, dollar price variation is relevant for them. Figure 3-11 shows dollar price monthly variation (in Chilean pesos) since 1987 until 2011.

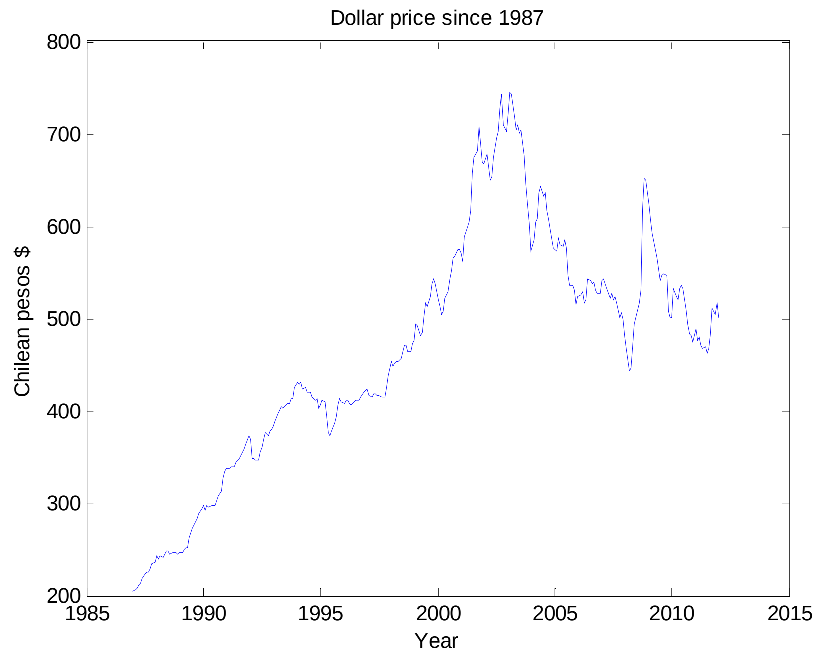

Figure 3-11 – Dollar price variation from 1987 to 2011

Source: *Central Bank of Chile*

3.5.3.- Price indexes

The aim of this section is providing a relationship between macro-economic variables and price indexes.

IPM price indexes (*Índices de precios al por mayor*) are one of the economical monthly indexes calculated by *Instituto Nacional de Estadísticas (INE)*, and they are used to determine price tendencies and price policies (Ahumada, 2005). Price concept used within IPM corresponds to sales net value without taxes.

The relationship between macro-economic variables and price indexes are described next. A robust regression procedure is considered.

3.5.3.1.- Asphalt index

Monthly records of crude oil price (in US$/barrel), UF value (in Chilean pesos), and asphalt index are considered since 2006. Data was processed in MATLAB. Using a robust regression model the following expression was obtained and it is plotted in figure 3-12.

$$I_{asphalt}(oil, UF) = -532.2638 + 0.3790 \cdot oil + 0.0363 \cdot UF$$

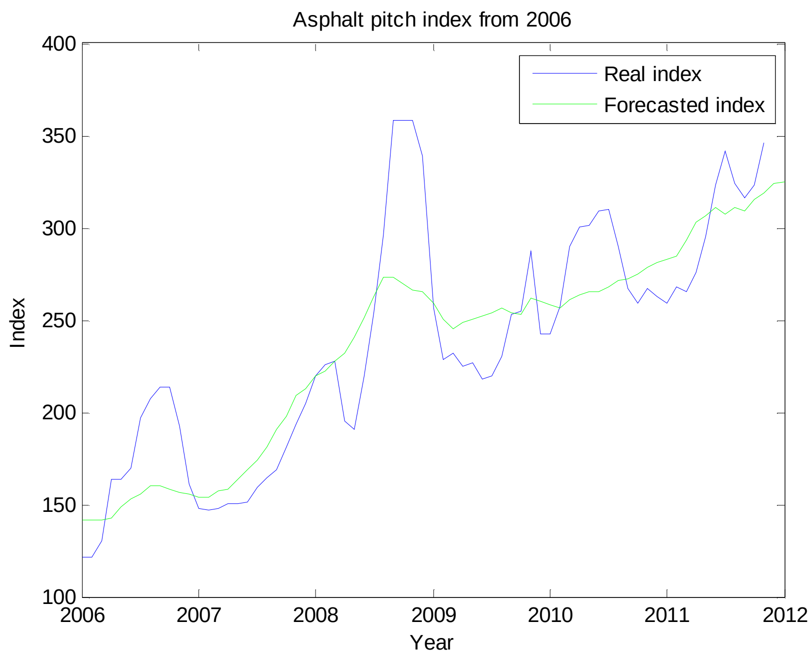

Figure 3-12 – Comparison between real asphalt index and forecasted asphalt index

3.5.3.2.- Cement index

Monthly records of dollar price (in Chilean pesos), crude oil price (in US$/barrel), UF value (in Chilean pesos), and cement index are considered since 1993. Data was processed in MATLAB. Using a robust regression the following expression was obtained and it is plotted in figure 3-13.

$$I_{cement}(USD, UF, oil) = 179.9236 + 2.1355 \cdot USD + 0.0720 \cdot UF + 17.9224 \cdot oil$$

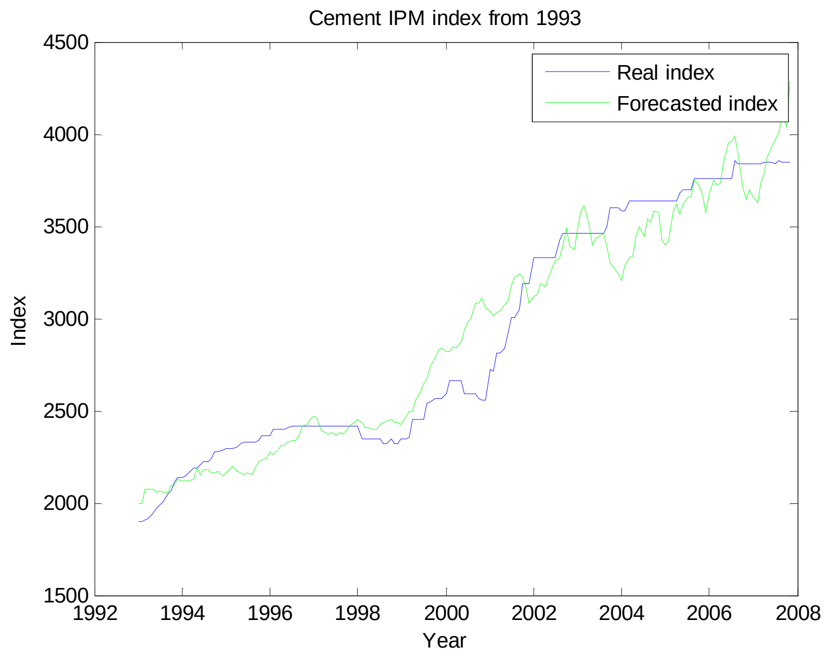

Figure 3-13 – Comparison between real cement index and forecasted cement index

3.5.3.3.- Diesel index

Monthly records of dollar price (in Chilean pesos), crude oil price (in US$/barrel) and diesel index are considered since 1993. Data was processed in MATLAB. Using a robust regression the following expression was obtained and it is plotted in figure 3-14.

$$I_{diesel}(USF, oil) = 77{,}535.5404 + 176.5556 \cdot USD + 4{,}422.6174 \cdot oil$$

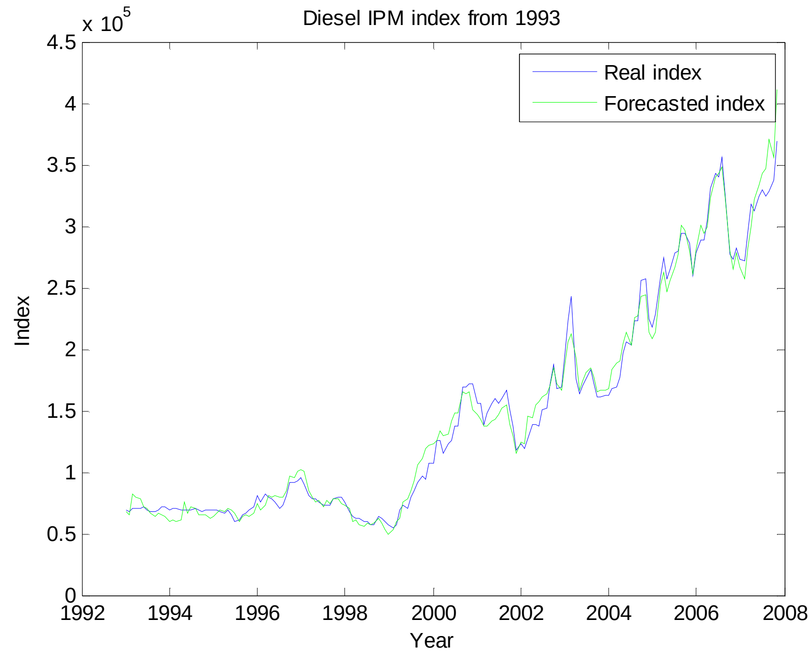

Figure 3-14 – Comparison between real diesel index and forecasted diesel index

3.5.3.4.- Gravel index

Monthly records of dollar price (in Chilean pesos) crude oil price (in US$/barrel), UF value (in Chilean pesos), and natural gravel index are considered since 1993. Data was processed in MATLAB. Using a robust regression the following expression was obtained and it is plotted in figure 3-15.

$$I_{gravel}(USD, UF, oil) = -69{,}823.1748 + 192.3243 \cdot USD + 11.3510 \cdot UF - 1{,}145.7319 \cdot oil$$

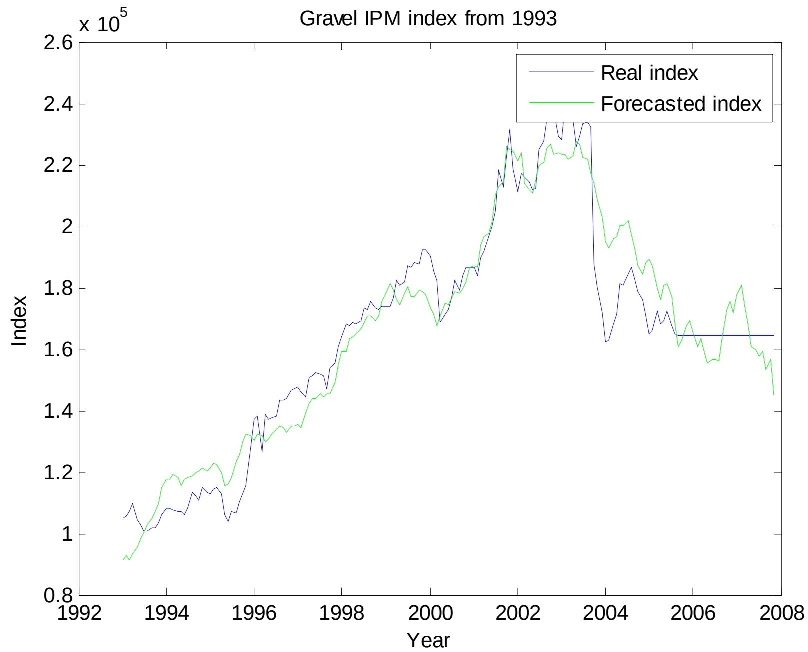

Figure 3-15 – Comparison between real gravel index and forecasted gravel index

3.5.3.5.- Steel index

Monthly records of crude oil price (in US$/barrel), UF value (in Chilean pesos), and steel index are considered since 1993. Data was processed in MATLAB. Using a robust regression the following expression was obtained and it is plotted in figure 3-16.

$$I_{steel}(UF, oil) = -77{,}474.3028 + 14.2870 \cdot UF + 4{,}886.0634 \cdot oil$$

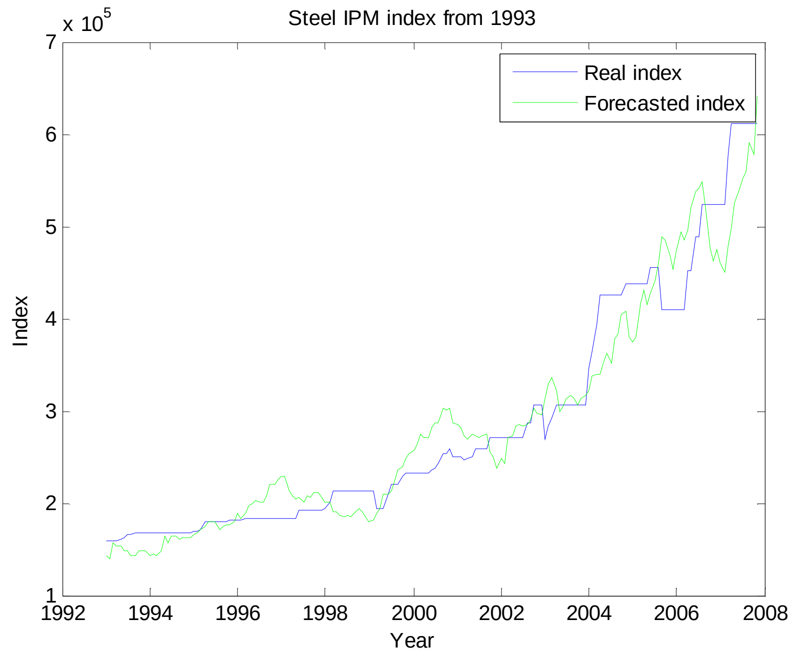

Figure 3-16 – Comparison between real steel index and forecasted steel index

3.5.3.6.- Curve adjustment comparison

Including unnecessary predictors can degrade the efficiency of the resulting estimation procedure and yield less accurate predictions. On the other hand, omitting an important explanatory variable may produce biased parameter estimates and prediction results (Arslan, 2011).

Optimal curve determination was performed by comparing R^2 coefficient of determination of regressions. This coefficient corresponds to variability proportion accounted in the model and it provides a measure of forecasting reliability. $R^2 = 1$ implies that all values are contained within the regressed curve, whereas $R^2 = 0$ implies that predictive variables are incapable to predict the objective function. For the same data, coefficients of similar magnitude imply that the additional variable is no significant and is preferred not to be included in the model.

Table 3-8 shows R^2 for all price indexes and different variable combinations. The highest value of R^2 for each price index, with the fewest amount of predictive variables, are highlighted to show which predictive variables were selected in each case.

Table 3-7 – R^2 values for each index regression

Indexes	Predictive variables						
	UF	Dollar	Oil	UF + Dollar	UF + Oil	Dollar + Oil	UF + Dollar + Oil
Asphalt	0.751	0.002	0.203	0.760	**0.826**	0.413	0.826
Cement	0.849	0.501	0.703	0.853	0.892	0.917	**0.932**
Diesel	0.708	0.254	0.947	0.729	0.970	**0.988**	0.988
Gravel	0.519	0.813	0.041	0.824	0.812	0.822	**0.917**
Steel	0.716	0.199	0.886	0.770	**0.925**	0.909	0.925

3.5.4.- Price updating methodology

These relationships between macro-economic variables and price indexes allow price updating methodology. To illustrate it an example case is presented.

Price updating of activity 7.304.4c *Sello Tipo Lechada Asfáltica* (asphalt slurry seal) is considered. This activity has units of $/m^2, and its price composition is presented in figure 3-17.

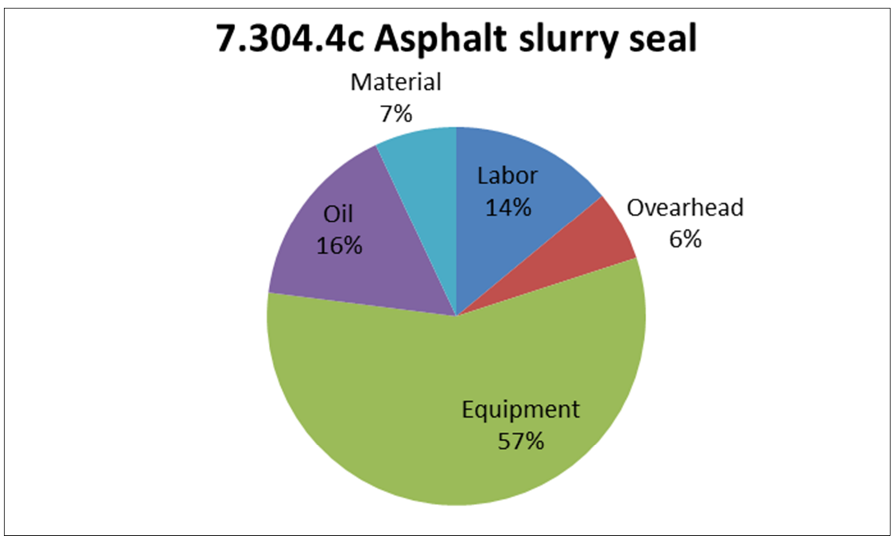

Figure 3-17 – 7.304.4c *Asphalt slurry seal* price composition

Data processing determines average indexes for each activity, while, at the same time, current indexes are computed. Both were obtained by equations mentioned earlier. Table 3-8 shows computed indexes.

Table 3-8 – 7.304.4c *Asphalt slurry seal* average and current price indexes

	Asphalt	**Cement**	**Diesel**
Regression average index	272	4,525	599,305
Current index	316	4,936	680,907
Factor	1.163	1.091	1.136

Activity 7.304.4c *Asphalt slurry seal* average unit price without adjustment is 0.0660 UF/m^2. It is composed by

- Labor 0.009 UF/m^2
- Overhead 0.004 UF/m^2
- Equipment 0.037 UF/m^2
- Oil 0.011 UF/m^2
- Material 0.005 UF/m^2

Then cost components are updated. Oil is multiplied by 1.136 and material is multiplied by 1.163. Labor, overhead and equipment updating is not shown because unit prices are in UF, which already considers general price inflation.

$$UP_{updated} = 0.009 + 0.004 + 0.037 + 0.011 \cdot 1.136 + 0.005 \cdot 1.163 = 0.0683$$

Finally, adjusted 7.304.4c *Asphalt slurry seal* unit price is 0.0683 UF/m^2.

This example illustrates how this methodology works. In following chapters, a comparison between real and predicted updating factors is performed to check forecasting capabilities.

3.6.- Summary

In this chapter, an explanation of the economies of scale problem was presented. A logarithmic/exponential transformation in combination to an IRLS robust regression is considered for the methodology.

In order to decrease price variability, a geographical division of data was proposed considering different climates and connectivity. A statistical analysis of the decrease in price variability is performed in next chapter.

Then, the random nature of unit prices and how cost contingency is included to reduce the over exceeding risk to an acceptable level were described. For each activity, three contracted volume ranges were defined in order to demonstrate that price variability decreases for higher contracted volumes.

Finally a price updating methodology was presented. Macro-economic variables of inflation, dollar price, and crude oil price were described. Relationships between these variables and prices indexes were presented. An illustrative example of the updating technique was presented.

4.- MODEL IMPLEMENTATION AND VALIDATION

4.1.- Introduction

In this chapter, statistical analyses are performed to corroborate or dismiss the research hypotheses. Then, some example cases are presented to illustrate methodology capabilities. Finally results are analyzed and recommendations are given.

4.2.- Statistical analysis

4.2.1.- Geographical division of data variance comparison

Statistical analysis that data zone division decreases variability in unit prices it is performed. To do it, unit price variances were compared considering 5 different groups:

- Zone 0: data from all regions of Chile
- Zone 1: data from regions XV, I, II, and III.
- Zone 2: data from regions IV, V, RM, VI, and VII
- Zone 3: data from regions VIII, IX, and X
- Zone 4: data from regions XI, and XII

Fisher variance ratio test is used to identify whether exists difference in variance between groups Z1, Z2, Z3 and Z4 with Z0. This test considers two samples taken from two populations with normal distribution, mutually independent. Sample standard deviations are denoted S_1, S_2. Hypothesis are:

- Null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$
- Alternative hypothesis: $H_1 : \sigma_1^2 > \sigma_2^2$

Under null hypothesis, test statistic corresponds to

$$f_0 = \frac{S_1}{S_2}$$

This corresponds to a one-tailed test. Null hypothesis is rejected if

$$f_0 > f_{\alpha, n_1 - 1, n_2 - 1}$$

Alpha (α) is the significance level, and n_1, n_2 are degrees of freedom.

Tests are performed for all road maintenance activities by comparing four zones to Zone 0. In order to see whether there is a tendency, tests are performed for both $\alpha = 0.05$ and $\alpha = 0.10$. These tests are performed using *vartest2* routine in MATLAB. Detailed explanation is shown in Appendix B.

After tests were performed, the ratio of road maintenance actions in which null hypothesis was rejected (i.e. zone variability is less than global variability) is computed. Results are shown in table 4-1.

Table 4-1 – Decrease in variability after geographical division

α	**Zone 1**	**Zone 2**	**Zone 3**	**Zone 4**	**Average**
0.05	60.7%	61.1%	55.9%	78.6%	64.1%
0.10	60.7%	63.9%	64.7%	78.6%	67.0%

For a 90% level of confidence ($\alpha = 0.10$), 67.0% of road maintenance actions statistically present a lower unit price variability if they are grouped in described zones. For the other 23.0% there is no significance difference in variability.

Considering only Zone 4 (Aysén and Magallanes), 78.6% of road maintenance activities statistically present a lower unit price variability. This is explained by poor connectivity and isolation of these regions with the rest of Chile, distorting and creating a “bubble” of prices a contractor may offer.

Then, for all maintenance activities an average unit price value was determined for each of four zones (Z1, Z2, Z3, and Z4) and also for the global country (Z0). Considering unit prices from all regions together, the following averages are computed:

- Unit prices from Zone 1 are 1.05 times greater than Z0
- Unit prices from Zone 2 are 0.97 times smaller than Z0
- Unit prices from Zone 3 are 1.38 times greater than Z0
- Unit prices from Zone 4 are 1.58 times greater than Z0

Zone 2 factor is nearly 1.00, whereas the rest of the factors are bigger. These results are partly explained by the large amount of data records from Zone 2, which means that most of the available the data corresponds to this zone. Zone 4 geographical isolation produces higher unit prices compared with the rest of the country.

These results justify analyzing the model separately for regions that may have different climate conditions, connectivity, and urban development.

However, factors were also computed for all Chilean provinces (except for *Isla de Pascua* and Antartic territory). It is worth noting that *Palena* province (*Los Lagos* region, Zone 3) does not have terrestrial connectivity, thus a high factor should be expected for itself. The unit prices of *Palena* province are 1.685 higher than the rest of the country (confirming its isolation), and this implies that Zone 3 factor might be overestimated. A list of price factors for all Chilean provinces is included in Appendix A.

4.2.2.- Exponential robust fitting

In order to assess the regression model two statistics test are performed (Bewick et al, 2005; Kutner et al., 1996; Montgomery & Runger, 2002).

4.2.2.1.- Regression significance test

This test consists of verifying regression significance using analysis of variance. It is used to check whether a statistical linear relation between response and at least one of the predictor variables exist. Hypothesis to test are the following:

$$H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$$

$$H_1 : \beta_i \neq 0 \qquad \text{for at least one i}$$

Where β_i is the i-regression coefficient that defines the regression.

Definitions:

- MSR: regression quadratic mean
- MSE: mean quadratic error
- k: MSR degree of freedom
- n-(k+1): MSE degree of freedom

The statistic is for H_0

$$F_0 = \frac{MS_R}{MS_E}$$

At least one predictor variable has to be rejected in order for regression to exist. Null hypothesis is rejected for α level of significance if F_0 statistic satisfies

$$F_0 > f_{\alpha,k,n-(k+1)}$$

A significance level of 0.05 is commonly used.

4.2.2.2.- Lack-of-fit test

Second test assesses lack-of-fit of the regression model. Hypothesis to test are the following:

- H_0: model does not present lack-of-fit.
- H_1: model presents lack-of-fit.

This test consist of dividing residual squared sum in

$$SS_E = SS_{LOF} + SS_{PE}$$

Definitions:

- SS_{LOF}: lack-of-fit squared sum.
- m: SS_{LOF} degrees of freedom.
- SS_{PE}: pure error squared sum
- n: SS_{PE} degrees of freedom.

The statistic for H_0 is

$$F_0 = \frac{SS_{LOF} / (m-2)}{SS_{PE} / (n-m)}$$

If lack-of-fit term does not have significant difference with pure-error term, null hypothesis is not rejected, verifying good model fit. Null hypothesis is rejected for α level of significance if F_0 statistic satisfies

$$F_0 > f_{\alpha,m-2,n-m}$$

A significance level of 0.05 is commonly used.

4.2.2.3.- Results

Calculations are performed in MATLAB by using *lofrtest.m* routine (Trujillo-Ortiz et al., 2005).

Results showed that 64.7% of regressed curves present goodness of fit for both tests. For the remaining 35.3% that presented lack-of-fit, 83.3% do not have more than 20 records. Then, it is concluded that a critical variable is the amount of input data.

More results and analysis are shown in the example cases. Summarized results are shown in Appendix A. Detailed explanation of MATLAB routine used is shown in Appendix B.

4.2.3.- Unit price variability decrease as volume increases

Statistical analysis that unit price variability decreases as contracted volume increases is performed. To do it, unit price variances are compared considering 3 different contracted volume intervals for each road maintenance activity.

- Range R1: low contracted volume
- Range R2: medium contracted volume
- Range R3: high contracted volume

Fisher variance ratio test is used to identify whether exists difference in variance between groups R1-R2, R2-R3, and R1-R3. This test considers two samples taken from two populations with normal distribution, mutually independent. Sample standard deviations are denoted S_1, S_2. Hypothesis are:

- Null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$
- Alternative hypothesis: $H_1 : \sigma_1^2 > \sigma_2^2$

Under null hypothesis, test statistic corresponds to

$$f_0 = \frac{S_1}{S_2}$$

This corresponds to a one-tailed test. Null hypothesis is rejected if

$$f_0 > f_{\alpha, n_1 - 1, n_2 - 1}$$

Alpha (α) is the significance level, and n_1, n_2 are degrees of freedom.

Tests are performed for all road maintenance activities, comparing range R1 with range R2, range R2 with range R3, and range R1 with range R3. In order to see whether there is a tendency, tests are performed for different significance levels ($\alpha = 0.01, 0.05, 0.10, 0.20$). These tests are performed using *vartest2* routine in MATLAB. Detailed explanation is shown in Appendix B.

After all test were performed, the ratio of road maintenance actions in which null hypothesis was rejected (i.e. variability of a low contracted volume is higher than variability of a high contracted volume) is computed. Results are shown in table 4-2.

Table 4-2 – Decrease in variability after contracted volume range division

	Ranges		
α	**R1 - R2**	**R2 - R3**	**R1 - R3**
0.05	65.5%	69.2%	75.0%
0.10	79.3%	69.2%	91.7%

For a 99% level of confidence ($\alpha = 0.01$), 65.5% of road maintenance activities statistically present a lower unit price variability in range R2 compared to R1; 69.2% present lower unit price variability in range R3 compared to R2; and 75.0% presents lower unit price variability in R3 compared to R1. These ratios are slightly increased as confidence level is relaxed.

Variability decrease as contracted volume increases can be explained by the ability to take advantage of economies of scale. In fact, when contract requires more material, the overall unit costs decreases and probably a more detailed price estimation is performed by the contractor.

Considering this fact (decrease in unit price variability) the cost contingency assignment for each specific range of contracted volume is justified. Otherwise cost contingency might be erroneously assigned.

4.2.4.- Price updating factors comparison

In order to determine price updating methodology reliability, price updating factors are compared. Price updating factor is defined as the times a unit price changes between time t_2 and t_1.

$$f_{t_2,t_1} = \frac{UP_{t_2}}{UP_{t_1}} \qquad t_2 > t_1$$

Where:

- UP_t corresponds to a unit price in time *t*.
- f_{t_2,t_1} corresponds to the price updating factor between time t_1 and t_2.

For the studied road maintenance activities, the average unit price for each contracted volume range, from year 2007 until 2011 are calculated. Then, real updating factors f_R are calculated for periods 2007-2011, 2008-2011, 2009-2011, 2010-2011.

On the other hand, for each regressed curve average index prices are computed, and forecasted updating factors f_F are calculated by each year average indexes (see 3.3.4 Price Updating Methodology).

For comparison, real factors minus forecasted factors differences are considered ($f_R - f_F$). The ratio (percentage) of number of activities with a maximum absolute factor difference of 0.15 divided by total number of activities is considered. This is a conservative value, considering that average contingency price variability ranges from 25% to 60% of mean values.

The more number of records considered the best the forecasted updating factor. Six different numbers of minimum amounts of records are considered: 0, 20, 50, 100, 250, 400 records.

In figures 4-1 and 4-2, the ratio of activities (*z axis*) against number of records (*x label*) and forecasted years (*y label*) are plotted, for low (R1) and high contracted volume (R2 and R3) respectively.

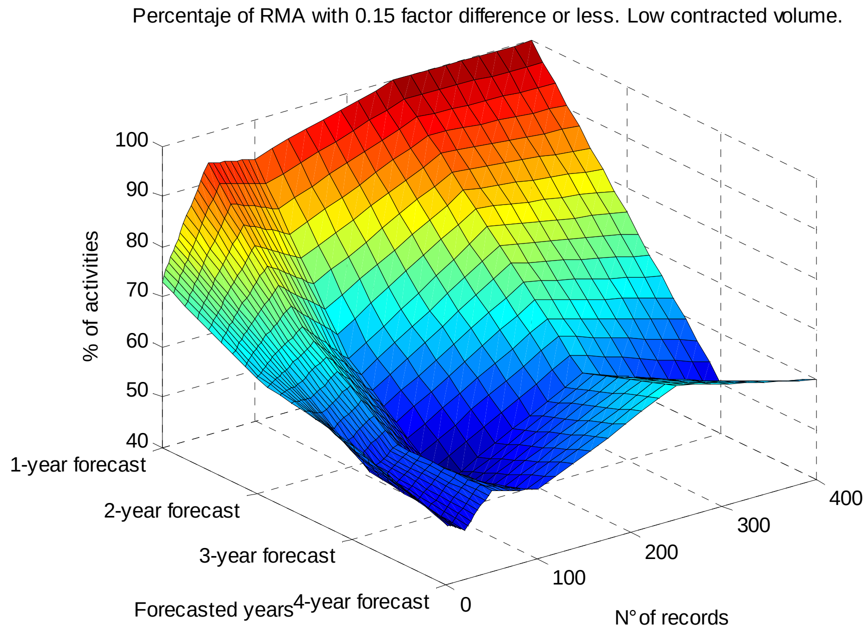

Figure 4-1 – Percentage of RMA with 0.15 factor difference or less, low volume

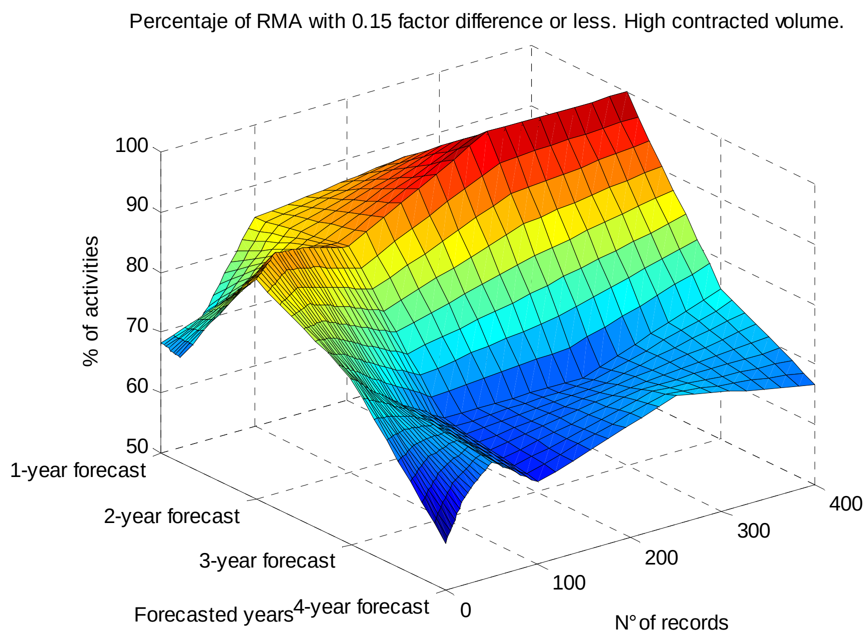

Figure 4-2 – Percentage of RMA with 0.15 factor difference or less, mid and high volume

Despite there is a strange pick at 2-year forecast (figure 4-2) probably due high rises in crude oil prices from mid-2010 to mid-2011, no significance difference is observed between different ranges of contracted volume (figure 4-1 and 4-2). This result confirms that average regression values were used in the calculations.

Forecasting tends to improve as the number of records increases. For each activity, for 300 records or more updating factor difference tends to stabilize; then this may be a recommended value.

The variable that seems to most affect this procedure is the forecasting time ahead. At a first glance, this methodology works fine for a period of 1 and 2 forecasted years, while it drops dramatically for 3 and 4 forecasted years. However, carefully analyzing the results, it is observed that crude oil price presents huge variations in 2008-2009 (see chapter 3, figure 3-10) which exactly coincides with this fall. Probably crude oil price big variations do not immediately reflect in activity unit prices, and that is the main explanation for this anomaly. However, as a greater time-range of information was not available when conducting this research, no other conclusion may be extrapolated.

In conclusion, this updating procedure works equally fine for high contracted volume unit prices as well as for low contracted volume. As long more data records are included in the model it tends to improve. At least 300 well time-distributed records are recommended to provide good updating factors. In case of abrupt changes in macro-economic variables (e.g. changes in crude oil price) this updating procedure must be taken with care. Considering the crude oil price changes of 2008, this procedure works fine for a period of 4-years forecasting. However, a deeper study with a long period of analysis should be conducted.

4.2.5.- Collinearity

One of the assumptions for the experimental design is that there should not exist a high collinearity between independent variables (SAS, 2005). This means that the average unit price of an activity for a specific contracted volume range should not depend on the geographical zone selected, and vice versa, the average unit price of an activity for a specific geographical zone should not depend on the contracted volume range selected.

To verify this, a hypotheses test for a difference in means with variance unknown is used.

Sample means are denoted $\bar{X}_1, \bar{X}_2$ and sample standard deviations are denoted S_1, S_2. Hypothesis are:

- Null hypothesis $H_0 : \mu_1 - \mu_2 = \Delta_0$
- Alternative hypothesis: $H_0 : \mu_1 - \mu_2 \neq \Delta_0$

In this case $\Delta_0 = 0$ is being looked for. Under null hypothesis, test statistic corresponds to

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

This corresponds to a two-tailed test. Null hypothesis is rejected if

$$t_0 > t_{\alpha/2,\nu} \text{ or } t_0 < -t_{\alpha/2,\nu}$$

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(S_1^2 / n_1\right)^2}{n_1 - 1} + \frac{\left(S_2^2 / n_2\right)^2}{n_2 - 1}}$$

Alpha (α) is the significance level, and n_1, n_2 are degrees of freedom of the samples.

For each activity, sample means considering 4 geographical areas (Z1, Z2, Z3, and Z4) and three contracted volume ranges (R1, R2, and R3) are compared with a base case (consisting of all the geographical zones and the complete contracted volume range). Then, for each case, the statistical test is performed and the number of hypothesis rejection is counted among all activities. Table 4-3 shows the percentage of activities whose hypotheses were not rejected, so there is no reason to believe correlation in their variables.

Table 4-3 – Percentage of activities with no-collinear variables

	R1	R2	R3
Z1	82.4%	85.3%	83.8%
Z2	76.5%	82.4%	82.4%
Z3	72.1%	82.4%	82.4%
Z4	76.5%	80.9%	80.9%

As it can be seen in table 4-3, in all situations more than 70% of the activities do not present significant difference, which means that there is no reason to believe there is a collinear relation between independent variables. It is interesting noting that the worst performance corresponds to volume range R1, probably due to the high variability existing in low contracted volume ranges. On the other hand, ranges R2 and R3 present very similar results suggesting that they could be treated as one entity.

4.2.6.- Sensitivity analysis

A sensitivity analysis is performed by changing independent variables under *ceteris paribus* assumption. For each activity, one variable is changed and the rest is kept fixed. Afterwards, a percentage variation is calculated. Finally, an average variation for that variable among all activities is calculated.

The variables considered for the sensitivity analysis are the following.

Contracted volume

- Low
- Medium
- High

Contingency

- Low
- Medium
- High

Variations of macro-economic variables are performed according historic standard deviations considering the last five years. These deviations correspond to:

- UF: $1119.46

- Dollar: $45.82
- Crude oil: US$/bb $24.56

When the sensitivity analysis was performed, macro-economic variable values were:

- UF: $22490.20
- Dollar: $484.53
- Crude oil: US$/bb $105.74

For each maintenance activity the variation in unit prices due to a change in a variable is measured. This variation is expressed in unit price percentage of its original value. Afterwards, an average variation for all activities is computed.

The results of the sensitivity analysis are presented next.

The variation from medium volume to low volume produces an average increase in prices of 112.2%, whereas a variation from medium volume to high volume produces an average decrease in prices of 32.0%. Similarly, an increase in contingency from medium to high produces an increase in prices of 14.3%, whereas a decrease in contingency from medium to low produces a decrease in prices of 11.2%.

Special attention must be paid when using a low contracted volume. High variability in prices for low-volume contracts produces an increase in prices that cannot be neglected.

A variation of 3.5 standard deviations of macro-economic variables produces the following average changes in unit prices:

- UF: 10.8%
- Dollar: 4.7%
- Crude oil: 10.7%

It is observed that prices are more sensitive to changes of inflation and crude oil price. The analysis shows that variations in dollar price are less important.

Figure 4-3 shows a tornado diagram of the different variables analyzed.

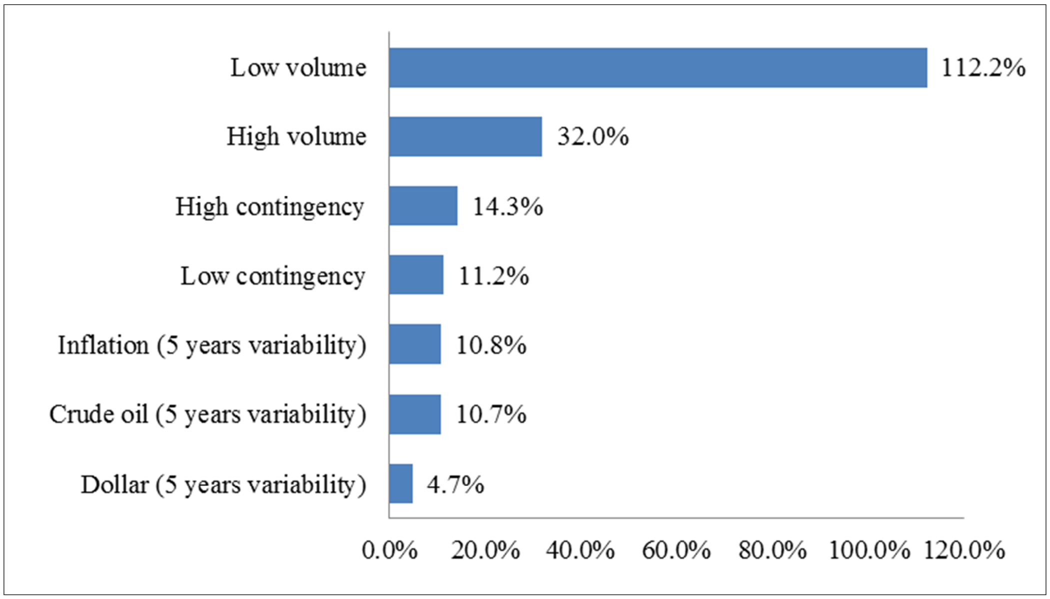

Figure 4-3 – Tornado diagram for percentage variations in unit prices

From the analysis, the following conclusions can be stated.

First, the most incident variable corresponds to contracted volume. For low contracted volumes price variability is considerably high, and for high contracted volumes the decrease in prices cannot be neglected. This confirms that the most incident variable is the contracted volume.

Second, cost contingency assignment may damp those variations produced by macro-economic variables, especially in those periods of high inflation or abrupt changes in crude oil prices.

The most incident macro-economic variables are inflation and crude-oil price. However, the latter currently presents a higher variability due to the global economic context, and its fluctuations are no immediately transferred into prices. Thus, its behavior is riskier than limited inflation.

4.3.- Case studies

4.3.1.- Cross validation

This section provides an example of cross validation for a road maintenance activity, considering 75% of the records for the parameters estimation and the remaining 25% for the model testing.

The activity refers to the rebuilding of road platform that have suffered scour, erosion, complete destruction or is in such state that cannot provide an adequate level of service. Soil filling is included to complete damaged areas, being composed by inorganic soils, free of organic matter, debris, garbage, frozen materials, pieces of rock and other items.

The records belong specifically to embankments (activity 7.302.5d). The unit of measure is cubic meter (m3) of compacted embankment, geometrically calculated according to topography transverse profiles. This activity includes material handling, transportation, and soil compaction.

Total number of records for this activity is 1497. By using a MATLAB random selection of approximately 75% of the data, 1095 records are used for parameter estimation whereas 402 records are left to validate de model.

Exponential robust regression is performed. Beta parameters obtained are b1 = -0.6277, b2 = -0.0405. Figure 4-4 shows the adjusted curve for activity 7.302.5d *Embankments*, considering the 75% of records for parameter estimation, whereas figure 4-5 plots the same curve and the 25% remaining data.

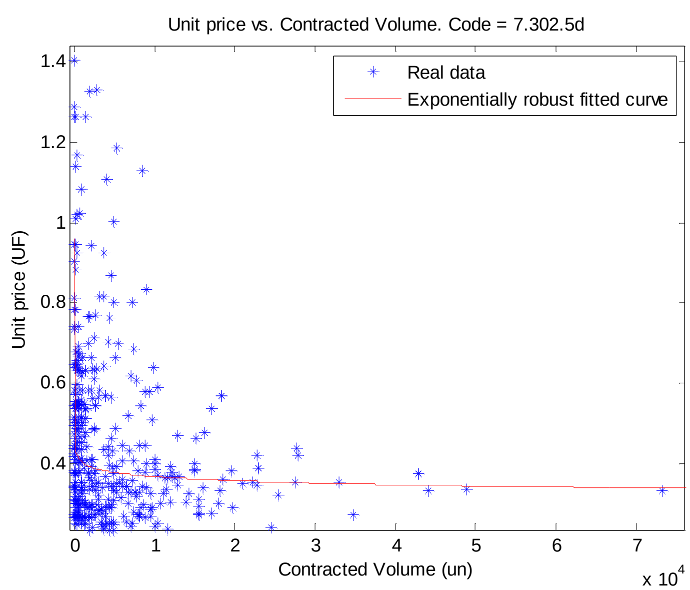

Figure 4-4 – Adjusted curve for activity 7.302.5d *Embankments* and 75% of the records

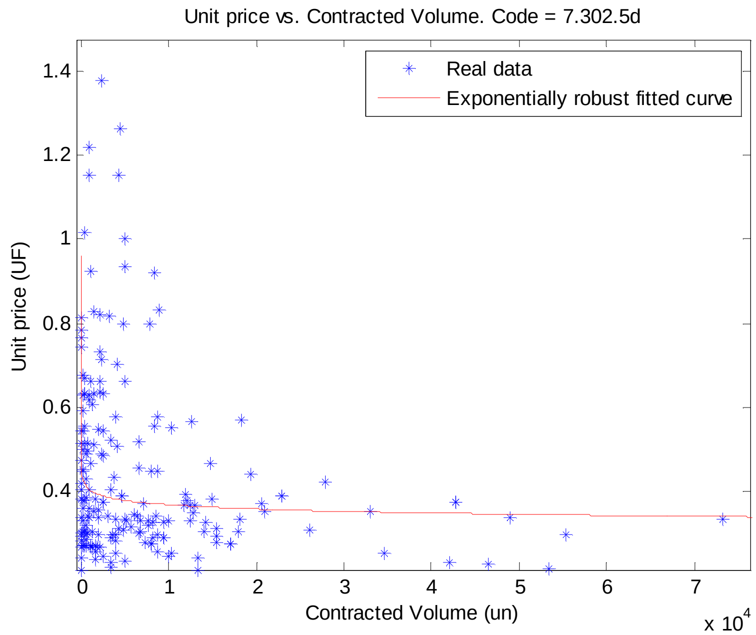

Figure 4-5 – Adjusted curve for activity 7.302.5d *Embankments* and 25% of the records

The residual sum of squares (SSE) is an overall measurement of the discrepancy between the data and the estimation model. It corresponds to the sum of squares of residuals, or the difference between the predicted values and the actual values. The mean squared error for n degrees of freedom is defined as

$$MSE = \frac{SSE}{n}$$

As it is shown next, mean squared error of the model for the estimated data (75% of the records) and for the validation data (25% of the records) are very similar.

$$MSE_{25\%} = 0.1694$$
$$MSE_{75\%} = 0.1615$$

Moreover, a fisher test of variance is used to check whether there is a significant difference between variance of estimation residuals and validation residuals. A significance level of $\alpha = 0.05$ is used.

$$f_{\alpha/2,n25\%,n75\%} = 0.8426$$
$$f_{1-\alpha/2,n25\%,n75\%} = 1.1781$$

$$F_0 = \frac{MSE_{25\%}}{MSE_{75\%}} = 1.0491$$

Thus null hypothesis cannot be rejected and there is no evidence suggesting that the model does not adequate properly.

4.3.2.- Numeric examples

The activities consider in the example cases are the following:

7.304.4c - Sello Tipo Lechada Asfáltica (asphalt slurry seal) (Lorenzen et al., 2001). This asphalt maintenance activity belongs to the group of bituminous seals. It refers to asphaltic covering through asphalt irrigation, in combination with some aggregates. This activity includes surface cleaning and preparation, asphalt material implementation, equipment, compaction, and other final actions. It is quantified in squared meters (m^2).

7.305.1b - Sellado de Juntas y Grietas (joint and crack sealing) (Lorenzen et al., 2001). This concrete maintenance activity consists of sealing or resealing existing cracks in concrete pavements. This operation includes crack cavity conformation, cleaning, and sealing with all required procedures. It is quantified in linear meters (m).

7.306.4a - Recebo de Carpetas Granulares (gravel road platform reshape) (Lorenzen et al., 2001). This gravel maintenance activity belongs to the group of activities intended to reshape road's platforms, and it considers replacing missing material. The aim of this activity is recovering initial geometry and serviceability of the road, and includes road preparation, and material supply and placement. It is quantified in cubic meters (m^3).

In order to test the methodology, the chosen activities have different amount of input data. Table 4-4 shows the number of valid data records for each activity.

Table 4-4 – Example activities input data records

Code	Activity	Unit	N° of records
7.304.4c	Asphalt slurry seal	m2	303
7.305.1b	Joint and crack sealing	m	26
7.306.4a	Gravel road platform reshape	m3	278

Robust regression is then performed in MATLAB software using *robustfit* routine. Detailed explanation of this routine is shown in Appendix B. The parameters obtained from the regression are:

- b_1: first beta coefficient
- b_2: second beta coefficient
- σ_1/μ_2: variation factor for the first volume range
- σ_2/μ_2: variation factor for the second volume range
- σ_3/μ_3: variation factor for the third volume range
- Vp_1: mean contracted volume for the first range
- Vp_2: mean contracted volume for the second range
- Vp_3: mean contracted volume for the third range

The following data was automatically dismissed in robust regression

- 7.304.4c *Asphalt slurry seal*: 3.96% of the data (13 outliers)
- 7.305.1b *Joint and crack sealing*: 3.70% of the data (1 outlier)
- 7.306.4a *Gravel road platform reshape*: 2.51% of the data (7 outliers)

For these particular cases outliers were a small number. However, as explained in 3.1, leaving just a few outliers in the regression can create very bad models due to their different order of magnitude.

Table 4-5 summarizes parameters obtained from MATLAB robust regression. It includes exponential regression coefficients, variation factors for each volume range, and mean contracted volume for each range.

Table 4-5 – Parameters obtained from MATLAB regression

Code	b1 (UF)	b2 (UF/un)	$\sigma 1/\mu 1$	$\sigma 2/\mu 2$	$\sigma 3/\mu 3$	Vp1 (un)	Vp2 (un)	Vp3 (un)
7.304.4c	-1.1965	-0.1215	29.32%	19.41%	15.65%	69,881	273,964	465,880
7.305.1b	-0.3917	-0.3611	65.43%	35.12%	31.07%	1,728	4,586	8,465
7.306.4a	-0.0906	-0.1411	45.88%	45.96%	40.87%	1,893	5,984	11,298

Results for each maintenance activity are shown next.

Results for 7.304.4c *Asphalt slurry seal*

Figure 4-6 shows unit prices for asphalt slurry seal plotted against squared meters contracted, and the exponentially robust fitted curve.

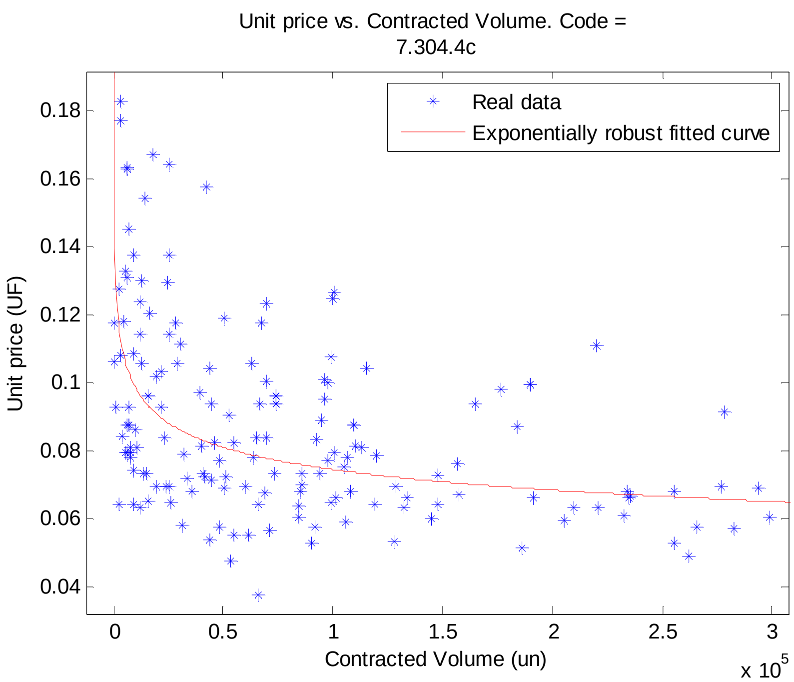

Figure 4-6 – Fitted curve for 7.304.4c *Asphalt slurry seal*

Lack-of-fit test for regression model was successfully carried out. It was performed by *lofrtest* routine in MATLAB, and its results are presented in table 4-6.

Table 4-6 – Lack-of-fit test for 7.304.4c *Asphalt slurry seal* regression

Lack-of-fit test for regression model with independent replicate values.

SOV	SS	df	MS	F	P
Model	7.358	1	7.358	135.837	0.0000
Residual	15.654	289	0.054		
Lack-of-fit	-1834.012	7	-262.002	-39.945	1.0000
Pure error	1849.666	282	6.559		
Total	23.271	290			

Results for 7.305.1b *Joint and crack sealing*

Figure 4-7 shows unit prices for joint and crack sealing plotted against linear meters contracted, and the exponentially robust fitted curve.

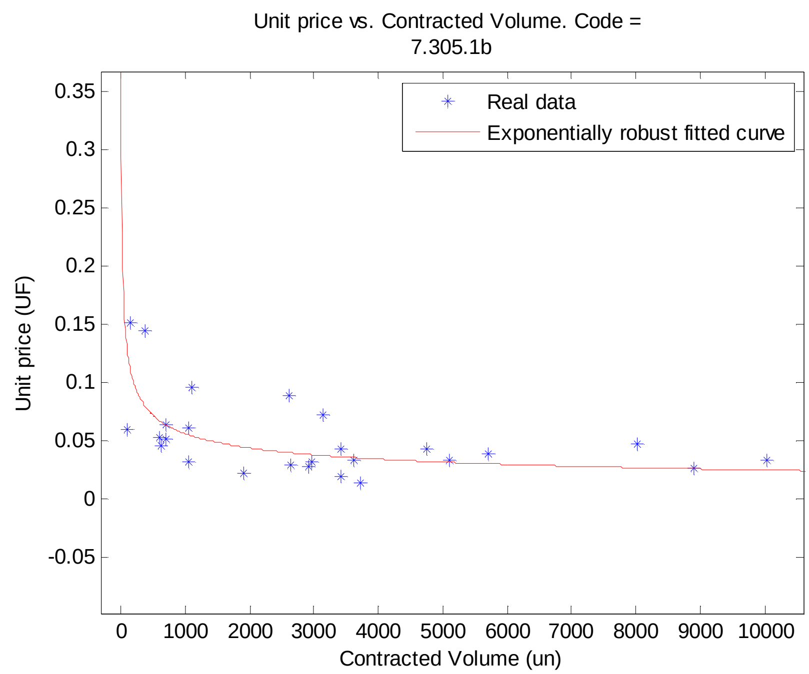

Figure 4-7 – Fitted curve for 7.305.1b *Joint and crack sealing*

Lack-of-fit test for regression model was successfully carried out. It was performed by *lofrtest* routine in MATLAB, and its results are presented in table 4-7.

Table 4-7 – Lack-of-fit test for 7.305.1b *Joint and crack sealing* regression

Lack-of-fit test for regression model with independent replicate values.

SOV	SS	df	MS	F	P
Model	4.669	1	4.669	19.636	0.0002
Residual	5.707	24	0.238		
Lack-of-fit	-167.932	7	-23.990	-2.349	1.0000
Pure error	173.638	17	10.214		
Total	8.453	25			

Results for 7.306.4a *Gravel road platform reshape*

Figure 4-8 shows unit prices for joint and crack sealing plotted against linear meters contracted, and the exponentially robust fitted curve.

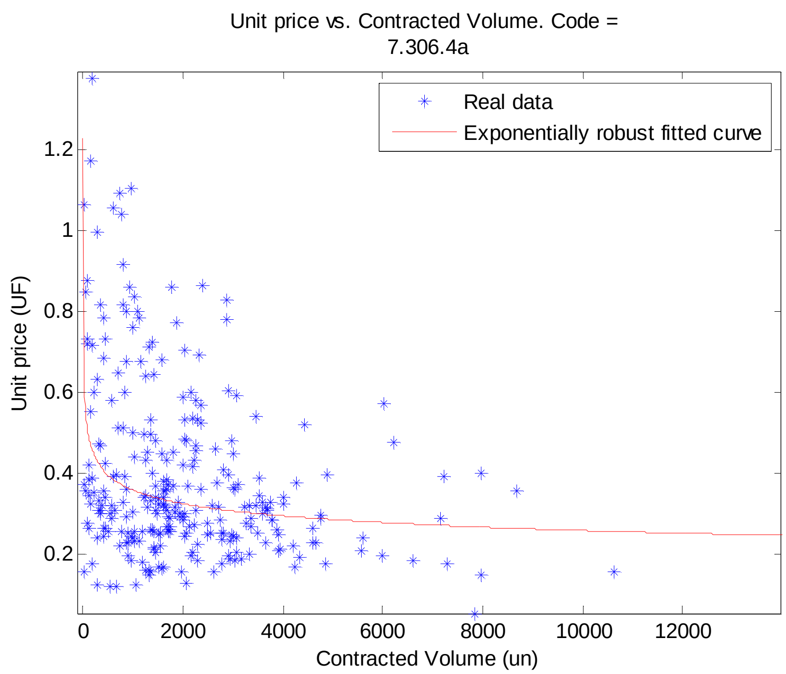

Figure 4-8 – Fitted curve for 7.306.4a *Gravel road platform reshape*

Lack-of-fit test for regression model was successfully carried out. It was performed by *lofrtest* routine in MATLAB, and its results are presented in table 4-8.

Table 4-8 – Lack-of-fit test for 7.306.4a *Gravel road platform reshape* regression

Lack-of-fit test for regression model with independent replicate values.

SOV	SS	df	MS	F	P
Model	5.955	1	5.955	29.906	0.0000
Residual	53.767	270	0.199		
Lack-of-fit	-305.943	7	-43.706	-31.956	1.0000
Pure error	359.709	263	1.368		
Total	60.432	271			

4.4.- Analysis of results

Unit price equations are obtained by performing an exponential transformation.

$$\log\left(UP\left(V\right)\right) = b_1 + b_2 \cdot \log\left(V\right) \quad \backslash \exp(\;)$$
$$\rightarrow \quad UP\left(V\right) = e^{b_1 + b_2 \log(V)}$$

Unit prices curves for example maintenance activities are shown next. Their accuracy can be confirmed by simply replacing example values and comparing them with graphs (figure 4-6, figure 4-7, figure 4-8).

$$UP_{7.304.4c}\left(V\right) = e^{-1.1965-0.1215\cdot\log(V)}$$

$$UP_{7.305.1b}\left(V\right) = e^{-0.3917-0.3611\cdot\log(V)}$$

$$UP_{7.306.4a}\left(V\right) = e^{-0.0906-0.1411\cdot\log(V)}$$

Now consider performing a cost contingency analysis by estimating the unit price associated with adding 1 (one) standard deviation to 7.305.1b *Joint and crack sealing*. Suppose that a volume (in this case "length") of 5,000 meters of sealing is required. Replacing 5,000 in $UP_{7.305.1b}\left(V\right)$ yields to

$$UP_{7.305.1b}\left(5000\right) = e^{-0.3917-0.3611\cdot\log(5000)} = 0.0312\,UF$$

According to table 4-5 this volume corresponds to a mid-range. From table 4-5, variation coefficient for mid-range is 35.12%, which means that the associated standard deviation is 35.12% of the mean. Thus the unit price associated to the contingency level is

$$UP_{c\;7.305.1b}\left(5000\right) = UP_{7.305.1b}\left(5000\right)\cdot\left(1+f\right) = 0.0312\cdot\left(1+0.3512\right) = 0.0422\,UF$$

This means that, for 5,000 meters of sealing, the unit price that has a probability of 15.9% of being exceeded corresponds to 0.0422 UF. This confidence level (84.1%) was arbitrarily chosen due to its simplicity in calculations (by just simply adding 1 standard deviation to the mean), but any value for confidence level between 0.00 and 1.00 can be chosen.

Fitted curves show to successfully model price variation within normal volume ranges (i.e. historic data range included in database). Applying these equations to extreme values (nearly zero volume or infinite volume) may cause anomalies. Special care must be taken when considering low-contracted volumes, because these exponential functions tend to infinity as they approach to zero. For this cases it is suggested to compare nearby unit price estimation to check how fast price changes.

Considering results from all activities 64.7% of them present goodness of fit. Table 4-9 shows road maintenance activities that present lack-of-fit and its number of input data records.

Table 4-9 – Activities with lack of fit, for all regions

Code	**Activity**	**Unit**	**N° of records**
7.304.3a	Bacheo Profundo con Mezclas en Caliente	m2	515
7.304.6d	Nivelación de Bermas en Pavimentos Asfálticos Pavimentadas con Mezcla Asfáltica en Frío	m2	5
7.304.9b	Reperfilado Simple con Riego de Bermas	km	0
7.304.9c	Reperfilado con Compactación	km	4
7.305.1a	Sellado de Juntas y Grietas	km	112
7.305.4a	Reparación de Espesor Parcial	m2	0
7.305.4b	Reparación de Espesor Parcial para Puesta en Servicio Acelerada	m2	5
7.305.6	Instalación de Drenes de Pavimento	m	7
7.305.7a	Cepillado de la Superficie Pavimento de Hormigón	m2	18
7.305.7b	Cepillado de la Superficie Pavimento Asfáltico	m2	5
7.305.9a	Nivelación de Bermas en Pavimentos de Hormigón Revestidas con un Tratamiento Superficial Simple	m2	0
7.305.9b	Nivelación de Bermas en Pavimentos de Hormigón Revestidas con un Tratamiento Superficial Doble	m2	2
7.305.9d	Nivelación de Bermas en Pavimentos de Hormigón Pavimentadas con Mezcla Asfáltica en Frío	m2	0
7.305.10	Reemplazo de Losas por Mezclas Asfálticas	m2	1
7.305.11	Bacheo Superficial Asfáltico en Frío en Pavimentos de Hormigón	m2	13
7.305.13a	Reperfilado Simple de Bermas	km	0
7.305.13b	Reperfilado Simple con Riego de Bermas	km	0
7.305.13c	Reperfilado con Compactación de Bermas	km	0
7.305.14	Colocación de Barras de Traspaso de Carga	N°	4
7.306.1b	Reperfilado Simple con Riego	km	116
7.306.2a	Reparación de Areas Inestables	m3	5
7.306.3b	Bacheo de Capas de Rodadura Granulares con Adición de Estabilizador	m3	1
7.306.4a2	Recebo de Capas de Rodadura Granulares Tamaño 3"	m3	17
7.306.4b	Recebo de Capas de Rodadura Granulares con Material Chancado	m3	601

As it can be seen from results, almost every single activity that did not present goodness of fit had very little input data. There are four exceptions (7.304.3a, 7.305.1a, 7.306.1b, 7.306.4b) in which there is enough input data but there is no regression. These are rare cases where robust fit procedure was not able to distinguish appropriateness of data. For example, activity 7.306.1b *Simple reshaping with irrigation* has [km] units, and when data was gathered some of it was considered in [m], creating two value groups. This kind of data errors cannot be handled by exponential robust regression.

On the other hand, activities goodness of fit has on average a number of 300-400 input records. A minimum between 50 and 100 data records are recommended in order to have goodness of fit.

Two suggestions arise after running the model. It is not very important to consider two volume ranges instead of three (joining R2 and R3), and this can simplify the model. Second, the smaller the geographical division of data, the better the results obtained. However this requires a big amount of good data. Then, zone division may be performed for smaller areas but considering only those having the necessary records.

4.5.- Summary

In this chapter variance comparison tests for the geographical division of data were carried out. In most of road maintenance activities the proposed division succeeded in decreasing price variability. However, zone division may be performed for smaller areas but considering only those having the necessary records.

Also, lack-of-fit tests were performed to validate regressed curves. Most of activities showed goodness of fit, whereas the majority of the remaining activities had fewer than 20 data records. Hence, the number of input data is a critical variable for the methodology.

At the same time, by using a variance comparison test it was proven that most of the activities statistically present lower price variability as contracted volume increases. This result justifies assigning different cost contingency levels for each of the volume ranges defined. However, it was verified that range R2 (mid volume) and R3 (high volume) may be treated as a unified entity, whereas range R1 (low volume) should be treated separately due to its high variability.

The predicting capacity of the price updating methodology was tested by comparing real and predicted updating factors. Although it was very sensitive to abrupt changes in macro-economic variables and to the number of input data records, it showed to work reasonably well for a forecasting period of 4 years.

Finally, some example cases were presented to illustrate the operation of the methodology.

5.- CONCLUSIONS AND RECOMMENDATIONS

5.1.- Conclusions

This research presented a methodology for the estimation of the unit prices of road maintenance activities, which fulfills the main objective. Robust fit regression (IRLS) in combination with a logarithmic transformation successfully allows to model economies of scale of the unit prices of road maintenance activities, which positively answers the main hypothesis of this research.

Secondary objectives are fulfilled because:

- Most of the studied activities presented good model fit and many of them presented outlier data which was well addressed.
- It was shown that a geographical division of data significantly reduces the variability in unit prices, particularly for isolated regions.
- Most of the activities presented an exponential decreasing behavior, and their variability tended to decreases for higher volumes. The latter allowed performing a cost contingency analysis for different ranges.
- It was shown that unit prices can be updated in time using relationships between macro-economic variables of inflation, dollar price, and crude oil price and price indexes of asphalt, cement, gravel, diesel, and steel.
- For the studied activities, parameters were generated and the methodology was validated.
- It was shown that the most incident variable within the model was the contracted volume of road maintenance activities.
- Case studies were developed to explain the methodology operation.

Economies of scale generate that road maintenance activities unit prices present an exponential decreasing behavior, which can be easily modeled with the methodology presented. Furthermore, the method showed to be an effective way to avoid price distortions due to outlier existence in data, which are typically present in this kind of database. Ignoring outlier presence can produce erroneous curve adjustment, and performing this kind of analysis requires negligible more effort in comparison to traditional ordinary least squares regression.

From all studied road maintenance activities 66.2% presented good model fit, whereas 27.9% did not have enough data. The rest 5.9% had wrong input data and were not able to be handled with this procedure. In order to have goodness of fit, a number of at least 50 to 100 input data is required. Moreover, 400 input values was the average number for curves that presented goodness of fit.

On the other hand, fitted curves showed to successfully model price variation within normal volume ranges. For rarely extreme values of contracted volume (nearly zero or infinity) model might provide abnormal unit prices. In order to correct this, total volume range was divided into three intervals, and for each a median volume value was assigned. This allows capturing price variation due to economies of scale, and avoids extreme values anomalies. In case a specific low-contracted volume is looked for, special care must be considered because these exponential functions tend to infinity as they approach to zero. For these cases, comparing with nearby estimations is recommended.

It was also shown that dividing input data into different zones with similar characteristics (similar climates, connectivity with urban cities) allows reducing price variability. Nearly 70% of the studied activities present this behavior. In order to create separate models for each area it is important to divide input data considering homogenous regions. Additionally adjusting factors can be calculated to easily estimate prices among different zones. However, caution must be paid when considering factors from Zone 3 because they may be overestimated by poor connectivity of *Palena* and *Chiloé* provinces. Finally, zone division may be performed for smaller homogeneous areas (e.g. mountain climate), considering only those having the necessary input records.

Road maintenance activities unit prices should not be treated as deterministic values, because in practice they behave like random variables. This research showed for 60–70% of the activities variability of unit prices decreases as contracted volume increases. In other words, for small contracts price variability tends to be considerably high, whereas for big contracts unit prices tend to be more homogeneous. Results suggest this can be explained by the ability to take advantage of economies of scale and due to a more detailed price estimation performed by contractors.

Cost contingency analysis allows assigning a normal probability cost distribution by creating for each activity three contracted volume intervals. This random variable presents higher variability for smaller contracted volumes and tends to stabilize for higher ones. In unit price estimation, where material suppliers and contractors past performance are ignored, it is highly recommended to include a contingency item within the estimation. Using a high contingency level does not ensure that real unit prices will not over exceed estimated value; contingency means that if the "experiment" is repeated infinite times with different sample spaces, similar confidence levels would be obtained.

The most incident variable within the model is the contracted volume of road maintenance activities. For low contracted volume, cost contingency level assignment also plays an important role due to great variability of prices.

Finally, unit prices can be updated in time using relationships between macro-economic variables of inflation, dollar price, and crude oil price and price indexes of asphalt, cement, gravel, diesel, and steel. Those relationships rely on a robust linear fit between macro-economic variables and price indexes. Unit prices are updated considering its price composition (personal, overhead, material, fuel, and equipment). This updating technique works equally fine for high contracted volume as well as for low contracted volume. As long as more records are included in the model it tends to improve the forecasting capacity. At least 300 well time-distributed records are recommended to provide good updating factors. In case of abrupt changes in macro-economic variables this updating procedure must be taken with care.

5.2.- Recommendations

After running the model some recommendations for its implementation arise.

- The most incident variable corresponds to contracted volume. For low contracted volumes, price variability is considerably high. In case of estimating unit prices for low volumes it is recommended to consider a contingency analysis. If the required volume is known, using low volume it is not recommended.
- Cost contingency assignment may damp those variations produced by macro-economic variables, especially in those periods of high inflation or abrupt changes in crude oil prices. Thus, for periods of economic crisis affecting the region, a high contingency is recommended.

- The most incident macro-economic variables are inflation and crude-oil price. However, the latter currently presents a higher variability due to the global economic context. Thus, its behavior is riskier than a bounded inflation.
- It was verified that, for the proposed contracted volume division, range R2 (mid volume) and R3 (high volume) may be treated as a unified entity, whereas range R1 (low volume) should be treated separately due to its high variability. This can simplify the implementation of the methodology.
- The application of the methodology should include a continuous incorporation of data. This can eventually allow dividing the data in smaller geographical areas and generating more accurate models. Also, by considering only the newest records, the use of the predictive functionality for longer periods may not be necessary.
- Database activity codes spelling is a problem not solved in this research, which left out many records from curve fitting. An artificial intelligence procedure could be studied and developed to uniform codes spelling.
- A visual inspection of data prior to regression is recommended, in order to find rare price anomalies (e.g. data with different units). Otherwise, methodology may not be able to handle such errors.

An approach for implementing the proposed methodology for its use in a highway agency is described next. First, generate parameters and export them into a spreadsheet, where unit price estimation and contingency assignment may be performed when required. This parameter generation and exporting process may be automated, and the program may be adapted for the estimation using data from a specific time-period range. If plenty of data is available the updating procedure could be dismissed by considering only historic prices from recent periods.

The spreadsheet may have the possibility to interact with the user, allowing choosing the desired activity to be studied by defining the values of its variables. The current values for macro-economic variables may be automatically gathered from internet. Results may be displayed only for selected activities, but an option to display a reference value for all activities in the database may be useful. Additionally, there could be an option to include the parameters for a new activity which was not previously considered.

5.3.- Future research

The model was tested with road maintenance projects developed in Chile during the last years. Future research should consider applying the model in other countries and places to check its adaptation ability to any kind of data. Also a study involving other geographical data segregation to test the impacts in unit price variability should be performed.

This research considered using robust regression (iteratively reweighted least squares) to deal with dependent variable outlier existence. However, if the outliers occur in the explanatory variables the performance of robust regression estimators are not better than the ordinary least squares estimators (Arslan, 2011). To deal with this problem, a weighted least absolute deviation regression (WLAD) estimation might be tested.

Because of road maintenance activities dependence to asphalt, equipment and labor mainly, macro-economic variables of inflation, dollar price, and crude oil price were considered. A deeper study should consider including other economic variables, or even characterizing each single road maintenance activities with specific variables.

The generated parameters are only valid for a mid-term period of time: 4-5 years from data collecting. Beyond that period new factors may arise (new technologies, new constructive methodologies) or important macro-economic changes may occur that will considerably affect results. This methodology should be tested for longer periods, or a different model should be proposed.

The methodology properly works for the studied road maintenance activities (asphalt, concrete, and gravel roads). However, its applicability for other road maintenance activities (e.g. signage, drainage) is not limited by theory, and should be matter of future research. Furthermore, this procedure might be applied to simulate any other problem where exponential fitting is needed and outlier data is frequently present (e.g., mining processes, general construction, or industrial activities).

Due to the focus of this research in modeling the relationship between unit prices and contracted volume, it was difficult to include all aspects from road management. As it is observed in figure 1-4, there are other factors that influence the project net present worth in an economic assessment. Apart of pavement performance (that is beyond the scope of this research) one of them not considered is discount rate, which influences the final preservation projects portfolio that maximizes social benefits.

In the case of economic analysis of investments it is best practice to forecast life-cycle costs and benefits of a project without inflation. Inflation is very hard to predict, particularly more than a few years into the future (Geiger, 2003). On the other hand, for longer periods of analysis, the discount rate becomes more important. The time value of resources reflects the fact that there is a cost associated with diverting the resources needed for an investment from other productive uses or planned consumption within the economy (Geiger, 2003).

These aspects should be considered when expanding this methodology as part of a broader benefit-cost analysis or life-cycle cost analysis program.

The final goal should be creating an application of benefit-cost analysis that includes this methodology as part of its "brain". Future research may consider including it within existent LCCA software (e.g. HDM-4) and comparing predicted life cycle analysis with real preservation experiences.

REFERENCES

AASHTO. (2007). *A Practical Guide To Estimating*. Design Task Force Technical Committee on Cost Estimating. American Association of State Highway and Transportation Officials

Abelson, P. & Flowerdew, A. (1975). Models for the Economic Evaluation of Road Maintenance. *Journal of Transport Economics and Policy*, 9 (2), 93-114. University of Bath and The London School of Economics and Political Science.

Acevedo J. & Muñoz, D. (2010). *Valor del patrimonio vial de la red vial nacional – Año 2009*. Departamento de Gestión Vial, Dirección de Vialidad, MOP. Santiago, Chile.

Ahumada, G. (2005). *Índice de precios al por mayor (Base junio 1992=100)*. Instituto Nacional de Estadísticas.

Anastasopoulos, P., Bhargava, A., Labi, S., Mannering, F. & Sinha, K. Three-Stage Least-Squares Analysis of Time and Cost Overruns in Construction Contracts. *Journal of Construction Engineering and Management*, 136 (11), 1207-1218. doi 10.1061/ ASCE CO.1943-7862.0000225

Andersen, R. (2008). *Modern Methods For Robust Regression*. SAGE, USA.

Anderson, S., Damnjanovic, I., Nejat, A., & Ramesh, S. (2009). *Synthesis on construction unit cost development*: technical report. Texas Transportation Institute, The Texas A&M University System.

Apanavičienė, R., Juodis, A., & Šiškina, A. (2009). Evaluation of the competitiveness of construction company overhead costs. *Journal of Civil Engineering and Management*, 15(2), 215–224. doi: 10.3846/1392-3730.2009.15.215-224

Archondo-Callao, R. (2011). Maintenance Impact on Economic Evaluation of Upgrading Unsealed Roads. Transportation Research Record: *Journal of the Transportation Research Board*, 2203, Transportation Research Board of the National Academies, Washington.

Arslan, O. (2011). Weighted LAD-LASSO method for robust parameter estimation and variable selection in regression. *Computational Statistics & Data Analysis*, 56(6), 1952–1965. doi: 10.1016/j.csda.2011.11.022

Baccarini, D. (2004). *Accuracy in estimating project cost construction contingency – A statistical analysis*. The international construction research conference of the Royal Institution of Chartered. Leeds Metropolitan University, England.

Bai, X., Yao, W., & Boyer, J. (2012). Robust fitting of mixture regression models. *Computational Statistics & Data Analysis*, 56(7), 2347–2359. doi: 10.1016/j.csda.2012.01.016

Bewick, V., Liz Cheek, L., & Ball, J. (2005). Statistics review 14: Logistic regression. Crit Care. *BioMed Central Ltd*. 9(1): 112-118. doi: 10.1186/cc3045

Björck, A. (1996). *Numerical Methods for Least Squares Problems*. Society for Industrial and Applied Mathematics.

Dakin, J., Stannard, E., & Wightman, D. (2006). *HDM-4 Volume three Software user guide*. Highway Development and Management Series. PIARC.

Geiger, D. (2003). *Economic Analysis Primer*. Federal Highway Administration Office of Asset Managment. U.S. Department of Transportation.

Geiger, D. (2005). *Pavement Preservation Definitions*. Memorandum. Federal Highway Administration Office of Asset Managment. U.S. Department of Transportation.

Gkritza, K. & Labi, S. (2008). Estimating Cost Discrepancies in Highway Contracts: Multistep Econometric Approach. *Journal of Construction Engineering and Management*, 134, (12), 953-962.doi 10.1061/ ASCE 0733-9364 2008 134:12 953

Gransberg, D. & Scheepbouwer, E. (2010) Infrastructure Asset Life Cycle Cost Analysis Issues. *2010 AACE International Transactions*. Texas, USA.

Haas, R., Tighe, S., & Falls, L. (2006). Determining Return on Long-Life Pavement Investments. Transportation Research Record: *Journal of the Transportation Research*

Board, Transportation Research Board of the National Academies, 1974, 10-17. doi: 10.3141/1974-04.

Heggie, I. (1994). *Management and Financing of Roads. An Agenda for Reform.* World Bank Technical Paper number 275. Africa Technical Series.

Hernandez-Sancho, F., Molinos-Senante, M., & Sala-Garrido, R. (2010). Cost modelling for wastewater treatment processes. *Desalination*, 268(1–3). doi:10.1016/j.desal.2010.09.042

Hesham, A. (2011). *Reclycling asphalt recues roads. Solving the maintenance riddle*. Engineering News Record. McGraw-Hill Construction.

Idrus, A., Nuruddin, F. & Rohman, M. (2010). Development of project cost contingency estimation model using risk analysis and fuzzy expert system. *Expert Systems with Applications*, 38(3), 1501-1508. doi: 10.1016/j.eswa.2010.07.061

Instituto Nacional de Estadísticas (INE). (s.f.). *Índice de precios al por mayor (Base junio 1992=100)*. Recuperado el 4 de enero de 2012, de http://www.ine.cl.

Instituto Nacional de Estadísticas. (2009). *Manual Metodológico de Índice de Precios al Consumidor (IPC) Nacional Base Anual 2009=100*. Departamento de Estadísticas de Precios, Subdirección de Operaciones, INE.

Jrade, A. & Alkass, S. (2007). Computer-Integrated System for Estimating the Costs of Building Projects. *Journal of Architectural Engineering*, 13(4). doi: 10.1061/ ASCE 1076-0431 2007 13:4 205

Knight, N., & Wang, J. (2009). A Comparison of Outlier Detection Procedures and Robust Estimation Methods in GPS Positioning. *Journal of Navigation*, 62(4), 699-709. doi: 10.1017/S0373463309990142

Kutner, M., Nachstheim, C., Neter, J., & Wasserman, W. (1996). *Applied Linear Statistical Models*, Fourth Edition. McGraw-Hill Irwin, USA.

Lawrence, G. 2007. Cost Estimate Contingency. Use and Misuse of Capital Cost Estimate Contingency – Why Deleting it Makes Projects More Expensive, Not Less. Pharmaceutical *Engineering® The Official Magazine of ISPE*. September/October 2007, 27 (5).

Lin, C. & Lin, K. (2010). Applying Utility Theory to Cost Allocation of Pavement Maintenance and Repair. *Journal of pavement research and technology*, 4 (4), 212-221

Liu, L., Wehbe, G., & Sisovic, J. (2010). The Accuracy of Hybrid Estimating Approaches: A Case Study of an Australian State Road & Traffic Authority. The Engineering Economist: A *Journal Devoted to the Problems of Capital Investment*, 55(3), 225-245. doi: 10.1080/0013791X.2010.502962

Lorenzen, C., Arriagada, E. & Belmar, N. (2001). *Manual de Carreteras Volumen 2 Procedimientos de Estudios Viales*. Dirección de Vialidad, Ministerio de Obras Públicas, Gobierno de Chile. Santiago, Chile.

Lorenzen, C., Arriagada, E. & Belmar, N. (2001). *Manual de Carreteras Volumen 7 Mantenimiento Vial*. Dirección de Vialidad, Ministerio de Obras Públicas, Gobierno de Chile. Santiago, Chile.

MATLAB Documentation, Statistics Toolbox. (2011). Robustfit. Mathworks

MATLAB Documentation. (2011). Linear Regression, Parametric Regression Analysis. Mathworks.

Maurer, M., Wolfram, M., & Anja, H. (2010). Factors affecting economies of scale in combined sewer systems. *Water science and technology a journal of the International Association on Water Pollution Research*, 62(1), 36-41.doi: 10.2166/wst.2010.241

Montgomery, D. & Runger, G. (2002). *Applied Statistics and Probability for Engineer*. 3rd ed. John Wiley & Sons, Inc., USA.

Orellana, R. (2006). *El mercado del asfalto en Chile y los avances producidos en la última década*. Instituto Chileno del Asfalto. Santiago, Chile.

Osman, H. (2005). Risk-Based Life-Cycle Cost Analysis of Privatized Infrastructure. Transportation Research Record: *Journal of the Transportation Research Board*, Transportation Research Board of the National Academies, 1924, 192-196. doi: 10.3141/1924-24

Pan, N., Ko, C., Yang, M., Hsu, K. (2011). Pavement performance prediction through fuzzy regression. *Journal Expert Systems with Applications: An International Journal archive*, 38 (8), 10010–10017. doi:10.1016/j.eswa.2011.02.007

Robinson, R., Danielson, U., & Snaith, M. (1998). *Road Maintenance Management*. 1st Edition. MacMillan Press.

SAS. (2005). *Concepts of Experimental Design*. Design Institute for Six Sigma.

Salem, O. (1999). *Infrastructure Construction and Rehabilitation: Risk-Based Life Cycle Cost Analysis*. (Doctoral dissertation). University of Alberta, Canada.

Schexnayder, C., Weber, S., & Fiori, C.. (2003). *Cost Estimating: A Synthesis of Highway Practice, Final Report*. National Cooperative Highway Research Program, Project 20-7, Task 152. Transportation Research Board.

Schliessler, A. & Bull, A. (1994). *Caminos. Un nuevo enfoque para la gestión y conservación de redes viales*. Naciones Unidas, Comisión Económica para America Latina y el Caribe. Santiago, Chile.

Tighe, S. (2001). Guidelines for Probabilistic pavement life cycle cost analysis. Transportation Research Record: *Journal of the Transportation Research Board*, Transportation Research Board of the National Academie, 1769, 28-38. doi: 10.3141/1769-04

Trujillo-Ortiz, A., R. Hernandez-Walls, A. Castro-Perez and F.J Marquez-Rocha. (2005). *lofrtest: Lack-of-fit test for regression model with independent replicate values.* A MATLAB file. Available at
http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=7074

Valdés, M., & Guerra, R. (2011). *Estudio básico metodología para la determinación de los costos de acciones de conservación*. Dirección de Vialidad, Ministerio de Obras Públicas, Gobierno de Chile. Santiago, Chile.

Villacres, J. (2005). *Pavement Life-Cycle Cost Studies Using Actual Cost Data, A Synthesis*. Asphalt Pavement Alliance, USA.

Whiteley, L., Tighe, S., & Zhang, Z. (2005). Incorporating Variability into Pavement Performance, Life-Cycle Cost Analysis, and Performance-Based Specification Pay Factors. Transportation Research Record: *Journal of the Transportation Research Board*, Transportation Research Board of the National Academies, 1940, 13-20. doi: 10.3141/1940-02

APPENDICES

APPENDIX A – General information and results

Main supplies for selected road maintenance activities

Code	Road maintenance activity	Unidad	Asphalt	Cement	Steel	Gravel	Diesel
7.304.1a	Sellado de Areas con Grietas de hasta 6 mm de ancho	m2	x				x
7.304.1b	Sellado de Grietas de Ancho entre 6 y 20 mm	m	x				x
7.304.1c	Sellado de Grietas de Ancho entre 20 y 70 mm	m	x				x
7.304.1d	Sellado de Grietas y Cavidades de Ancho superior a 70 mm	m	x				x
7.304.2a	Bacheo Superficial Manual con Mezclas en Caliente	m2	x			x	x
7.304.2b	Bacheo Superficial Manual con Mezcla en Frío	m2	x			x	x
7.304.2c	Bacheo Superficial Manual con Mezclas Preparadas	m2	x			x	x
7.304.2d	Bacheo Superficial Mecanizado	m2	x			x	x
7.304.2e	Bacheo con Tratamiento Superficial Doble	m2	x			x	x
7.304.2f	Parche Superficial	m2	x			x	x
7.304.3a	Bacheo Profundo con Mezclas en Caliente	m2	x			x	x
7.304.3b	Bacheo Profundo con Mezclas en Frío	m2	x			x	x
7.304.3c	Bacheo Profundo con Tratamiento Superficial	m2	x			x	x
7.304.3d	Bacheo Profundo con Mezclas en Frío Predosificadas	m2	x			x	x
7.304.4a	Sello Tipo Riego de Neblina	m2	x			x	x
7.304.4b	Sello Tipo Tratamiento Superficial Simple	m2	x			x	x
7.304.4c	Sello Tipo Lechada Asfáltica	m2	x			x	x
7.304.4d	Sello Localizado con Gravilla	m2	x			x	x
7.304.4e	Sello Localizado con Lechada	m2	x				x
7.304.5	Nivelación de Bermas Granulares no Revestidas en Pavimentos Asfálticos	m2				x	x
7.304.6a	Nivelación de Bermas en Pavimentos Asfálticos Revestidas con un Tratamiento Superficial Simple	m2	x			x	x
7.304.6b	Nivelación de Bermas en Pavimentos Asfálticos Revestidas con un Tratamiento Superficial Doble	m2	x			x	x
7.304.6c	Nivelación de Bermas en Pavimentos Asfálticos Pavimentadas con Mezcla Asfáltica en Caliente	m2	x			x	x
7.304.6d	Nivelación de Bermas en Pavimentos Asfálticos Pavimentadas con Mezcla Asfáltica en Frío	m2	x			x	x
7.304.7	Imprimación Reforzada	m2	x			x	x
7.304.8	Reposición de Capa de Rodadura de concreto Asfáltico	m2	x			x	x
7.304.9a	Reperfilado Simple de Bermas	km					x
7.304.9b	Reperfilado Simple con Riego de Bermas	km					x
7.304.9c	Reperfilado con Compactación	km					x
7.305.1a	Sellado de Juntas y Grietas	km	x				x
7.305.1b	Sellado de Juntas y Grietas	m	x				x
7.305.2	Reparación en Todo el Espesor	m2		x	x	x	x
7.305.3	Reparación en Todo el Espesor para Puesta en Servicio Acelerada	m2		x	x	x	x
7.305.4a	Reparación de Espesor Parcial	m2		x		x	x
7.305.4b	Reparación de Espesor Parcial para Puesta en Servicio Acelerada	m2		x		x	x
7.305.5	Pavimentos Insertados	m2		x	x	x	x
7.305.6	Instalación de Drenes de Pavimento	m				x	x
7.305.7a	Cepillado de la Superficie Pavimento de Hormigón	m2					x
7.305.7b	Cepillado de la Superficie Pavimento Asfáltico	m2					x
7.305.8	Nivelación de Bermas Granuladas no Revestidas en Pavimento de Hormigón	m2				x	x
7.305.9a	Nivelación de Bermas en Pavimentos de Hormigón Revestidas con un Tratamiento Superficial Simple	m2				x	x
7.305.9b	Nivelación de Bermas en Pavimentos de Hormigón Revestidas con un Tratamiento Superficial Doble	m2				x	x
7.305.9c	Nivelación de Bermas en Pavimentos de Hormigón Pavimentadas con Mezcla Asfáltica en Caliente	m2				x	x
7.305.9d	Nivelación de Bermas en Pavimentos de Hormigón Pavimentadas con Mezcla Asfáltica en Frío	m2				x	x
7.305.10	Reemplazo de Losas por Mezclas Asfálticas	m2	x			x	x
7.305.11	Bacheo Superficial Asfáltico en Frío en Pavimentos de Hormigón	m2	x			x	x
7.305.12	Reposición de Losas de Hormigón	m2		x	x	x	x
7.305.13a	Reperfilado Simple de Bermas	km					x
7.305.13b	Reperfilado Simple con Riego de Bermas	km					x
7.305.13c	Reperfilado con Compactación de Bermas	km					x
7.305.14	Colocación de Barras de Traspaso de Carga	N°		x	x		x
7.306.1a	Reperfilado Simple	km					x
7.306.1b	Reperfilado Simple con Riego	km					x
7.306.1c	Reperfilado con Compactación	km					x
7.306.1d	Reperfilado con Compactación con Riego que Incorpora Estabilizador	km					x
7.306.1e	Reperfilado con Compactación de Calzada con Adición de Material Plástico	km					x
7.306.2a	Reparación de Areas Inestables	m3				x	x
7.306.2b	Tela Geotextil para Refuerzo de la Subrasante	m2					x
7.306.3a	Bacheo de Capas de Rodadura Granulares	m3				x	x
7.306.3b	Bacheo de Capas de Rodadura Granulares con Adición de Estabilizador	m3				x	x
7.306.4a1	Recebo de Capas de Rodadura Granulares Tamaño 1 1/2"	m3				x	x
7.306.4a2	Recebo de Capas de Rodadura Granulares Tamaño 3"	m3				x	x
7.306.4b	Recebo de Capas de Rodadura Granulares con Material Chancado	m3				x	x
7.306.4c	Recebo de Capas de Rodadura Granulares con Adición de Suelo Plástico	m3				x	x
7.306.4d	Recebo de Capas de Rodadura con Maicillo	m3				x	x
7.306.4e	Recebo de Capas de Rodadura Granulares con Provisión Externa de Material	m3				x	x
7.306.4f	Recebo de Capas de Rodadura Granulares con Adición de Estabilizador	m3				x	x
7.306.5	Reparación de la Calzada con Material Integral	m3				x	x

Exponential robust fit summary table for 7.304 activities

Code	Activity	Unit	N° of records	Regression fitness	% of outliers
7.304.1a	Sellado de Areas con Grietas de hasta 6 mm de ancho	m2	19	Good fit	0.0%
7.304.1b	Sellado de Grietas de Ancho entre 6 y 20 mm	m	131	Good fit	15.3%
7.304.1c	Sellado de Grietas de Ancho entre 20 y 70 mm	m	41	Good fit	4.9%
7.304.1d	Sellado de Grietas y Cavidades de Ancho superior a 70 mm	m	7	Good fit	14.3%
7.304.2a	Bacheo Superficial Manual con Mezclas en Caliente	m2	890	Good fit	3.0%
7.304.2b	Bacheo Superficial Manual con Mezcla en Frío	m2	482	Good fit	13.5%
7.304.2c	Bacheo Superficial Manual con Mezclas Preparadas	m2	652	Good fit	30.1%
7.304.2d	Bacheo Superficial Mecanizado	m2	13	Good fit	7.7%
7.304.2e	Bacheo con Tratamiento Superficial Doble	m2	242	Good fit	20.0%
7.304.2f	Parche Superficial	m2	40	Good fit	5.0%
7.304.3a	Bacheo Profundo con Mezclas en Caliente	m2	515	Lack of fit	
7.304.3b	Bacheo Profundo con Mezclas en Frío	m2	68	Good fit	11.1%
7.304.3c	Bacheo Profundo con Tratamiento Superficial	m2	145	Good fit	13.3%
7.304.3d	Bacheo Profundo con Mezclas en Frío Predosificadas	m2	11	Good fit	36.4%
7.304.4a	Sello Tipo Riego de Neblina	m2	13	Good fit	15.4%
7.304.4b	Sello Tipo Tratamiento Superficial Simple	m2	259	Good fit	38.9%
7.304.4c	Sello Tipo Lechada Asfáltica	m2	344	Good fit	4.9%
7.304.4d	Sello Localizado con Gravilla	m2	20	Good fit	20.0%
7.304.4e	Sello Localizado con Lechada	m2	42	Good fit	4.8%
7.304.5	Nivelación de Bermas Granulares no Revestidas en Pavimentos Asfálticos	m2	34	Good fit	11.8%
7.304.6a	Nivelación de Bermas en Pavimentos Asfálticos Revestidas con un Tratamiento Superficial Simple	m2	17	Good fit	0.0%
7.304.6b	Nivelación de Bermas en Pavimentos Asfálticos Revestidas con un Tratamiento Superficial Doble	m2	9	Good fit	33.3%
7.304.6c	Nivelación de Bermas en Pavimentos Asfálticos Pavimentadas con Mezcla Asfáltica en Caliente	m2	28	Good fit	7.1%
7.304.6d	Nivelación de Bermas en Pavimentos Asfálticos Pavimentadas con Mezcla Asfáltica en Frío	m2	5	Lack of fit	
7.304.7	Imprimación Reforzada	m2	222	Good fit	0.5%
7.304.8	Reposición de Capa de Rodadura de concreto Asfáltico	m2	183	Good fit	2.2%
7.304.9a	Reperfilado Simple de Bermas	km	6	Good fit	0.0%
7.304.9b	Reperfilado Simple con Riego de Bermas	km	0	Lack of fit	
7.304.9c	Reperfilado con Compactación	km	4	Lack of fit	

Exponential robust fit summary table for 7.305 activities

Code	Activity	Unit	N° of records	Regression fitness	% of outliers
7.305.1a	Sellado de Juntas y Grietas	km	112	Lack of fit	
7.305.1b	Sellado de Juntas y Grietas	m	52	Good fit	9.8%
7.305.2	Reparación en Todo el Espesor	m2	59	Good fit	11.6%
7.305.3	Reparación en Todo el Espesor para Puesta en Servicio Acelerada	m2	19	Good fit	21.1%
7.305.4a	Reparación de Espesor Parcial	m2	0	Lack of fit	
7.305.4b	Reparación de Espesor Parcial para Puesta en Servicio Acelerada	m2	5	Lack of fit	
7.305.5	Pavimentos Insertados	m2	33	Good fit	0.0%
7.305.6	Instalación de Drenes de Pavimento	m	7	Lack of fit	
7.305.7a	Cepillado de la Superficie Pavimento de Hormigón	m2	18	Lack of fit	
7.305.7b	Cepillado de la Superficie Pavimento Asfáltico	m2	5	Lack of fit	
7.305.8	Nivelación de Bermas Granuladas no Revestidas en Pavimento de Hormigón	m2	15	Good fit	21.1%
7.305.9a	Nivelación de Bermas en Pavimentos de Hormigón Revestidas con un Tratamiento Superficial Simple	m2	0	Lack of fit	
7.305.9b	Nivelación de Bermas en Pavimentos de Hormigón Revestidas con un Tratamiento Superficial Doble	m2	2	Lack of fit	
7.305.9c	Nivelación de Bermas en Pavimentos de Hormigón Pavimentadas con Mezcla Asfáltica en Caliente	m2	4	Good fit	0.0%
7.305.9d	Nivelación de Bermas en Pavimentos de Hormigón Pavimentadas con Mezcla Asfáltica en Frío	m2	0	Lack of fit	
7.305.10	Reemplazo de Losas por Mezclas Asfálticas	m2	1	Lack of fit	
7.305.11	Bacheo Superficial Asfáltico en Frío en Pavimentos de Hormigón	m2	13	Lack of fit	
7.305.12	Reposición de Losas de Hormigón	m2	155	Good fit	13.3%
7.305.13a	Reperfilado Simple de Bermas	km	0	Lack of fit	
7.305.13b	Reperfilado Simple con Riego de Bermas	km	0	Lack of fit	
7.305.13c	Reperfilado con Compactación de Bermas	km	0	Lack of fit	
7.305.14	Colocación de Barras de Traspaso de Carga	N°	4	Lack of fit	

Exponential robust fit summary table for 7.306 activities

Code	Activity	Unit	N° of records	Regression fitness	% of outliers
7.306.1a	Reperfilado Simple	km	11584	Good fit	12.6%
7.306.1b	Reperfilado Simple con Riego	km	116	Lack of fit	
7.306.1c	Reperfilado con Compactación	km	494	Good fit	12.2%
7.306.1d	Reperfilado con Compactación con Riego que Incorpora Estabilizador	km	122	Good fit	21.3%
7.306.1e	Reperfilado con Compactación de Calzada con Adición de Material Plástico	km	20	Good fit	5.0%
7.306.2a	Reparación de Areas Inestables	m3	5	Lack of fit	
7.306.2b	Tela Geotextil para Refuerzo de la Subrasante	m2	43	Good fit	0.0%
7.306.3a	Bacheo de Capas de Rodadura Granulares	m3	965	Good fit	22.1%
7.306.3b	Bacheo de Capas de Rodadura Granulares con Adición de Estabilizador	m3	1	Lack of fit	
7.306.4a1	Recebo de Capas de Rodadura Granulares Tamaño 1 1/2"	m3	22	Good fit	0.0%
7.306.4a2	Recebo de Capas de Rodadura Granulares Tamaño 3"	m3	17	Lack of fit	
7.306.4b	Recebo de Capas de Rodadura Granulares con Material Chancado	m3	601	Lack of fit	
7.306.4c	Recebo de Capas de Rodadura Granulares con Adición de Suelo Plástico	m3	16	Good fit	0.0%
7.306.4d	Recebo de Capas de Rodadura con Maicillo	m3	79	Good fit	39.2%
7.306.4e	Recebo de Capas de Rodadura Granulares con Provisión Externa de Material	m3	14	Good fit	14.3%
7.306.4f	Recebo de Capas de Rodadura Granulares con Adición de Estabilizador	m3	77	Good fit	18.2%
7.306.5	Reparación de la Calzada con Material Integral	m3	226	Good fit	9.7%

Price factors for Chilean provinces

ID	Provincia	Región	Factor
1	Arica	Arica y Parinacota	1.194
2	Parinacota	Arica y Parinacota	0.761
3	Iquique	Tarapacá	0.726
4	Tamarugal	Tarapacá	0.818
5	Antofagasta	Antofagasta	1.082
6	El Loa	Antofagasta	1.175
7	Tocopilla	Antofagasta	1.841
8	Copiapó	Atacama	1.306
9	Chañaral	Atacama	1.861
10	Huasco	Atacama	0.671
11	Elqui	Coquimbo	0.771
12	Choapa	Coquimbo	0.884
13	Limarí	Coquimbo	0.788
14	Valparaíso	Valparaíso	0.777
15	Isla de Pascua	Valparaíso	
16	Los Andes	Valparaíso	0.783
17	Petorca	Valparaíso	0.991
18	Quillota	Valparaíso	0.708
19	San Antonio	Valparaíso	1.657
20	San Felipe	Valparaíso	1.070
21	Marga Marga	Valparaíso	0.744
22	Cachapoal	Lib. Gral. Bernardo O'Higgins	0.994
23	Cardenal Caro	Lib. Gral. Bernardo O'Higgins	1.077
24	Colchagua	Lib. Gral. Bernardo O'Higgins	0.924
25	Talca	Maule	0.925
26	Cauquenes	Maule	1.024
27	Curicó	Maule	1.052
28	Linares	Maule	0.897
29	Concepción	Biobío	1.417
30	Arauco	Biobío	2.415
31	Biobío	Biobío	1.739
32	Ñuble	Biobío	1.601
33	Cautín	La Araucanía	0.750
34	Malleco	La Araucanía	0.590
35	Valdivia	Los Ríos	0.959
36	Ranco	Los Ríos	1.058
37	Llanquihue	Los Lagos	0.920
38	Chiloé	Los Lagos	1.244
39	Osorno	Los Lagos	1.048
40	Palena	Los Lagos	1.685
41	Coihaique	Aysén del Gral. C. Ibáñez del Campo	1.802
42	Aysén	Aysén del Gral. C. Ibáñez del Campo	1.896
43	Capitán Prat	Aysén del Gral. C. Ibáñez del Campo	1.554
44	General Carrera	Aysén del Gral. C. Ibáñez del Campo	0.962
45	Magallanes	Magallanes y Antártica Chilena	1.153
46	Antártica Chilena	Magallanes y Antártica Chilena	
47	Tierra del Fuego	Magallanes y Antártica Chilena	1.120
48	Última Esperanza	Magallanes y Antártica Chilena	1.601
49	Santiago	Metropolitana de Santiago	1.017
50	Cordillera	Metropolitana de Santiago	0.821
51	Chacabuco	Metropolitana de Santiago	1.082
52	Maipo	Metropolitana de Santiago	1.128
53	Melipilla	Metropolitana de Santiago	1.021
54	Talagante	Metropolitana de Santiago	0.675

APPENDIX B – MATLAB scripts

Robustfit.m

```
%ROBUSTFIT Robust linear regression
%   B = ROBUSTFIT(X,Y) returns the vector B of regression coefficients,
%   obtained by performing robust regression to estimate the linear model
%   Y = Xb.  X is an n-by-p matrix of predictor variables, and Y is an
%   n-by-1 vector of observations.  The algorithm uses iteratively
%   reweighted least squares with the bisquare weighting function.  By
%   default, ROBUSTFIT adds a column of ones to X, corresponding to a
%   constant term in the first element of B.  Do not enter a column of ones
%   directly into the X matrix.
%
%   B = ROBUSTFIT(X,Y,'WFUN',TUNE) uses the weighting function 'WFUN' and
%   tuning constant TUNE.  'WFUN' can be any of 'andrews', 'bisquare',
%   'cauchy', 'fair', 'huber', 'logistic', 'talwar', or 'welsch'.
%   Alternatively 'WFUN' can be a function that takes a residual vector as
%   input and produces a weight vector as output.  The residuals are scaled
%   by the tuning constant and by an estimate of the error standard
%   deviation before the weight function is called.  'WFUN' can be
%   specified using @ (as in @myfun).  TUNE is a tuning constant that is
%   divided into the residual vector before computing the weights, and it
%   is required if 'WFUN' is specified as a function.
%
%   B = ROBUSTFIT(X,Y,'WFUN',TUNE,'CONST') controls whether or not the
%   model will include a constant term.  'CONST' is 'on' (the default) to
%   include the constant term, or 'off' to omit it.
%
%   [B,STATS] = ROBUSTFIT(...) also returns a STATS structure
%   containing the following fields:
%       'ols_s'     sigma estimate (rmse) from least squares fit
%       'robust_s'  robust estimate of sigma
%       'mad_s'     MAD estimate of sigma; used for scaling
%                   residuals during the iterative fitting
%       's'         final estimate of sigma, the larger of robust_s
%                   and a weighted average of ols_s and robust_s
%       'se'        standard error of coefficient estimates
%       't'         ratio of b to stats.se
%       'p'         p-values for stats.t
%       'covb'      estimated covariance matrix for coefficient estimates
%       'coeffcorr' estimated correlation of coefficient estimates
%       'w'         vector of weights for robust fit
%       'h'         vector of leverage values for least squares fit
%       'dfe'       degrees of freedom for error
%       'R'         R factor in QR decomposition of X matrix
%
%   The ROBUSTFIT function estimates the variance-covariance matrix of the
%   coefficient estimates as V=inv(X'*X)*STATS.S^2.  The standard errors
%   and correlations are derived from V.
%
%   ROBUSTFIT treats NaNs in X or Y as missing values, and removes them.
%
%   Example:
%      x = (1:10)';
%      y = 10 - 2*x + randn(10,1); y(10) = 0;
%      bls = regress(y,[ones(10,1) x])
%      brob = robustfit(x,y)
%      scatter(x,y)
%      hold on
%      plot(x,brob(1)+brob(2)*x,'r-', x,bls(1)+bls(2)*x,'m:')
%
%   See also REGRESS, ROBUSTDEMO.

% References:
```

```
%   DuMouchel, W.H., and F.L. O'Brien (1989), "Integrating a robust
%     option into a multiple regression computing environment,"
%     Computer Science and Statistics:  Proceedings of the 21st
%     Symposium on the Interface, American Statistical Association.
%   Holland, P.W., and R.E. Welsch (1977), "Robust regression using
%     iteratively reweighted least-squares," Communications in
%     Statistics - Theory and Methods, v. A6, pp. 813-827.
%   Huber, P.J. (1981), Robust Statistics, New York: Wiley.
%   Street, J.O., R.J. Carroll, and D. Ruppert (1988), "A note on
%     computing robust regression estimates via iteratively
%     reweighted least squares," The American Statistician, v. 42,
%     pp. 152-154.

%   Copyright 1993-2006 The MathWorks, Inc.
%   $Revision: 1.4.2.5 $  $Date: 2006/11/11 22:55:45 $
```

Vartest2.m

```
%VARTEST2 Two-sample F test for equal variances.
%   H = VARTEST2(X,Y) performs an F test of the hypothesis that two
%   independent samples, in the vectors X and Y, come from normal
%   distributions with the same variance, against the alternative that
%   they come from normal distributions with different variances.
%   The result is H=0 if the null hypothesis ("variances are equal")
%   cannot be rejected at the 5% significance level, or H=1 if the null
%   hypothesis can be rejected at the 5% level.  X and Y can have
%   different lengths.
%
%   X and Y can also be matrices or N-D arrays.  For matrices, VARTEST2
%   performs separate tests along each column, and returns a vector of
%   results.  X and Y must have the same number of columns.  For N-D
%   arrays, VARTEST2 works along the first non-singleton dimension.  X and Y
%   must have the same size along all the remaining dimensions.
%
%   VARTEST2 treats NaNs as missing values, and ignores them.
%
%   H = VARTEST2(X,Y,ALPHA) performs the test at the significance level
%   (100*ALPHA)%.  ALPHA must be a scalar.
%
%   H = VARTEST2(X,Y,ALPHA,TAIL) performs the test against the alternative
%   hypothesis specified by TAIL:
%       'both'  -- "variances are not equal" (two-tailed test)
%       'right' -- "variance of X is greater than variance of Y" (right-
%                  tailed test)
%       'left'  -- "variance of X is less than variance of Y" (left-
%                  tailed test)
%   TAIL must be a single string.
%
%   [H,P] = VARTEST2(...) returns the p-value, i.e., the probability of
%   observing the given result, or one more extreme, by chance if the null
%   hypothesis is true.  Small values of P cast doubt on the validity of
%   the null hypothesis.
%
%   [H,P,CI] = VARTEST2(...) returns a 100*(1-ALPHA)% confidence interval for
%   the true ratio var(X)/var(Y).
%
%   [H,P,CI,STATS] = VARTEST2(...) returns a structure with the following
%   fields:
%      'fstat'  -- the value of the test statistic
%      'df1'    -- the numerator degrees of freedom of the test
%      'df2'    -- the denominator degrees of freedom of the test
%
%   [...] = VARTEST2(X,Y,ALPHA,TAIL,TESTTYPE,DIM) works along dimension DIM
%   of X and Y.
%
%   Example:  Is the variance significantly different for two model years,
```

```
%             and what is a confidence interval for the ratio of these
%             variances?
%      load carsmall
%      [h,p,ci] = vartest2(MPG(Model_Year==82),MPG(Model_Year==76))
%
%   See also ANSARIBRADLEY, VARTEST, VARTESTN, TTEST2.

%   Copyright 2005-2009 The MathWorks, Inc.
%   $Revision: 1.1.6.6 $  $Date: 2009/11/05 17:03:46 $
```

Lofrtest.m

```
% LOFRTEST Lack-of-fit test for regression model with independent replicate values.
%  LOFRTEST(D,alpha) is a statistical test that gives information on the form
%  of the model under consideration. A significant lack-of-fit suggest that there
%  may be some systematic variation unaccounted for in the hypothesized model
%  (chosen model does not well describe the data). It arises when there are exact
%  replicate values of the independent variable in the model that provide an estimate
%  of pure error. Pure error is in essence the amount of error that cannot be
accounted
%  for by any model. Then allows a test on whether there is error present aside
%  from pure error. For the construction of the lack-of-fit test we need to examine
%  three common types of linear models:
%      - single mean (one parameter)
%      - slope and intercept or common regression model (two parameters)
%      - separate means for each x-value or one-way ANOVA (many parameters).
%  So, the pure error is the error of the separate means on ANOVA and the total error
%  in the residual resulting in the regression analysis: the lack-of-fit results
%  to be the difference between this two sources of error,
%
%                  SS(LOF) = SSR(Model) - SSE(ANOVA).
%
%  Syntax: lofrtest(D,alpha)
%
%  Inputs:
%       D - matrix data (=[X Y]) (last column must be the Y-dependent variable).
%           (X-independent variable entry can be for a simple [X], multiple
[X1,X2,X3,...Xp]
%           or polynomial [X,X^2,X^3,...,X^p] regression model).
%   alpha - significance level (default = 0.05).
%
%  Outputs:
%       A complete summary (table) of analysis of variance partitioning sources of
%       variation for testing lack-of-fit.
%
%  Created by A. Trujillo-Ortiz, R. Hernandez-Walls, A. Castro-Perez and
%             F.J. Marquez-Rocha
%             Facultad de Ciencias Marinas
%             Universidad Autonoma de Baja California
%             Apdo. Postal 453
%             Ensenada, Baja California
%             Mexico.
%             atrujo@uabc.mx
%  Copyright (C) March 4, 2005.
%
%  To cite this file, this would be an appropriate format:
%  Trujillo-Ortiz, A., R. Hernandez-Walls, A. Castro-Perez and F.J Marquez-Rocha.
%    (2005). lofrtest:Lack-of-fit test for regression model with independent replicate
%    values. A MATLAB file. [WWW document]. URL
http://www.mathworks.com/matlabcentral/
%    fileexchange/loadFile.do?objectId=7074
%
%  References:
%
%  Department of Psychology of the York University. Available on Internet at the
%          URL address http://www.psych.yorku.ca/lab/psy3030/assign/assign3.htm
```

```
%  Zar, J. H. (1999), Biostatistical Analysis (2nd ed.).
%          NJ: Prentice-Hall, Englewood Cliffs. p. 345-350.
```

Main routine

```
% ROAD MAINTENANCE UNIT COSTS ESTIMATING THROUGH AN EXPONENTIAL ROBUST REGRESSION
MODEL. Cristóbal Moena Madrid.
clear all
close all
clc

%% CASOS
buenos = 1:68; % Todas las ADC
nom_base = {'B1Z0','B1Z1','B1Z2','B1Z3','B1Z4';
    'B2Z0','B2Z1','B2Z2','B2Z3','B2Z4'};

%% Procesamiento de datos
% Se carga la lista de códigos de Acciones de Conservación
[codigos_n codigos_t codigos_r] = xlsread('Acciones de conservación.xlsx','Manual de
carreteras');
codigos = codigos_t(2:69,2);
nombres = codigos_t(2:69,3);
N = length(codigos);

% Se carga la base de datos "Contratos de Conservación"
[datos_b1 datos_t_b1 datos_r] = xlsread('BASE1 v3.0.xlsx','Base');
datos_t_b1 = datos_t_b1(2:end,:); % Se cargan los códigos de cada ADC
N_ADC_B1 = zeros(size(codigos));
for i = 1:N,
    N_ADC_B1(i) = size(find(strcmp(datos_t_b1(:,14),codigos(i))==1),1);
end

% Se carga la base de datos "Costos de administración directa"
[datos_b2 datos_t_b2 datos_r] = xlsread('Costos Administración Directa 2007-
2010.xlsx','Todos');
datos_t_b2 = datos_t_b2(2:end,:); % Se cargan los códigos de cada ADC
N_ADC_B2 = zeros(size(codigos));
for i = 1:N,
    N_ADC_B2(i) = size(find(strcmp(datos_t_b2(:,5),codigos(i))==1),1);
end

% Se cargan los indicadores económicos. El índice 1 corresponde a Enero de 1987
indicadores = xlsread('Indicadores.xls','Indices');
asfalto = indicadores(:,3);
cemento = indicadores(:,4);
diesel = indicadores(:,5);
arena = indicadores(:,6);
acero = indicadores(:,7);
uf = indicadores(:,8);

VRVZ0 = cell(N,2); % Para validación reducción variabilidad geográfica
CONTZ0 = cell(N,2,4); % Para validación disminución variabilidad contingencia
PROMS = zeros(N,5); % Almacena los promedios por zona

for base = 1,
    for zona = 0,

        MODELO = zeros(N,16);
        MODELO2 = zeros(N,16);
        COMPOS = zeros(N,10);
        for ii = 1:length(buenos)
            i = buenos(ii);

            if base==1
```

```
                MODELO(:,1) = N_ADC_B1;
                % Se obtienen los datos de PU y Volumen
                D = datos_b1(find(strcmp(datos_t_b1(:,14),codigos(i))==1),:);

                if zona == 0 % Filtro todas las regiones
                    DD = D(:,[18,35]);
                    DDt = D(:,[5,6]);
                elseif zona == 1 % Filtro por región Zona 1
                    DD = D([find(D(:,1)==1)' find(D(:,1)==2)'
find(D(:,1)==3)'],[18,35]);
                    DDt = D([find(D(:,1)==1)' find(D(:,1)==2)'
find(D(:,1)==3)'],[5,6]);
                elseif zona == 2 % Filtro por región Zona 2
                    DD = D([find(D(:,1)==4)' find(D(:,1)==5)' find(D(:,1)==6)'
find(D(:,1)==7)' find(D(:,1)==13)'],[18,35]);
                    DDt = D([find(D(:,1)==4)' find(D(:,1)==5)' find(D(:,1)==6)'
find(D(:,1)==7)' find(D(:,1)==13)'],[5,6]);
                elseif zona == 3 % Filtro por región Zona 3
                    DD = D([find(D(:,1)==8)' find(D(:,1)==9)'
find(D(:,1)==10)'],[18,35]);
                    DDt = D([find(D(:,1)==8)' find(D(:,1)==9)'
find(D(:,1)==10)'],[5,6]);
                elseif zona == 4 % Filtro por región Zona 4
                    DD = D([find(D(:,1)==11)' find(D(:,1)==12)'],[18,35]);
                    DDt = D([find(D(:,1)==11)' find(D(:,1)==12)'],[5,6]);
                end

            elseif base==2
                MODELO(:,1) = N_ADC_B2;
                % Se obtienen los datos de PU y Volumen
                D = datos_b2(find(strcmp(datos_t_b2(:,5),codigos(i))==1),:);
                if zona == 0 % Filtro todas las regiones
                    DD = D(:,[8,18]);
                    DD0 = DD; % Validación reducción variabilidad por zona
                    DDt = D(:,[1,2]);
                    DDc = D(:,9:14)./(D(:,8)*ones(1,6));

                    % Composición de precios
                    COMPOS(i,1:5) = [nanmean(DDc(:,1)) nanmean(DDc(:,2))
nanmean(DDc(:,3)) nanmean(DDc(:,4)) nanmean(DDc(:,5))];
                    COMPOS(i,1:5) = COMPOS(i,1:5)/sum(COMPOS(i,1:5));

                    % Factor de variación
                    COMPOS(i,6:10) = [nanstd(DDc(:,1))/nanmean(DDc(:,1))
nanstd(DDc(:,2))/nanmean(DDc(:,2)) nanstd(DDc(:,3))/nanmean(DDc(:,3))
nanstd(DDc(:,4))/nanmean(DDc(:,4)) nanstd(DDc(:,5))/nanmean(DDc(:,5))];

                elseif zona == 1 % Filtro por región Zona 1
                    DD = D([find(D(:,1)==1)' find(D(:,1)==2)'
find(D(:,1)==3)'],[8,18]);
                    DDt = D([find(D(:,1)==1)' find(D(:,1)==2)'
find(D(:,1)==3)'],[1,2]);
                elseif zona == 2 % Filtro por región Zona 2
                    DD = D([find(D(:,1)==4)' find(D(:,1)==5)' find(D(:,1)==6)'
find(D(:,1)==7)' find(D(:,1)==13)'],[8,18]);
                    DDt = D([find(D(:,1)==4)' find(D(:,1)==5)' find(D(:,1)==6)'
find(D(:,1)==7)' find(D(:,1)==13)'],[1,2]);
                elseif zona == 3 % Filtro por región Zona 3
                    DD = D([find(D(:,1)==8)' find(D(:,1)==9)'
find(D(:,1)==10)'],[8,18]);
                    DDt = D([find(D(:,1)==8)' find(D(:,1)==9)'
find(D(:,1)==10)'],[1,2]);
                elseif zona == 4 % Filtro por región Zona 4
                    DD = D([find(D(:,1)==11)' find(D(:,1)==12)'],[8,18]);
                    DDt = D([find(D(:,1)==11)' find(D(:,1)==12)'],[1,2]);
                end
```

```
            end


            % --------
            % REGRESIÓN
            % Asignación de datos
            magnitud = DD(:,1);
            pu = DD(:,2);
            mu = nanmean(pu);
            sigma = nanstd(pu);

            PROMS(i,zona+1) = mu;

            % Validación "reducción variabilidad por zona" y "contingencia"
            if zona == 0,
                VRVZ0(i,base) = {pu};
                CONTZ0(i,base,1) = {pu};
            end

            % Transformación logarítmica de los datos
            log_DD = log(DD);
            log_magnitud = log_DD(:,1);
            log_pu = log_DD(:,2);

            % Regresión robusta y exponencial
            try
                [b stats] = robustfit(log_magnitud,log_pu);
                ind_w = find(stats.w > min_weigth);
                % Se valida estadísticamente el modelo
                [base zona i]
                validador(log_magnitud,log_pu,b,stats,1)
            catch
                b = [0;0];
                stats = 0;
                ind_w = 1:length(pu);
            end
            pu_levg = pu(ind_w);
            magnitud_levg = magnitud(ind_w);



            magnitud_reg = [0.01*min(magnitud):0.001*(max(magnitud)-
min(magnitud)):max(magnitud)*1.5]';
            log_magnitud_reg = log(magnitud_reg);
            log_pu_reg = [ones(length(log_magnitud_reg),1) log_magnitud_reg]*b;

            pu_reg = zeros(size(magnitud_reg));
            for j = 1:length(magnitud_reg),
                pu_reg(j,1) = exp(log_pu_reg(j)); % Se transforma a espacio real
            end
            % --------

            % --------
            % ACTUALIZACIÓN
            ano = DDt(ind_w,2);
            mes = DDt(ind_w,1);
            t = (ano - 1987)*12 + mes;

            t(find(t>301)) = ones(length(find(t>301)))*301;

% Índices de precios de la regresión
            MODELO(i,14:19) = [nanmean(asfalto(t)) nanmean(cemento(t))
nanmean(diesel(t)) nanmean(arena(t)) nanmean(acero(t)) nanmean(uf(t))];
            MODELO2(i,14:19) = [mean(asfalto(t)) mean(cemento(t)) mean(diesel(t))
mean(arena(t)) mean(acero(t)) mean(uf(t))];
```

```
            % --------


            % --------
            % CONTINGENCIA: División por rangos
            N_div = 3;
            rango = (max(magnitud_levg) - min(magnitud_levg))/N_div;

            for j = 1:N_div,
                lim_inf = min(magnitud_levg) + rango*(j-1);
                lim_sup = min(magnitud_levg) + rango*j;

                indx = find(magnitud_levg>lim_inf);
                indx2 = find(magnitud_levg(indx)<lim_sup);
                indx = indx(indx2);

                pu_rango = pu_levg(indx);
                magnitud_rango = magnitud_levg(indx);
                mu_rango = nanmean(pu_rango);
                sigma_rango = nanstd(pu_rango);

                % Validación de "contingencia"
                if zona == 0,
                    CONTZ0(i,base,j+1) = {pu_rango};
                end

                % Se calcula el porcentaje de desviación de PU de cada rango
                MODELO(i,j+5) = sigma_rango/mu_rango;

                % Se calcula el promedio de Magnitud de cada rango
                if isempty(magnitud_levg)
                else
                    MODELO(i,8+j) = nanmean(magnitud_rango);
                end
            end
            % --------

            % --------
            % Gráfico de ajuste
            if 0
                figure
                plot(magnitud,pu,'*')
                hold on
                plot(magnitud_reg,pu_reg,'r')
                title(['Unit price vs. Contracted Volume. Code = ',codigos(i)])
                xlabel('Contracted Volume (un)')
                ylabel('Unit price (UF)')
                legend('Real data','Exponentially robust fitted curve')
            end
            % --------

            % Se graban los parámetros "b", "mu" de PU general, y "sigma" de PU
general.
            MODELO(i,[2 3]) = b;
            MODELO(i,[4 5]) = [mu sigma];
            if isempty(magnitud_levg)
            else
                MODELO(i,[12 13]) = [nanmean(magnitud_levg) nanstd(magnitud_levg)];
            end
        end


        % Escribir los parámetros del modelo
```

```
                xlswrite('basevialV18.xlsx',MODELO(:,1:13),nom_base{base,zona+1},'D2:P6
                9');

                xlswrite('basevialV18.xlsx',MODELO(:,14:19),nom_base{base,zona+1},'AA2:
                AF69');

                xlswrite('basevialV18.xlsx',MODELO2(:,14:19),nom_base{base,zona+1},'AA2
                :AF69');
    end
end
```

Cross validation

```
clear all
close all
clc

% Se cargan los datos
D = xlsread('Accion.xlsx','7.302.5d');

% Se define la muestra de estimación y la muestra de test
n = size(D,1);
RR = rand(n,1);
indr = find(RR<0.75);
indrt = find(RR>=0.75);
nr = size(indr,1);

PUtest = D(indrt,4);
Vtest = D(indrt,1);
PU = D(indr,4);
V = D(indr,1);

% Transformación logarítmica de los datos
log_PUtest = log(PUtest);
log_Vtest = log(Vtest);
log_PU = log(PU);
log_V = log(V);

% Regresión robusta
[b stats] = robustfit(log_V,log_PU);

% Curva regresionada
V_reg = [0.0000001*min(V):0.001*(max(V)-min(V)):max(V)*1.5]';
log_V_reg = log(V_reg);
log_PU_reg = [ones(length(log_V_reg),1) log_V_reg]*b;
PU_reg = zeros(size(V_reg));
for j = 1:length(V_reg),
    PU_reg(j,1) = exp(log_PU_reg(j));
end

% Curva de test 75%
V_reg_test = Vtest;
log_V_reg_test = log(V_reg_test);
log_PU_reg_test = [ones(length(log_V_reg_test),1) log_V_reg_test]*b;
PU_reg_test = zeros(size(V_reg_test));
for j = 1:length(V_reg_test),
    PU_reg_test(j,1) = exp(log_PU_reg_test(j));
end

% Gráfico de ajuste
if 1
    figure
    plot(V,PU,'*')
    hold on
```

```
    plot(V_reg,PU_reg,'r')
    title(['Unit price vs. Contracted Volume. Code = 7.302.5d'])
    xlabel('Contracted Volume (un)')
    ylabel('Unit price (UF)')
    legend('Real data','Exponentially robust fitted curve')

    figure
    plot(Vtest,PUtest,'*')
    hold on
    plot(V_reg,PU_reg,'r')
    title(['Unit price vs. Contracted Volume. Code = 7.302.5d'])
    xlabel('Contracted Volume (un)')
    ylabel('Unit price (UF)')
    legend('Real data','Exponentially robust fitted curve')
end

% Residuales
resid_est = stats.resid;
resid_test = log_PU_reg_test - log_PUtest;

MSE = sum(resid_est.^2)/(length(resid_est)-2)
MSEt = sum(resid_test.^2)/(length(resid_test)-2)
F0 = MSEt/MSE
finv(0.025,(length(resid_test)-2),(length(resid_est)-2))
finv(0.975,(length(resid_test)-2),(length(resid_est)-2))

[H,P,CI,STATS]= vartest2(resid_test.^2,resid_est.^2,0.05,'right')
```

Collinearity

```
clear all
close all
clc

load contz0
H = 2*ones(68,4,3);
PROM = zeros(68,4,3);
for i = 1:68, % 68 actividades
    for j = 1:4, % 4 zonas geográficas: Z1, Z2, Z3 y Z4
        for k = 1:3, % 3 rangos de volumen: R1, R2, y R3
            Y = CONTZ0{i,1,1};
            X = CONTZ0{i,j+1,k+1};
            if(isempty(X) || isempty(Y))
            else
                h = ttest2(X,Y); % Test-t de comparación de medias de muestras
distintas
                prom = nanmean(X);
            end

            H(i,j+1,k+1) = h; % h=0-> fail to reject null hypothesis; h=1-> null
hypothesis rejection
            PROM(i,j+1,k+1) = prom;
        end
    end
end

% Calculo de porcentaje de actividades sin colinearidad
SUMA = zeros(4,3);
for j = 1:4,
    for k = 1:3,
        SUMA(j,k) = sum(H(:,j+1,k+1),1);
    end
end
1-SUMA/68
```