Project-Specific Cost Escalation Modeling: Crafting a Stochastic Tool for Predicting escalation

Modeling cost escalations

MSc. Construction Management and Engineering

Frenk van der Kraan

Delft University of Technology

Project-Specific Cost Escalation Modeling: Crafting a Stochastic Tool for Predicting escalation Modeling cost escalations

Graduation thesis

University	Delft University of Technology
Faculty	Civil Engineering and Geosciences (CiTG)
Curriculum	MSc Construction Management and Engineering

Author

Name	Frenk van der Kraan
Student number	5415306

Graduation committee

Chairman, TU Delft	Dr.ir. Oswaldo Morales Napoles
1st supervisor, TU Delft	Prof.dr.ir. P.H. van Gelder
2nd supervisor, TU Delft	Dr. Ir. O Kammouh
Company supervisor, Linesight	J. Phillips

In partial fulfilment of the Master of Science in **Construction Management and Engineering**.

17th January 2024

Acknowledgements

Writing this thesis proved to be a significant challenge and rewarding experience for me, as I worked on putting my ideas and knowledge into a final piece of work. Despite it being a time-consuming process, I am confident and hopeful that I have managed to articulate my thoughts as clearly as I perceived them.

I'm excited to present this final piece to my committee and extend my sincere thanks for their guidance and feedback. A thank you to Omar – our discussions on structuring and reporting my thesis were incredibly insightful. Each meeting seemed to help me improve my work in terms of clarity and organization. Pieter, your support in the mathematical aspects and shared enthusiasm were a great source of motivation – thank you. Oswaldo, your insights helped me in shaping my thesis and keeping it focused. And Jonathan, your support and timely reminders of the importance of my work were indispensable throughout this journey. Lastly, I would like to thank my family and friends for their unconditional support and encouragement during this period.

I hope you find the read enjoyable!

Frenk van der Kraan

17th January 2024

Abstract

Gaining insights into anticipated future expenses is essential for both the planning and construction phases of a project. A significant factor in future costs, particularly over extended periods, is the variability of material and labor prices. the change in costs over time, known as 'cost escalation,' has been the focus of numerous research efforts. These studies primarily aim to forecast the Construction Cost Index (CCI), a composite index representing a standardized array of materials and labor typical in construction projects.

This research presents a novel approach to forecast cost escalations, tailored to individual construction projects, addressing the shortcomings of predictions of a generic Construction Cost Index (CCI). Traditional CCI predictions, while providing some foresight, are limited by their generic nature and often overlook the specific material and labor variations within different projects. Additionally, current research generally fails to account for the uncertainty in the forecasts of the change in the material price and the estimate of the construction cost.

In response to these challenges, this research pivots around the following research question:

> **How can cost escalation for various types of construction projects be predicted, accounting for uncertainties in final construction costs and the forecast?**

Addressing this question involves selecting distinct indices for various project resources, such as steel, concrete, and labor, and combining these predicted indices in line with each resource's cost. To achieve this, the study evaluates several time-series forecasting models, namely the vector error correction model (VECM), the vector autoregression (VAR) model, and the Holt-Winters model. For each model an automatic forecasting process is created,which automatically checks the suitability of the model, select the appropriate variables (in case of a multivariate model) and selects parameters. These models are evaluated for their accuracy across different time-frames and forecasting horizons. The Holt-Winters model, in particular, showed promise in providing reliable confidence intervals and point forecasts.

The study's testing phase revealed varying degrees of success, as volatile indices like copper and steel proved to be challenging, whereas forecasts for less volatile indices achieved higher accuracy. This outcome suggests room for improvement in refining these forecasting methods. The project-specific forecasts, including cost uncertainties, are developed by inputting cost estimates associated to the material price at current day value. The cost estimates input includes material and labor costs and their uncertainties. The uncertainties are represented through a three

point estimate (optimistic, pessimistic, and most likely values). This input is then transformed into a PERT distribution which is a transformation of the Beta distribution. The final stage involves combining the PDFs of each project activities cost (considering the cost of each specific resource) with the monthly forecast PDFs in a Monte Carlo simulation, providing a detailed cost distribution histogram of total cost and the individual resources cost.

The tool's functionality was demonstrated through a case study on a highway construction project. In this demonstration, specific project data, including material and labor cost estimates, were inputted into a hybrid web and Excel interface. This setup facilitates visualization and manipulation of project information. The tool processes these inputs via the AFP and Monte Carlo simulation, yielding comprehensive outputs such as histograms and statistical properties of the output. This is visible for total project costs and resource-specific escalations. This demonstration effectively showcased the tool's capability to offer detailed insights into cost escalation, addressing the variability and uncertainty in construction projects. It underscored the tool's alignment with the study's objective of providing nuanced, project-specific cost escalation forecasts, moving beyond traditional CCI predictions.

In conclusion, this study introduces a tool that caters to specific project resources, timelines, and uncertainties in cost escalations. While current limitations prevent its immediate practical application, this proof-of-concept lays the groundwork for future improvements. Focus of future research should be on refining accuracy and comparing the tool's forecasts with escalation of historic projects to establish more robust insight into the actual escalation of projects.

Contents

List of Figures

List of Tables

Acronyms

ADF Augmented Dickey-Fuller. 17, 18

AFP Automated Forecasting Process. 7, 21, 32, 33, 41, 56, 59, 61

ARIMA Autoregressive integrated moving average. 2, 10

CCI Construction Cost Index. 1, 3

CV Coefficient of Variation. 52

LSTM Long Short-Term Memory. 2, 11

MAPE Mean Absolute Percentage Error. 10, 19, 38

MBB Moving Block Bootstrap. 31, 34, 56

PDF Probability Density Function. 4, 21, 41, 61

PI Prediction Interval. 30

PPI Produces Price Index. 22, 55, 61

STL Seasonal and Trend decomposition using Loess. 12, 56

US United States. 5, 22

VAR vector autoregression. iv, 11, 55, 61

VECM vector error correction model. iv, 11, 55, 61

Glossary

base cost The base cost represent the estimate of an acitivity or an object without accounting for escalation. . 2

cost escalation Cost escalation refers to the variation in the cost of materials, labor, or equipment within a specific economy over a designated period. This fluctuation is typically influenced by market conditions. This change could be expressed in a percentage change or in the change in specific monetary value, such as euros.. 1, 2

data leakage Data leakage refers to a situation in statistical modeling where information from outside the training dataset is inadvertently used to create the model. This typically occurs when data that would not be available at the time of prediction is used during training, leading to overly optimistic performance estimates. Data leakage can compromise the model's validity and its ability to generalize to new, unseen data. Preventing data leakage involves separation of training and testing data and ensuring that the model only uses information available at the time of prediction.. 57

distributional forecast A distributional forecast refers to a method of forecasting that provides a range of possible future values along with their associated probabilities, rather than a single-point prediction. This approach quantifies the uncertainty in predictions by presenting a spectrum of potential outcomes and their likelihoods. . 21, 30, 47, 57, 61

explained variable The explained variable, also known as the dependent or response variable, is the variable that one seeks to explain or predict. It is the primary focus of the model's output. . 11, 22

explanatory variable The explanatory variable is used to assist in predicting or explaining the behavior of the explained variable. It possesses predictive qualities that enhance the accuracy of forecasts for the explained variable. . 22

model class The model class categorizes a statistical model based on its structure, assumptions, and application. Examples of model classes in the context of the thesis are, the VAR model class, the VECM class and the Holt-Winters model class.. 21

model configurations Model configurations define the structure and components of a statistical model, including variable selection, lag determination, and data preparation. In VAR models, this involves choosing variables, setting lag lengths, and ensuring data stationarity.. 27

multivariate In this thesis, 'multivariate' pertains to a multivariate time-series model. This model incorporates multiple variables to enhance forecasting. It features both 'explained variables' and 'explanatory variables', leveraging the interplay between them for more nuanced predictions. . 8, 11

order of integration The order of integration of a time series indicates the number of times the series needs to be differenced to achieve stationarity. For instance, if a time series becomes stationary after being differenced once, it is said to have an order of integration of 1. . 9

resources In project management, "Resources" collectively refers to the elements required for task execution, encompassing:

1. **Man Hours:** The labor required, quantified in hours or days.
2. **Materials:** Tangible components such as raw materials and products.
3. **Equipment:** Tools and machinery necessary for the project.

This term covers the combined human, material, and equipment inputs needed for project success. . 1

stationarity Stationarity is a characteristic of a time series where its mean and standard deviation are constant over time. It's common practice to difference a time series to attain stationarity, in the case the time-series is non-stationary. . 8, 9

univariate In this thesis, 'univariate' refers to a univariate time-series model. Such a model utilizes a single variable, the 'explained variable', and employs its historical data to predict future values. This prediction is based on identifying patterns and trends from its past data. . 8, 11

Chapter 1

Introduction

Construction projects frequently face budget and schedule overruns (Flyvbjerg, 2003). This issue is partially driven by substantial price volatility and frequent design modifications (Tang, Han, feng Guo, Chen & Zhang, 2019), making it crucial for construction managers to anticipate future price changes (Jiang, Xu & Liu, 2013). Understanding prospective costs is vital during both planning and execution stages of a construction project (Mir, Kabir, Nasirzadeh & Khosravi, 2021; Ashuri, Asce & Lu, 2010; wei Xu & Moon, 2013). A component of these changes in price is due to market factors, where material and labor costs change over time. The change in price of a material or labor is often referred to as cost escalation .

To predict the cost escalation of a project, several researchers have developed models for forecasting a Construction Cost Index (CCI) (Moon & Shin, 2018; Ashuri et al., 2010; Dong, Chen & Guan, 2020; wei Xu & Moon, 2013; Choi, Ryu & Shahandashti, 2021). A CCI often consists of the average cost of man-hours and the materials steel, lumbar and cement (Grogan, 1992) over different suppliers. While these models provide some insight into cost escalation, limitations remain. Among which, not taking into account specific project resources by using a generic index. Additionally the incorporation of the uncertainty of both the prediction and the project cost itself is often neglected. This research aims to address these two limitations by developing a tool for forecasting escalation that allows for the consideration of the project specific resources and incorporating these uncertainties.

This chapter begins with an introduction to the study's background, then outlines the problem statement and research objective. It also defines the scope and presents the research questions and concludes with a reading guide.

1.1 Background

The forecasting of cost escalation is done by predicting an index that should represent the projects resources, which is often a generic index as the CCI. This prediction is used to calculate the percentage change over two time periods, indicating the escalation as a percentage. This percentage change can be multiplied by the cost estimate that uses the material prices at current price levels, often referred to as base cost (wei Xu & Moon, 2013; Touran, Asce & Lopez, 2006). This results in the expected escalation in some monetary value as euros, see equation 1.1. In the context of this study, the term 'cost escalation' consistently denotes a variation in monetary value, except where it is explicitly specified as a percentage change.

$$e_n = C \times \left(\frac{CCI_{\text{forecast, n}} - I_{\text{current}}}{CCI_{\text{current}}} \right) \tag{1.1}$$

Where:

- e_n represents the forecasted escalation at time n in expressed in a monetary value e.g. \$.
- C is the base cost estimate, this estimate accounts for resource prices at the current moment.
- $CCI_{\text{forecast, n}}$ is the forecasted CCI at time n.
- CCI_{current} is the current value of the CCI.

The forecast of the CCI ($CCI_{\text{forecast, n}}$) is generally done with time-series models e.g. Autoregressive integrated moving average (ARIMA), Holt-Winters or more complex models as an Long Short-Term Memory (LSTM). However, relying solely on a forecast of a generic index, may overlook characteristics of the project specific resources. Projects might rely heavily on a specific resource (wei Xu & Moon, 2013), and as such the CCI may not provide an accurate representation of the project's particular cost dynamics. For example, the difference between a steel or a concrete frame changes the risk exposure to material fluctuations of the project. For this reason, using an accurate prediction of the CCI might be providing an inaccurate representation of the actual cost escalation for a given project (Choi et al., 2021; Faghih & Kashani, 2018; Hwang, 2011; wei Xu & Moon, 2013; Mir et al., 2021).

To overcome the limitation of the generic approach Wilmot, Asce and Mei (2005) took a project specific approach by creating a highway construction index for Louisiana. This index was created using past contract data of the different components of a highway. Another research worth mentioning is that of Elfahham (2019), who took a country specific approach and developed a construction index representing typical Egyptian construction. These tailored indices, show clear advancements in the field however they remain only applicable to a narrow amount of projects and can not be changed towards different project characteristic . Additionally, these studies utilized non-publicly available data, making replication of their methods challenging for most practitioners.

When examining the input values of equation 1.1 it can be noted that two values are forecast-based and therefore likely to differ from actual values. The forecast of the future index value ($CCI_{\text{forecast, n}}$) will deviate. Where the uncertainty of the future values will remain completely hidden when using a point forecast.

The second value that is forecast based is the deviation of the estimated base cost from the actual base cost. A few researchers proposed tools that model these uncertainties Touran et al. (2006);

wei Xu and Moon (2013); Mir et al. (2021) . Notably, wei Xu and Moon (2013) introduced a stochastic model for predicting the CCI ENR, incorporating a Monte Carlo simulation to generate samples from both the monthly cost and the future CCI forecasts. This approach addresses the uncertainties of future costs and index values, yielding more robust and comprehensive forecasts for future escalation.

Previous attempts to predict costs changes/variations have also investigated the practicality of these models. For instance, Hwang, Park, Lee and Kim (2012) targeted the usability of forecasts by introducing an automated forecasting tool that uses ARIMA models to estimate escalations for specific resources, allowing inputs through a familiar Excel interface. This approach marked a step towards creating user-friendly forecasting methods. However, the research highlighted intrinsic shortcomings, such as reliance on point estimates and the need for a large database with historical material prices.

1.2 Problem statement

The Background section indicates the shortcomings of current forecasting methods. These methods typically use a CCI point forecast combined with a fixed project base cost estimate (as per formula 1.1). This approach has several limitations:

1. It offers a limited view of the escalation, failing to show the expected escalation of the individual resources.
2. It uses a generic index to represent each project, which does not capture the project specific resources and their corresponding inherent volatility and changes in price.
3. It overlooks the uncertainty inherent in both the forecast of the index and the costs of resources themselves.

These issues highlight the necessity for a tailored tool that can accurately compute escalation costs for diverse projects, addressing these specific limitations.

1.3 Research objectives

The primary objective of this research is to develop a tool designed to predict cost escalation across diverse construction projects. 'Diverse construction projects' refers to projects with different resources and costs associated with these resources. The tool will integrate forecasts of indices that represent these resources and the related costs of the construction project. The tool must account for uncertainties in both project costs and resource forecasts. In order to effectively allow for utilizing the tool it should be accessible to the target audience, ensuring accessibility without necessitating advanced skills in programming, econometrics, or statistics.

This objective is represented in a conceptual overview of the tool (Figure 1.1). The first element involves data: project base cost and indices. The project base cost estimates should detail resource costs and their uncertainty. Indices reflect resource cost changes, indicating escalation as a percentage.

The second element involves modeling the input data. This includes forecasting the various indices that represent the price changes of the project resources with an appropriate time-series model, while indicating the uncertainty in the forecast. As this process needs to be integrated into a tool, and allow for forecasting various indices over various time-frames it needs to be automated. Additionally, the base cost estimate needs to be converted to reflect the uncertainty of this estimate and the corresponding resource cost. An example of indicating the 'uncertainty' is by using a normal distribution for representing the probability of varying values.

After, the forecasted indices and resource costs, along with their associated uncertainties need to be combined. Given that the project's cost is calculated based on its current value, and the forecasted indices indicate the future variations in resource costs, their combination facilitates the determination of cost escalation. The selected method for combining these uncertainties is a Monte-Carlo Simulation. Where the predicted indices and resource cost would be represented by a Probability Density Function (PDF), sampled and combined with an appropriate calculation. This is in line with the research of wei Xu and Moon (2013); Touran et al. (2006). The calculation for combining the predicted change of the indices and the corresponding cost would be similar to formula 1.1, but adapted for different resources and the moment that these resources will be bought.

Finally, this combination should be translated into a comprehensive output that gives insight into the escalation of the overall project, and the corresponding resources.

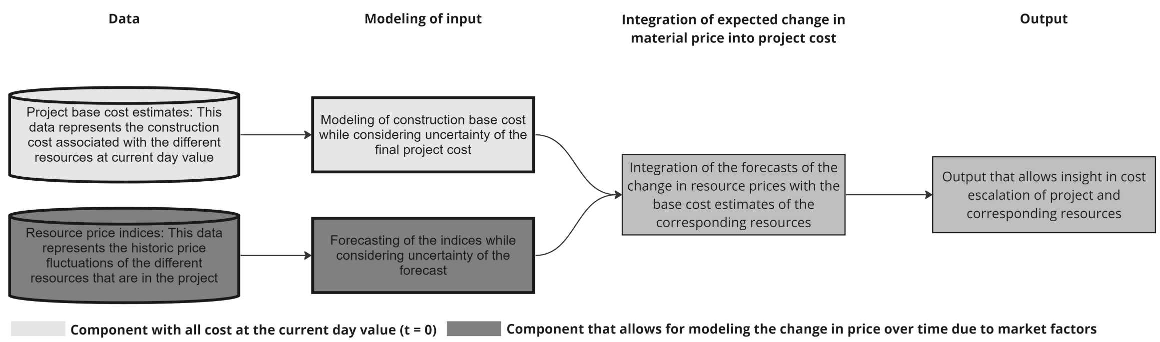


Figure 1.1: Conceptual overview of the tool

1.4 Research scope

This thesis aims to develop a proof of concept tool, setting the stage for subsequent refinements. While ensuring a baseline accuracy is vital, detailed optimization of the model's precision and elaborate exploration of alternative model architectures are beyond this study's ambit and are reserved for future research. The tool focuses on standard conditions, modeling atypical events (e.g., the 2008 financial crash, COVID-19 pandemic) falls outside its scope. Additionally, it is designed for the construction industry. The historical data selected for this research originates from the United States (US), with the objective of designing the tool to be adaptable for application in other countries.

1.5 Research questions

In pursuit of the outlined research objectives, the primary research question of this study is presented below:

> **How can cost escalation for various types of construction projects be predicted, accounting for uncertainties in final construction costs and the forecast?**

The following sub-questions are formulated to direct the research in addressing the primary research question:

SQ1 What specific indices (such as indices for labor cost, material prices and macro economic factors) are required to forecast cost escalation of various types of construction projects?

SQ2 Within the given research context, which time-series models (e.g., ARIMA, Holt-Winters) are most suitable for forecasting these identified indices?

SQ3 How can the chosen time-series model be applied to forecast the different selected indices while considering forecast uncertainty?

SQ4 How can the construction costs for various activities and associated resources be modeled while considering uncertainties in construction costs?

SQ5 Taking all associated uncertainties into account, how can the forecasts from the time-series model be integrated with the corresponding construction costs into an accessible tool that provides insight into future cost escalation in various types of construction projects?

1.6 Reading guide

The thesis is structured into four primary components. The first component, consists of the introduction and develops the broad ideas of this research.

The second component focuses on the development of a method to forecast the cost indices. This includes a Background: time-series modeling chapter that lays the groundwork with essential information on time-series modeling and selects relevant models for application. Building on this foundation, an Automated Forecasting Process (AFP) for each chosen model is developed. These processes are then evaluated for their accuracy in point forecasts and their ability to generate upper and lower bounds. The section concludes with the selection of the most effective AFP, capable of predicting various indices across different timeframes, while also most accuratly illustrating associated uncertainties.

The third component concentrates on merging the AFP with the modeling of resource cost into an accessible tool. This involves first developing a way to model the cost of different resources, including their uncertainties. Then, a method is formulated to combine these forecasts with project costs. Following the development of these methods, the integration into a practical tool is demonstrated through its application in a cost estimate.

The final component presents the conclusion of the research and a discussion of the findings. The structure of the thesis and its corresponding research questions are graphically represented in figure 1.2.

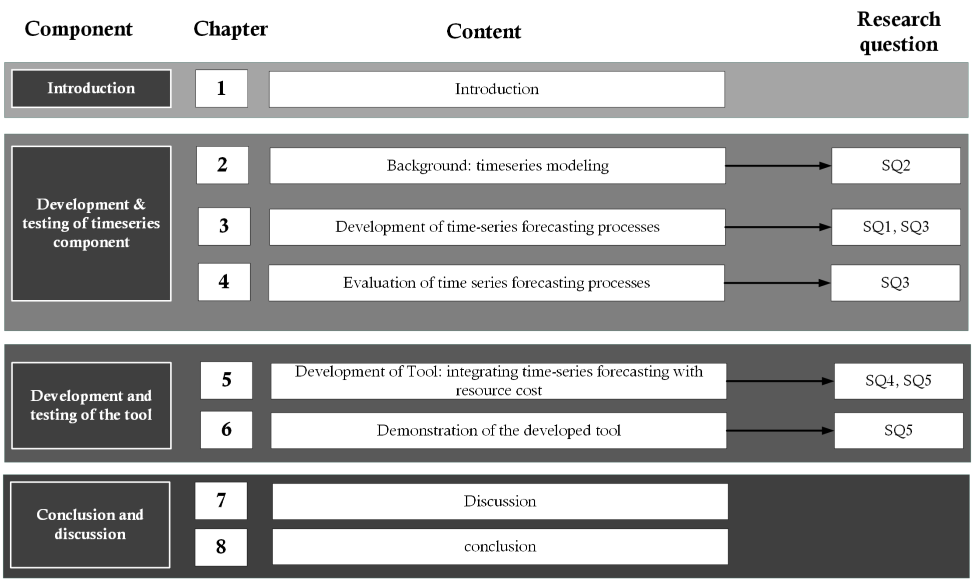


Figure 1.2: Reading guide

Chapter 2

Background: time-series modeling

A significant portion of the final tool is dependent on forecasting the cost indices. Consequently, this chapter seeks to establish the fundamental groundwork of time series modeling practices, which will serve as a basis for knowledge in subsequent chapters. A key aspect involves selecting time series models that are appropriate for further testing and evaluating their suitability in this context. This chapter is not intended to cover every intricate detail of time series models, but rather to offer a sufficient understanding that enables readers to grasp the further reasoning and methodologies presented later.

Initially, the chapter provides a comprehensive overview of time series modeling. It focuses on concepts such as transformations of the time-series and stationarity it also explores existing literature to identify potential time series models, including both univariate and multivariate models, that are suitable for this research.

After selecting the relevant time series models, we will introduce their most basic mathematical formulations. The chapter will conclude with a presentation of various statistical tests that will be utilized later in the research.

2.1 Overview of time-series modeling

This section will look into the different time-series models from a high-level perspective. Specifically, forecasting models utilized for indices related to construction are explored. Based on the existing literature time-series models are selected that are used in the rest of the research. these selected time-series models are then further detailed in section 2.2.

Table 2.1 shows an overview of the models that will be discussed. The further chapters, will elaborate on the content in the table.

Table 2.1: Review of Time Series Models

Model	**Linearity**	**Prediction Horizon**	**Interpret-ability**	**Multi-Univariate**	**Model assump-tions**
Long Short-Term Memory NN	Nonlinear	Long-term	Low	Both	None (flex-ible with non-stationary data)
Vector Auto Re-gression Model	Linear	Short-term	High	Multivariate	Stationarity, and causal relation
Vector Error Cor-rection Model	Linear	Long-term	High	Multivariate	Cointegration, causal relation and same order of integration
Holt-Winters	Linear	Short to Medium-term	High	Univariate	No specific re-quirement, but constant variance helps
Seasonal ARIMA	Linear	Short to Medium-term	High	Univariate	stationary after differencing

2.1.1 Stationarity, differencing & power transformations

Many economic time series, such as labor and material indices, exhibit trending behaviour that evolve over time (Jiang et al., 2013). These are termed non-stationary time series, which introduce specific considerations in time series modeling. Understanding stationarity is essential before delving into time series models. A time series is considered stationary if its mean and standard deviation remain constant over time.

Handling stationarity accordingly is an important aspect of time series modeling because the absence of stationarity can lead to spurious correlations between unrelated variables (Faghih & Kashani, 2018). For instance, there is a notable correlation (0.96) between per capita consumption of mozzarella cheese and the number of Civil Engineering doctorates awarded (Vigen, 2023). This high correlation likely arises not from a direct relationship but due to similar trends in both series. This correlation is likely a direct cause of the non stationary property of both time-series. Many time series models and tests require stationary data for valid statistical inferences (P. & Jenkins, 1984).

To achieve stationarity, differencing is often employed. This technique concists of subtracting consecutive observations in the series. A timeseries may need to be differenced multiple times in case this is needed to reach a stationary time series. The number of differencing steps required to achieve stationarity is referred to as the 'order of integration'. In situations where non-stationarity is attributed to non-constant variance (heteroskedasticity), applying a transformation is a common approach to stabilize the variance. (Anvari, Tuna, Canci & Turkay, 2016; Bandara, Bergmeir & Smyl, 2020; Hyndman & Athanasopoulos, 2018). Two commonly used transformations are the logarithmic transformation and the Box-Cox transformation. The Box-Cox transformation, specifically, is detailed in formula 2.1 (Hyndman & Athanasopoulos, 2018). This transformation relies on a parameter λ, which is selected to bring the data closer to a normal distribution.

$$w_t = \begin{cases} \ln(y_t) & \text{if } \lambda = 0; \\ \frac{y_t^\lambda - 1}{\lambda} & \text{otherwise.} \end{cases} \tag{2.1}$$

Where:

- w_t represents the transformed variable.
- y_t is the original variable.
- λ is the transformation parameter.

2.1.2 Univariate time series models

Univariate time series models are forecasting techniques that utilize a single variable's historical data to predict its future values. These variables are predicted on the principle that past patterns of the variable—such as trends and seasonality—are indicative of its future behavior. Techniques used for such univariate time series methods are the simple moving average, exponential smoothing, and Autoregressive integrated moving average (ARIMA) models. A study comparing different Univariate time series methods for predicting the CCI found that the Holt-Winters exponential smoothing model is the most-accurate for out-of-sample forecasting with a reported MAPE of 1% for a 1 year out of sample forecast. The ARIMA model showed the most accurate in-sample forecasts, while still having an accuracy of 1.3% for the out of sample forecast. This research also showed that several univariate models had more accurate forecasts for the CCI then subject matter experts (Ashuri et al., 2010).

Exponential smoothing and ARIMA models stand as the two predominant methods in time series forecasting, each offering a unique approach to the task. Exponential smoothing models primarily focus on identifying and utilizing the trends and seasonality present in the data. In contrast, ARIMA models are geared towards understanding and modeling the autocorrelations within the data. One notable advantage of the Holt-Winters model over the ARIMA model, is its applicability without the need for stationarity after differencing. This characteristic simplifies its use, especially in cases where first-order differencing does not achieve stationarity. Moreover, the Holt-Winters model has demonstrated a higher level of accuracy, as evidenced by its lower Mean Absolute Percentage Error (MAPE), making it a preferred choice for this research. Therefore, the Holt-Winters model will be employed further in this study due to its ease of use and proven effectiveness in forecasting.

2.1.3 Multivariate time series models

The multivariate time series model overcomes the limitation of the univariate time series model that it can can not utilize information of other variables (wei Xu & Moon, 2013). The Multivariate time series model however considers the relationship between explained variable and glsexplanatory variable. Where the explained variable is the variable that is aimed to be predicted and the explanatory variable is supposed to aid this prediction By considering this relationship, the model captures the dynamic behaviour of variables and can consider the long and short-run interactions (wei Xu & Moon, 2013). Multivariate models, requiring estimation of numerous parameters, often lead to higher specification and forecasting errors compared to the less error-prone univariate models with fewer parameters. (Cook & Doh, 2019; Giraitis, Kapetanios & Yates, 2018; Han, Zhang, Qiu, Xu & Ren, 2019).

In multivariate modelling, three approaches have been explored. These are the vector autoregression (VAR), vector error correction model (VECM), Long Short-Term Memory (LSTM)) model. The VAR and VECM model are Econometric models used specifically for time-series analyses. Whereas the VECM model is an extension of the VAR model. LSTM is a machine learning model that is a subtype of a Recurrent Neural Network. The LSTM model has broader application then just modeling timeseries.

One of the main trade-offs within the model selection process is the black box vs glass box trade off, which determines the explainability of the model. As explained by Burkov (2020), the black-box models generally have a lower prediction error but make explaining the prediction more difficult. Burkov makes the distinction between the neural network models that are more of black box models, the linear regression and decision tree models that are easier to interpret and explain. Tying this to the 3 models mentioned in the previous paragraph, the LSTM model will be a black box as it is a Neural Network, the VAR and VECM models are both linear models, which are easier to explain.

The LSTM model seems promising in terms of accuracy, particularly due to its capability to interpret nonlinear patterns in data, which the linear VAR and VEC model are not capable of. However, a significant limitation of the LSTM model is its 'black box' nature, which restricts the Explainability of its results. A downside of a model with low Explainability is a barrier for adopting the model (Duval, 2019). Given this argument and the exploratory nature of this research with the emphasis on developing a conceptual model, Explainability is prioritized over model accuracy. Therefore, this study focuses exclusively on the VAR and VECM models, as they offer greater clarity in understanding the results.

Within the literature related to construction index forecasting the VECM model has proven to reach a high prediction accuracy. The VEC Model has been used by Faghih and Kashani (2018) to forecast material prices 36 months into the future. His forecasts reached a MAPE as low as 0.56% when forecasting steel prices. The VAR model has been used by Hwang (2011) for predicting the CCI ENR 2 years into the feature with a MAPE of 1.01%. When comparing the research of Faghih and Kashani (2018) and Hwang (2011) the VEC model seems superior over the VAR model, as it has a higher accuracy of a more volatile index . This enhanced performance of the VEC model is not surprising, considering that it is an augmentation of the VAR model designed to encapsulate long-term relationships among the variables. This, in theory makes a VECM model better at longer term predictions (Faghih & Kashani, 2018). Overall research finds that in general a VECM model should be preferred as long as there are no substantial model mis-specifications (Clements & Hendry, 1998). However, from a forecasting point of view, if either the error correction term is not significant or the cointegration relationship changes over the forecast sample, the VAR in differences could produce more accurate forecasts than the VECM (Ghysels & Marcellino, 2018). Another claim made against VAR models is that the differencing for achieving stationarity (see subsection 2.1.1) can distort the model (Fanchon & Wendel, 1992). Although the VEC model has the potential to yield superior results in theory, the VAR model's simplicity sometimes makes it a more suitable choice, particularly when there is no cointegrated relationship among variables. Consequently, this research will further explore both the VAR and VEC models.

2.2 The selected time-series models.

In this chapter, the previously selected models will be further elaborated. With a focus on the specific formulas that underpin these models. The models in question are the Holt-Winters, VEC, and VAR models. Each of these models play a role in forecasting time-series data within the context of our research.

Additionally, the Seasonal and Trend decomposition using Loess (STL) model will be detailed. While the STL model differs from the primary forecasting models, its inclusion is needed. This model is important as it will be utilized in a subsequent chapter to assist in establishing the upper and lower bounds for the various time series models under discussion.

For each of the models some of the formulas are given in there most basic form. The Python module of Statsmodels is utilized for the implementation of the various models (Seabold & Perktold, 2010). In this module the Holt-Winters implementation is according to Hyndman and Athanasopoulos (2018), the VAR and VECM model follows Lütkepohl (2005), and STL decomposition follows the approach outlined by Cleveland, Cleveland, McRae and Terpenning (1990). Readers seeking an in-depth explanation of the different model are referred to these sources.

2.2.1 The Holt-Winters model

The Holt-Winters model is an extension of the Simple exponential smoothing model proposed in the late 1950s (Holt, 2004; Winters, 1960). Forecasts with exponential smoothing methods are weighted averages of past observations, with the weights getting smaller the older the observation is. In other words, the more recent the observation the higher weight of the observation. Holt-winters seasonal method is a method that extends the exponential smoothing with a trend and a seasonal component.The Holt-Winters forecasting model is an advanced form of time series analysis that extends simple exponential smoothing by incorporating both trend and seasonal components.

It provides two versions of forecast equations: additive for constant seasonal effects and multiplicative for when seasonal variations are proportional to the level of the series. In the additive model, the forecast ($\hat{y}_{t+h|t}$) is a sum of the estimated level at time t (ℓ_t), the product of the slope of the trend component at time t and the number of periods ahead (h) being forecasted (hb_t), and the seasonal deviation ($s_{t+h-m(k+1)}$), which is adjusted for the season's length (m). The multiplicative model, on the other hand, multiplies the sum of the level and trend components by the seasonal index. Both methods aim to predict future values by considering the most recent data points more heavily and adjusting for patterns that repeat over a fixed period. The multiplicative model, on the other hand, multiplies the sum of the level and trend components by the seasonal index. Both methods aim to predict future values by considering the most recent data points more heavily and adjusting for patterns that repeat over a fixed period. For the seasonal and trend parameter it is also possible to damp the component. This means that further in the future the component will approach a constant.

The forecast equation for the Holt-Winters' additive method is:

$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)} \tag{2.2}$$

The forecast equation for the Holt-Winters' multiplicative method is:

$$\hat{y}_{t+h|t} = (\ell_t + hb_t) \times s_{t+h-m(k+1)} \tag{2.3}$$

Where:

- $\hat{y}_{t+h|t}$ The forecast for h.
- ℓ_t The level component at time t.
- hb_t The trend component at time t.
- $s_{t+h-m(k+1)}$ The seasonal component, representing the seasonal effect t.
- k is in integer that helps to align the seasonal index from a past cycle to the forecast period it is calculated in a way that $t+h-m(k+1)$ point to the correct past period that had the same position in its seasonal cycle as the forecast period will have in its cycle.

2.2.2 The vector autoregression model (VAR)

The VAR model accounts for the relationships between variables and their previous values, as well as the relationships between past values of the variable itself. In contrast, a univariate AR model only considers the variable itself and its own lagged values. The system of equations (2.4) and (2.5) below represents a VAR(1) model involving two variables. In this VAR(1) model, the values of both variables will be examined one lag (one time step) before to predict the current value. The a terms in the equations denote the coefficients multiplied by the lagged values of the two time series. The aim is to optimize this system of equations to minimize the error term, denoted as $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$, which represent the unexplained variation or residuals.

$$y_{1,t} = C_1 + \phi_{1,1} y_{1,t-1} + \phi_{1,2} y_{2,t-1} + \varepsilon_{1,t} \tag{2.4}$$

$$y_{2,t} = C_2 + \phi_{2,1} y_{1,t-1} + \phi_{2,2} y_{2,t-1} + \varepsilon_{2,t} \tag{2.5}$$

Where:

- the values of variables $y_{1,t}$ and $y_{2,t}$ at time t.
- the values $y_{1,t-1}$ and $y_{2,t-1}$ at one time lag (previous time step).
- $\phi_{1,1}, \phi_{1,2}, \phi_{2,1}, \phi_{2,2}$ are the coefficients representing the influence of the lagged values of the variables on their current values.
- $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are the error terms for the two equations, capturing unexplained variation or residuals.

The above VAR(1) model can be translated into the more general Vector form (Wooldridge, 2009):

$$\mathbf{Y}_t = \mathbf{C} + \Phi_1 \mathbf{Y}_{t-1} + \Phi_2 \mathbf{Y}_{t-2} + \cdots + \Phi_p \mathbf{Y}_{t-p} + \varepsilon_t \tag{2.6}$$

Where:

- $\mathbf{Y}_t$ is a 2×1 vector of variables $\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix}$.
- $\mathbf{C}$ is a 2×1 vector of constants, the intercept term $\begin{pmatrix} C_1 \\ C_2 \end{pmatrix}$.
- Φ_i (for $i = 1, 2, \ldots, p$) are 2×2 matrices that describe the relation between the considered variables at each lag.
- ε_t is a 2×1 vector of error terms $\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$.
- p is the number of lag terms in the VAR model.

2.2.3 The vector error correction model (VECM)

The general VAR model, presented in equation (2.6), can be converted into a VEC model , which includes a term to account for cointegration among the variables. This in theory allows for capturing both the long and short term relation, between variables. While the VAR model, as shown in equation (2.6), effectively captures the dynamics of variables through their own past values and the past values of other variables in the system, it does not explicitly incorporate a long-term equilibrium relationships between these variables, especially when they are non-stationary. This is where the Vector Error Correction Model (VECM) becomes relevant.

The VECM reformulates the VAR model by including an error correction term, which represents the long-term equilibrium relationship between the variables. This term, allows the model to account for deviations from this equilibrium and adjust the variables in the short term to return to the long-term stable state. The corresponding VECM for a VAR(1) model, assuming the variables are cointegrated, is given by:

$$\Delta \mathbf{Y}_t = \Pi \mathbf{Y}_{t-1} + \Gamma_1 \Delta \mathbf{Y}_{t-1} + \mathbf{c} + \varepsilon_t \tag{2.7}$$

Where:

- $\Delta \mathbf{Y}_t$ represents the first differences of the variables, capturing the short-term dynamics.
- Π is the matrix reflecting the long-run equilibrium relationships between the variables, derived from the cointegration analysis.
- Γ_1 represents the short-term adjustments, similar to the lagged coefficients in the VAR model.
- $\mathbf{c}$ and ε_t are the intercept and error terms, respectively.

The VECM, thus, not only captures the short-term dynamics of the variables, as in the VAR model, but also integrates the long-term equilibrium constraints.

2.2.4 STL decomposition

Time series analysis often involves identifying and separating different underlying patterns within the data. Typically, a time series is decomposed into three components: a Trend-cycle component (commonly referred to as the 'trend'), a seasonal component depicting periodic changes, and a remainder encompassing elements that do not fit into the previous categories (Hyndman & Athanasopoulos, 2018). This decomposition can be either multiplicative, as shown in equation (2.9), where the seasonality's magnitude varies with the time series level, or additive, as in equation (2.8), where seasonality remains consistent regardless of the level.

The STL (Seasonal and Trend decomposition using Loess) method is particularly effective for time series decomposition (Hyndman & Athanasopoulos, 2018). It will be employed in section Moving Block Bootstrapping to facilitate the creation of prediction intervals. The STL decomposition enables the isolation of residuals from the trend and seasonality components, allowing for the use of bootstrapping techniques to generate new series.

An illustrative example of STL decomposition is applied to Google search data for ice cream (Ritvik, 2020). It reveals a distinct increase in seasonality during summer months and an overall rising trend, likely due to the general increase in Google searches over time. However, a limitation of the STL method is its inability to perform additive decomposition directly (Hyndman & Athanasopoulos, 2018). To address this, a logarithmic transformation of the data is required before the STL application. Given this study's focus on economic time series, which typically exhibit multiplicative trends, this transformation becomes essential for applying the STL method effectively (Hyndman & Athanasopoulos, 2018).

Additive Decomposition:

$$y_t = S_t + T_t + R_t \tag{2.8}$$

Multiplicative Decomposition and its Logarithmic Transformation:

$$\begin{aligned}
&\text{Original Multiplicative Model:}\\
&y_t = S_t \times T_t \times R_t\\
&\\
&\text{Logarithmic Transformation (Equivalent Additive Model):}\\
&\log(y_t) = \log(S_t \times T_t \times R_t) = \log(S_t) + \log(T_t) + \log(R_t)
\end{aligned} \tag{2.9}$$

Where:

- y_t represents the data at period t.
- S_t is the seasonal component, indicating repetitive patterns within specific periods.
- T_t is the trend-cycle component, reflecting long-term progression or cyclic movements.
- R_t is the remainder component, encapsulating irregular fluctuations not explained by the seasonal or trend-cycle components.

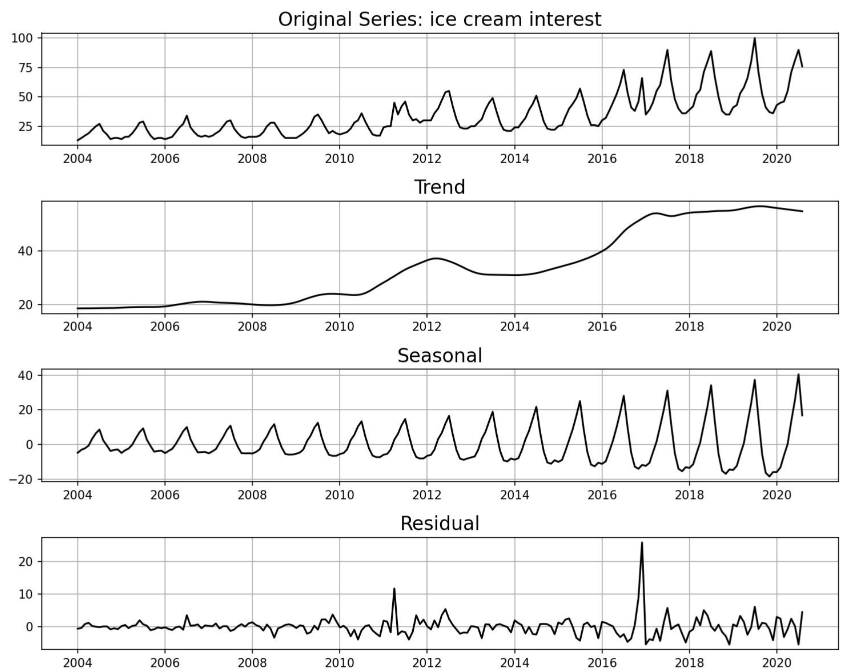


Figure 2.1: STL decomposition: example of decomposition for ice cream searches on google

2.3 Statistical tests utilized

This section provides a detailed elaboration of the various statistical tests that will be employed in the study. The statistical tests relate to the different models and specific data requirements. For a multi-variate model the explanatory variable needs have predictive strength before utilizing them in the model. The Granger Causality test indicates a causal relationship between the variables. The Johanson Cointegration test is used for indicating a cointegrated relation, which means that there exists a linear combination of variables that is stationary. Additionally as described in section 2.1.1 , the stationarity property that needs to be satisfied, will be tested by using the Augmented Dickey-Fuller (ADF) test. Similar to the timeseries models, the implementation of the statistical test are utilized using the Statsmodels python module (Seabold & Perktold, 2010). A more detailed elaboration on the Johanson Cointegration test can be found in Lütkepohl (2005), for the implementation of the Granger Causality test and ADF test Greene (2012) provides further detail.

A graphical representation of the different requirements/properties, and the corresponding tests for the different models can be found in Figure 2.2. As illustrated, the Vector Error Correction Model (VECM) emerges as the most complex in terms of requirements, whereas the Holt-Winters method presents no requirements related to the aspects in the visualization.

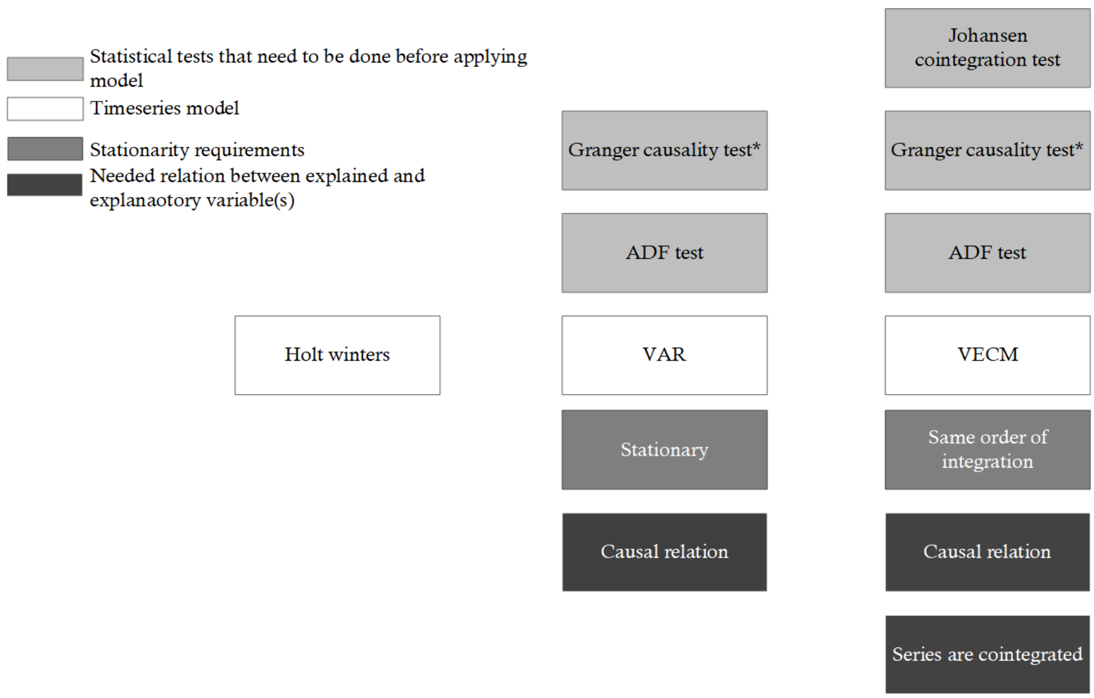


Figure 2.2: Overview of tests and relation to model and model requirements

2.3.1 Augmented Dickey–Fuller test

The Augmented Dickey-Fuller (ADF) test is a statistical test used for assessing if time-series data is stationary. The null hypothesis assumes that a unit root is present, implying non-stationary. The test statistic is calculated as follows:

$$\text{ADF Statistic} = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \tag{2.10}$$

where $\hat{\gamma}$ is the estimate of the lag coefficient in the test regression, and $SE(\hat{\gamma})$ is the standard error of the estimate. A more negative ADF statistic leads to the rejection of the null hypothesis, suggesting the time series is stationary(Dickey & Fuller, 1979).

2.3.2 Granger Causality test

The Granger Causality test assesses whether one time series can be used to forecast another . It is based on the principle that if a variable X causes Y, then changes in X will systematically occur before changes in Y. The basic bivariate model for testing Granger causality from X to Y is:

$$Y_t = \alpha + \sum_{i=1}^{k} \beta_i Y_{t-i} + \sum_{i=1}^{k} \gamma_i X_{t-i} + \varepsilon_t \tag{2.11}$$

where Y_t is the current value of the dependent time series, X_{t-i} are past values of the independent series, and ε_t is the error term. The null hypothesis is that the coefficients γ_i are all zero, indicating that X does not Granger-cause Y(Granger, 1969).

2.3.3 Johansen cointegration test

The Johansen cointegration test is used to determine the cointegration relationship between two or more time series. The test is based on the maximum likelihood estimation of the vector error correction model. This corresponds to the equation presented in 2.7, but extended to include higher lag terms. The null hypothesis involves the rank of Π, with a rank of zero indicating no cointegration and a rank equal to the number of series minus one indicating full cointegration(Johansen, 1988).

2.4 Testing time-series models

In this section, an outline of the methodologies employed for validating the forecasts of the time-series models will be given. Primarily, the approach involves describing the rolling-window train-test split method. This method allows for testing the model over various time-frames, making it a robust approach for testing. Furthermore, the performance metrics to be utilized in evaluating the model will be detailed.

2.4.1 The Rolling Window train-test split

For the validation of the testing set of the time series, the rolling window split technique is employed to ensure a robust and consistent evaluation. The method allows for train-test splitting over different time-frames, and with this generates a robust estimate on the accuracy. This approach facilitates the creation of sequential training and testing datasets, maintaining the chronological integrity of the data, and providing insights into the model's performance across different segments. This method is used to always ensure the same training size and due to this results in a clear comparison.

Figure 2.3: Rolling window train-test split

2.4.2 Model performance Metrics

Different model performance metrics can be utilized to validate the performance of the model. Mean Absolute Percentage Error (MAPE) (see equation: 2.12) allows us to asses the accuracy of the forecast, which is a common used metric in time-series modeling, as used in Hwang (2011); Dong et al. (2020); Faghih and Kashani (2018); Faghih, Gholipour and Kashani (2021); Wilmot et al. (2005). The bias (see equation: 2.13) is used to understand if the model makes a systematic error by generally over or underestimates the values.

Finally, a less common formula is used to score the created prediction intervals. Most often the prediction interval is scored based on coverage i.e., whether actual values fall within the intervals (Wilmot et al., 2005; wei Xu & Moon, 2013; Liu, 2007). Contrarily to the research of Mir et al. (2021), who looks at both the coverage and the width of the prediction interval. The second method of Mir et al. (2021) intuitively is a lot more sensible, as the width of the interval determines how practical it is to use. Since at a very large width the only function the interval serves is to indicate the high degree of uncertainty (Chatfield, 2001). In order to take into account both width and coverage, the interval score, as proposed by Gneiting and Raftery

(2007); Bracher, Ray, Gneiting and Reich (2021), is adopted. The interval score consists of three components: the base, which is the width of the interval; and the second and third terms, which incorporate a factor of 2 divided by α to account for deviations when actual values fall outside the interval. where alpha is 0.05 for a 95% confidence interval.

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \tag{2.12}$$

$$\text{Bias} = \frac{1}{n} \sum_{t=1}^{n} (F_t - A_t) \tag{2.13}$$

Where:

- A_t is the actual value at time t.
- F_t is the forecasted value at time t.
- n is the number of forecasted points.

$$\text{Score} = (U - L) + \frac{2}{\alpha}(L - O) \cdot \mathbf{1}_{(O<L)} + \frac{2}{\alpha}(O - U) \cdot \mathbf{1}_{(O>U)} \tag{2.14}$$

Where:

- O is the observed value.
- U is the upper limit of the acceptable range.
- L is the lower limit of the acceptable range.
- α is a weighting factor.
- $\mathbf{1}_{(\cdot)}$ is an indicator function.

Chapter 3

Development of time-series forecasting processes

This section describes the method for forecasting the cost indices. Which are the indices associated with the project's resources. The forecasting method needs to be integrated into a tool which allows for forecasting different indices, over different time periods. Additionally the developed tool, while initially focusing on the US, should allow for application to data of different countries. For these reasons, a method of forecasting that requires as little manual adjustments as possible is desired. Or in other words, an Automated Forecasting Process (AFP) is desired. An AFP will be created for each of the previously selected model classes, being the VAR, VECM and Holt-Winters model. An overview of the AFP is visible in figure 3.1

First the time series data will be selected, focusing on both the cost indices (explained variables) and the data potentially influencing these resources (explanatory variables). Following this, the approach to data preparation will be detailed.

Later, the method of determining which model classes are suitable given the data is explained. For example, among other requirements, a VECM model needs a cointegrated relation between the variables for the model to be utilized. This process, will determine which of the AFP's can be used given the cost index.

This will be followed by an explanation of the process used for creating candidate models with different configurations. After which the methodology for identifying the most effective model is given.

Finally, the section concludes with an explanation of the method for generating the final distributional forecast of the monthly Probability Density Function (PDF) through the application of bootstrapping techniques.

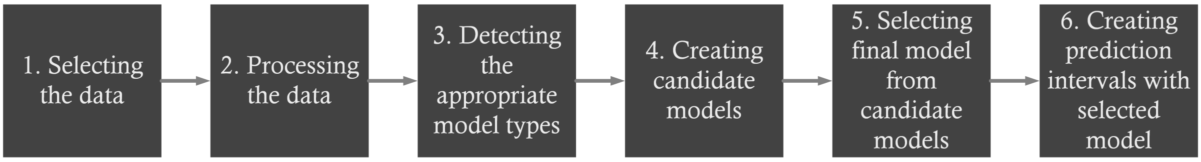


Figure 3.1: Overview of the automated time-series modelling process

3.1 Time-series data acquisition

This section describes the process of selecting the variables for the model. A distinction is made between the explained variable (i.e. predicting variables) and the explanatory variable (i.e. the variable used for predictions). The explained variable will represent the resources e.g. labour, steel or concrete. The explanatory variables are needed for the multivariate time-series models, and will be macro economic indicators that could help predicting the resource data. As the proposed tool will be demonstrated using time series data from the United States (US), data from this country will be gathered.

3.1.1 Explained variables

Since the final model will be created for the prediction of the escalation of various project types, a range of resources needs to be selected. For the representation of the material types the Produces Price Index (PPI) of various materials is used. The PPI indicates the avarage change in the selling price of the producer over time (of Labor Statistics, 2023). This indicator has been used in previous research to represent material price changes, Amongst which (Faghih & Kashani, 2018) and (Mir et al., 2021). For the wage index the statistics from 'Average Hourly Earnings of Production and Nonsupervisory Employees, Construction' is used. The PPI material and labour indexes can be collected from the US bureau of Labor statistics (of Labor Statistics, 2023)us bureau of labor statistics. The PPI is calculated through systematic sampling of specific industries. The future price of a resource using the PPI index can be calculated using the following formula:

$$\text{future resource price} = \text{current resource price} \times \left(\frac{\text{future PPI}}{\text{current PPI}} \right) \tag{3.1}$$

All the PPI data is collected from the dates 1993-01-01 until 2023-11-01. figure 3.2 shows all the explained variables selected, with each index standardized to a baseline value of 100 for simplified interpretation. Due to the large range of available indices representing various materials and labour types, expanding the selected indices for future research should be straightforward. The mean, standard deviation min and max of the pct change over the months are given in table 3.1. Note the strong differences in standard deviation between the series, which underscores the importance of modeling each resource individually. This approach provides a more accurate understanding of the specific uncertainties associated with each resource.

Table 3.1: Summary Statistics for resource data

Statistic	**Mean**	**Std**	**Min**	**Max**
PPI: Steel	0.01	0.49	-0.1794	0.3817
PPI: Concrete	0.0032	0.0061	-0.0143	0.89
Construction Wage	0.0025	0.0038	-0.0147	0.0182
PPI: Copper	0.01	0.21	-0.20	0.2645
PPI: Asphalt	0.0036	0.0132	-0.0514	0.11
PPI: Sand	0.0034	0.01	-0.0083	0.0352

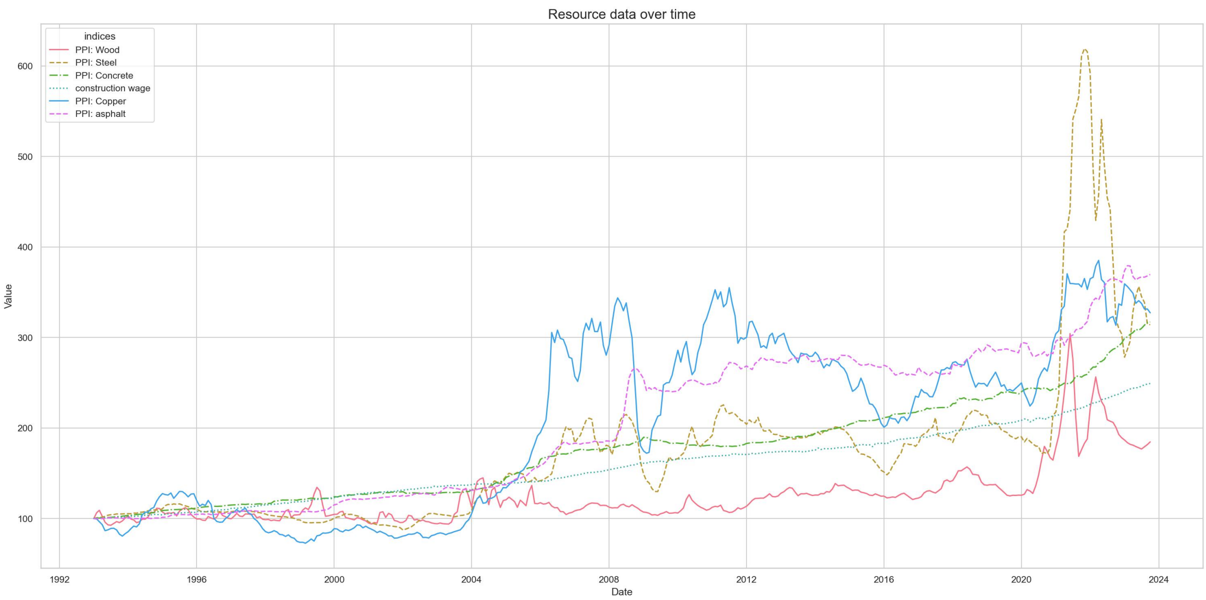


Figure 3.2: Explained material and labor data

3.1.2 Explanatory variables

The explanatory variables of the model should have a predictive power for the changes in labour costs and material costs, together forming the changes in cost for the construction project. Construction supply and demand or market conditions influence the changes in construction prices (Runeson, 1988; Skitmore, 1987; R. G. Taylor & Bowen, 1987). For this reason, picking indicators that have a relation with these factors is relevant.

In Table 3.2 an overview of the selected variables is given, whereby the data is recorded monthly, tracked from 1982-06-01 and collected through the FRED st. Louis' API. The selected variables stem from previously used variables in successful econometric models by the following researches (Faghih & Kashani, 2018; wei Xu & Moon, 2013; Faghih et al., 2021; Mir et al., 2021; Akintoye, Bowen & Hardcastle, 1998). An elaboration on the chosen explanatory variables can be found in appendix A.

The mean, standard deviation min and max of the pct change of the explanatory variables over the months are given in table 3.2.

Table 3.2: Summary statistics for explanatory variables

Statistic	Mean	Std	Min	Max
PPI	0.0022	0.0111	-0.0533	0.0321
CPI	0.0020	0.0013	-0.09	0.0077
Money Supply	0.0028	0.0069	-0.0118	0.0722
GDP	2.4619	6.2428	-68.5529	46.2616
Unemployment Rate	0.0020	0.1274	-0.1765	2.3409
Interest Rate Long Term	0.0519	1.8915	-12.7421	29.64
Total Construction Spending	0.01	0.0120	-0.0373	0.69
Housing Starts	0.0034	0.0785	-0.2699	0.2396
Iron Ore Price	0.0096	0.0873	-0.2984	0.7151
Building Permits	0.0019	0.0510	-0.2211	0.1933

3.2 Time-series data preparation

This chapter describes the data processing steps used in the AFP. These processing steps are needed to get stationary data. Stationary data allows for (1) applying the Granger Causality test, (2) Allow for application to the VECM and VAR model, and (3) reduce variance within the data.

To streamline the process of achieving stationarity within the data, the same transformations will be employed to all the indices, mitigating some complexity within the Python implementation, without making significant sacrifices on the quality of the models. This approach averts the need for managing multiple series with diverse transformations, thereby simplifying the tracking and application of transformations, as well as the reversion process post-prediction. The chosen transformations include first-order differencing and a logarithmic transformation. The logarithmic transformation aims to reduce noise, achieve constant variance over time (homoscedasticity) and allow a multiplicative model in the STL decomposition (see equation 2.9), thus serving three purposes.

3.3 Model class selection

Building upon the uniform approach for achieving stationarity, further details on the process of identifying the appropriate models will now be given. Since different model classes and tests necessitate unique requirements, application of them needs to be done with care. Table 3.3 provides a comprehensive overview of the specific requirements and functions associated with each test and model. This table shows the various requirements for applying VAR and VECM models. It also show that the Holt-Winters models requirements are very limited.

Given the various requirements, the flowchart in figure 3.3 is developed, which describes the possible model classes given the data properties. Note that regardless of the outcomes of the tests, the Holt-Winters model can always be utilized.

The initial phase of the model class selection involves assessing if the data is stationary after the log transform and differencing. When the data can be made stationary the Granger Causality test can be applied. The Granger Causality test assesses the explanatory variables predictive strength. If the test indicates that the explanatory variables lack predictive power for the response variables, the use of a multivariate model is deemed unnecessary. On identifying a Granger causal relationship, the next step is to examine the potential for cointegration, which if present allows for employing a VECM model.

While the process of achieving stationarity is generic, there is a difference in the data that is used for the models. The Holt-Winters model handles transformations of the data within the model fitting and for this reason uses the original data (more on this in subsection 3.4.3). The VAR model needs the data to be stationary and for this reason does use the data with a log transform and differencing. Finally, the VECM model uses only the log transform, as the model requires data with an order of integration of 1.

From figure 3.3, it can be noted that the amount of model classes possible increases when moving to the right of the flow chart. Where in the case of cointegration, Granger Causality and stationarity all three model classes are possible. Given this flowchart, 4 possible combinations can be created for an AFP, given that a fall-back model needs to be in place, see table 3.4. The selection between these 4 options will be tested in chapter 4.

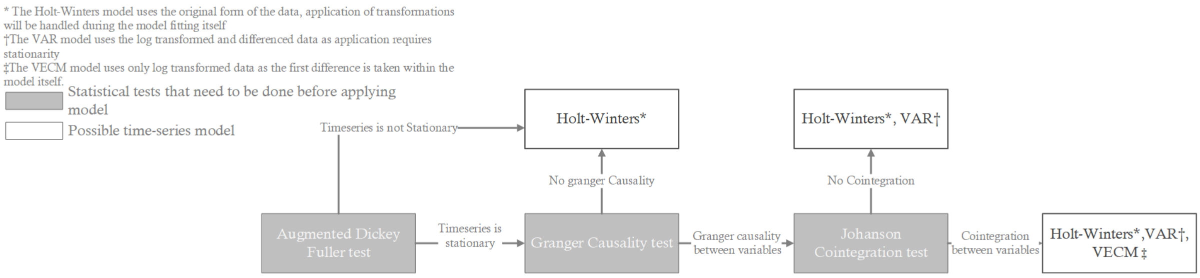


Figure 3.3: Flowchart that indicates the possible models

Table 3.3: Summary of Tests/Models and Their Functions and Requirements

Test/Model	**Function**	**Requirement**
ADF test	tests for stationarity	No requirements*
Granger Causality Test	Indicates a causal relation between variables	Stationary of both variables
Johanson Cointegration Test	Indicates that variables are cointegrated i.e., have a long term relation	Same order of integration
VECM	Allows for modeling long and short term interaction between variables	cointegrated variables, same order of integration, Outlier/Noise Adjustment
VAR	Allows for modeling short term causal relation	stationary, Outlier/Noise Adjustment
Holt-Winters	Applies exponential smoothing to capture level, trend, and seasonal components	Outlier/Noise Adjustment

* This test requires standard time series data prerequisites: the dataset must be a continuous time series without missing values and should have an adequate sample size for reliability. No specific transformations or unique properties are necessary for application.

Table 3.4: Options of Model classes in AFP

Option	**Preferred Model**	**1st Fall-Back**	**2nd Fall-Back**
Option 1	VECM	VAR	Holt-Winters
Option 2	VECM	Holt-Winters	None
Option 3	VAR	Holt-Winters	None
Option 4	Holt-Winters	None	None

3.4 Development of candidate models

After the model class is selected, the model configurations need to be determined. The Holt-Winters, VAR and VECM model all have some unique configurations. When utilizing the model, the objective is to select the configurations that approximate the true underlying system. The process of finding these configurations will be done by first creating candidate models, each with different configurations. After the creation of the candidate models the final 'best' model can be selected, which ideally approximates the underlying system. The method for creating the candidate models for Holt-Winters, VAR and VECM will be elaborated in the following three subsections.

3.4.1 Candidate model selection: VAR

The key configurations of the VAR model are the selection of variables, and their associated lags (Hyndman & Athanasopoulos, 2018). The process of selecting the possible variables and lags of interest is done by using the Granger Causality test. This approach has been done in previous research by amongst other (Ashuri & Shahandashti, 2012; Faghih & Kashani, 2018). Following the selection of explanatory variables with a leading relation, a model is created with all possible combinations of these variables and their time lags. This approach guarantees that all potential models are examined.

A fictive example of this is given in figure 3.4. In this figure, the variable 'steel' is analyzed for Granger Causality with various explanatory variables. The fictive findings indicate that both CPI and PPI have a predictive relationship with 'steel', making them useful for forecasting. Following this, all potential combinations of lags identified in the Granger Causality test are combined into different candidate models.

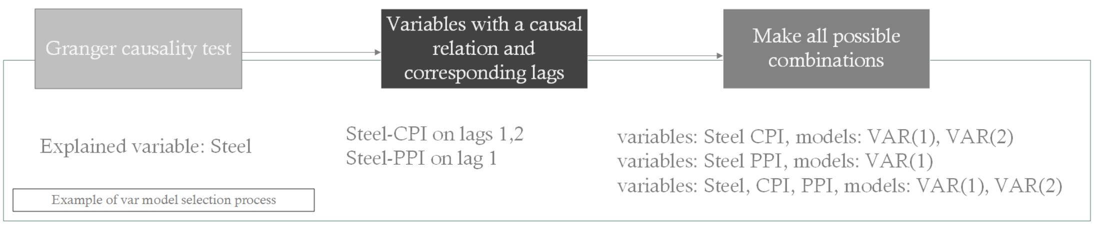


Figure 3.4: Model selection process VAR

3.4.2 Candidate model selection: VECM

The VECM model, as an extension of the VAR framework, inherits its selection process. All steps applicable to the VAR model's variable and lag selection, are similar in the creation of candidate VECM models. The process extends to examining the rank of cointegration among variable combinations. Should the rank exceed zero, indicating a cointegrated relationship, a VECM is formulated corresponding to that rank.

3.4.3 Candidate model selection: Holt-Winters

The Holt-Winters model can come in different forms, allowing for the creation of different models. This is due to its capability of incorporating both trend and seasonal components. These components can be structured in additive or multiplicative terms. The categorization of these models was originally proposed by (McCormick, 1969), who included the concept of a multiplicative trend. This framework was further expanded by (Gardner, 1985), introducing the additive damped trend, and by (J. W. Taylor, 2003) with the multiplicative damped trend.

In this study, the multiplicative trend is excluded, aligning with (Hyndman & Athanasopoulos, 2018) recommendations. Multiplicative trends assume that growth continues at an increasing rate, leading to exponential forecasts. This assumption is considered unrealistic for long-term predictions in construction indices.

Table 3.5 outlines the nine viable smoothing techniques. Each of these techniques will be further explored through variations, including the application of a Box-Cox transform and a log transform. This transform encompasses three scenarios: a fitted Box-Cox transform where the value for lambda is optimized to transform the data to a normal distribution, a Box-Cox transform with lambda set to 0 (equivalent to a logarithmic transform), and the absence of any Box-Cox transform. Additionally, we will investigate different seasonal patterns, including no seasonality, and seasonal cycles of 2, 3, 4, 6, and 12 periods in a year.

Table 3.5: Different Holt winter configurations

	Seasonal Component		
Trend Component	**None (N)**	**Additive (A)**	**Multiplicative (M)**
None (N)	(N,N)	(N,A)	(N,M)
Additive (A)	(A,N)	(A,A)	(A,M)
Additive Damped (A_d)	(A_d,N)	(A_d,A)	(A_d,M)

3.5 Selection of final model

After the candidate models have been created, a final model that will make the future predictions will be selected. This process consists of two steps, the first one is to filter out the 'bad' models based on residuals, often termed residual diagnostics. After, the final model is selected based on an information criterion.

3.5.1 Residual diagnostics filter on different models

After the candidate models are created and fitted a residual diagnostics filter is applied to filter out models. Using appropriate diagnostic checks will likely lead to a model that is at minimum a good approximation of the 'True' model (Chatfield, 2001). The residuals of a model are obtained by calculating the difference between the observation and the corresponding fitted values.

$$e_t = y_t - \hat{y}_t. \tag{3.2}$$

Where e_t is the residual, y_t is the actual value, and $\hat{y}_t$. is the fitted value. The residuals are ideally uncorrelated and have a mean of zero (Hyndman & Athanasopoulos, 2018). In the case that the residuals do not have these properties a better model can be created, since certain patterns are not captured.

To verify the first criterion of uncorrelated residuals, we employ the Ljung-Box Q-statistic (Ljung & Box, 1978). This statistical test operates under the null hypothesis that the time series exhibits no serial correlation. We apply the Ljung-Box Q-statistic up to a lag of 12 to assess serial correlation. This approach aligns with methodologies used in various studies (Faghih et al., 2021; Hwang et al., 2012; wei Xu & Moon, 2013; Sekma, Elleuch & Dridi, 2016).

For the second criterion, ensuring a zero mean in the residuals, we conduct a standard t-test with a significance level of 5%. The null hypothesis of this t-test hypothesizes that the mean of the residuals is zero. Rejection of this hypothesis implies a mean significantly different from zero, indicating that the model may not be suitable for use. This t-test assumes that the residuals are approximately normally distributed.

3.5.2 Information criterion

After applying residual diagnostics to filter models out, the next step is selecting an optimal model, here the bias-variance trade-off will be utilized. High bias indicates model under-fitting, leading to poor performance on training data and low predictive accuracy for testing data. Conversely, high variance suggests over-fitting to training data, enabling excellent predictions for these specific data points but poor generalizability. Both high bias and variance typically result in reduced predictive accuracy on test data, as highlighted by (Burkov, 2020; Chatfield, 2001). Figure 3.5 graphically represents this trade-off, highlighting the 'zone of solution'—an ideal equilibrium between bias and variance. One approach to achieving this balance is employing an information criterion. This criterion penalizes excessive model complexity while rewarding good fit, aiding in finding the balance. The Akaike Information Criterion (AIC) is employed for final model selection, the AIC is formulated in formula 3.3. The model with the lowest AIC will be utilized for forecasting.

$$AIC = 2k - 2\ln(\hat{L}) \tag{3.3}$$

Where:

- k is the number of parameters in the statistical model.
- $\hat{L}$ is the maximum value of the likelihood function for the model.

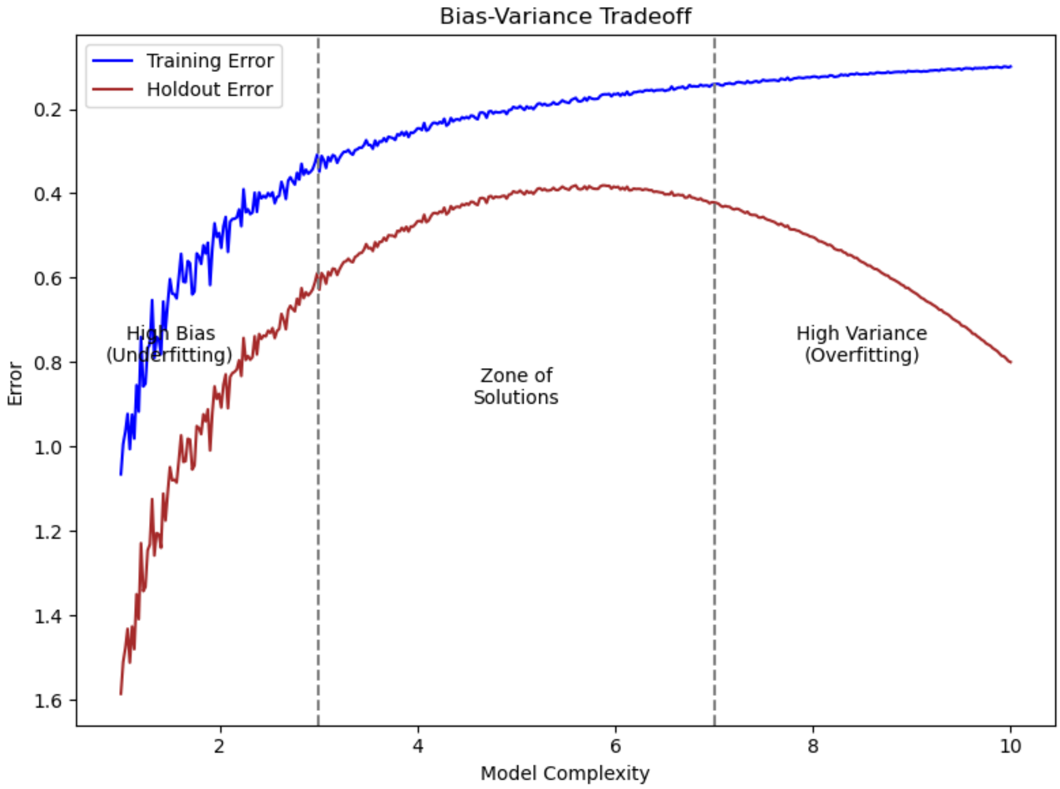


Figure 3.5: Bias variance trade-off

3.6 Predicting with the selected model

This section will focus on transferring the initial point forecasts of the resource indices to a distributional forecast. A distributional forecast refers to the range of possible future values with there associated probabilities. These distributions can then later be integrated into a Monte Carlo simulation for the final tool, as mentioned in section Research objectives. Distributional forecast can be transferred into a Prediction Interval (PI) which specifies an interval within which we expect the future value to lie in with a specified probability. While the Monte Carlo simulation will not utilize a prediction interval, they are relevant within this study as they allow for assessing the quality of the intervals created, thereby facilitating an assessment of the distribution forecast's accuracy. Additionally, most literature on accounting for uncertainty is related to prediction intervals rather then distributional forecasts, as often this is the final result of interest.

This chapter will first provide some background information on creating prediction intervals and distributional forecasts. After this the method for creating the distributional forecast and prediction intervals will be explained.

3.6.1 Distributional forecasts & prediction intervals

Since the indices we are forecasting are unknown, it can be thought of as a random variable (Hyndman & Athanasopoulos, 2018). For example, the aim could be to forecast the index value of the steel price for next month, or for next year. When forecasting the value of next month, the range of possible values will be a lot smaller opposed to when next year would be forecasted. This range of possible values with associated uncertainty could be expressed by the distributional forecast. The creation of distributional forecasts is an important procedure because decisions may depend on the level of uncertainty (Scott, 2001). When the range of possible future values is very broad, the choice can be made to look for alternatives, or use methods to be robust against these uncertainties. For example if the decision is between a steel frame and a concrete frame, and the uncertainty of the future cost for steel is a lot higher, the choice could be made to use a concrete frame. Similarly a different contract form could be used to cover some of this risk for large cost escalation.

The forecast that we obtain with the selected model represent the mean of this distributional forecast. A time-series forecast consists of at least four sources of uncertainty, as stated by Hyndman and Athanasopoulos (2018): "

1. The random error term;
2. The parameter estimates;
3. The choice of model for the historical data;
4. The continuation of the historical data generating process into the future.

"

Overconfident forecasts often arise from neglecting key sources of uncertainty, as emphasized by (Scott, 2001). Prediction intervals are frequently too narrow, a problem recurrently noted in literature (Hyndman, Koehler, Snyder & Grose, 2002; Hyndman & Athanasopoulos, 2018; Chatfield, 2001). This narrowness primarily stems from omitted uncertainty sources (Hyndman & Athanasopoulos, 2018; Chatfield, 2001). Common practice involves constructing PIs considering only the random error term, using residuals to estimate future forecast errors. However, this approach is limited as out-of-sample forecasts often exhibit greater variance (Makridakis & Winkler, 1989; Chatfield, 1993; Hyndman & Athanasopoulos, 2018) and it assumes normal distribution of residuals, which is not always valid (Hyndman & Athanasopoulos, 2018; Chatfield, 2001). To overcome some of these limitations, while using a method that is applicable to all three timeseries models a method that utilizes the Moving Block Bootstrap (MBB) is used.

3.6.2 Moving Block Bootstrapping

The method utilized was initially proposed by Bergmeir, Hyndman and Benítez (2016). The method combines STL decomposition, Moving Block Bootstrap, and Box-Cox transformation. For this study, Box-Cox is substituted with logarithmic transformation. This substitution was made because applying a Box-Cox transform to every time series in a multivariate model can lead to significant disparities between the time-series values within the model. These disparities are due to differences in the lambda values used in the transformation.

The first step in this method involves STL decomposition, dissecting the time series into seasonality, trend, and residuals. Then, the residuals are transformed into a new series using Moving Block Bootstrapping, which includes: (1) randomly selecting a point in the series, (2) choosing a predetermined number of subsequent data points (block size), and (3) appending this block to the residuals (MBB is visualized in appendix D). This process repeats until reaching the training data length, and it is performed 'n' times (e.g., 1000) to generate n slightly different bootstrapped time series. When considering VAR and VECM models, this method is applied to each series in the model.

The generated bootstrapped time series are then used to fit new models, extending predictions to the planned horizon and producing 'n' distinct forecasts. The selected model resulting from the AFP will be used to fit to the bootstrapped data. Every fitted model will have different parameters resulting in different forecasts. After this process, n amount of forecasts will be available.

The generated forecasts need to be converted into a (1) prediction interval for assessing the quality and (2) a distributional forecast for the implementation in the Monte Carlo simulation. Both of these will be done using the assumption that the bootstraps are normally distributed, which is a common assumption (Hyndman & Athanasopoulos, 2018). This assumption allows us to generate a 95% prediction intervals with formula 3.4 (Hyndman & Athanasopoulos, 2018). In a similar vein, it enables the construction of a normal distribution for the distributional forecasts, where the mean is derived from the forecast of the original series, and the standard deviation is obtained from the variability observed in the bootstrapped forecasts.

$$y_{T+h|T} \pm 1.96 \times \hat{\sigma}_h \tag{3.4}$$

Where:

- $y_{T+h|T}$ is the forecasted value at time $T+h$ given the information up to time T.
- $\hat{\sigma}_h$ is the estimated standard deviation of the forecast error at horizon h.

While ideally the process of selecting a model would be repeated for each bootstrapped series, allowing for the covering of model uncertainty. This process would become too computationally intensive. For this reason the utilized bootstrapping approach attempts to account for the uncertainty coming from the parameters and the random error term (source 1 and 2)(Hyndman & Athanasopoulos, 2018). Note here that source 3 and 4 are not accounted for, most likely still resulting in over-optimistic intervals. Also the assumption of a normal distribution might not be the most accurate representation. these two limitations leave room for improvement within the process of creating PI's and distributional forecasts.

A key benefit of the described method lies in its ability to maintain the statistical properties and patterns of the series, notably stationarity, as outlined by (G. S. H. Li & Maddala, 1996). While it is anticipated that stationarity will be preserved, the retention of Granger causality and cointegration remains unverified to date. This aspect represents an area of potential exploration in future research. Additionally the method can be applied in a relatively similar manner to the Holt-Winters, VAR and VECM model, which allows for straight forward implementation into the Automated Forecasting Process (AFP).

Chapter 4

Selection and validation of automated time series forecasting processes

The primary goal of this chapter is to identify the most suitable set up for the Automated Forecasting Process (AFP) that can later be integrated into the final tool. Where the 'most suitable set up' refers to the model classes that will be considered in the AFP (see 3.4 for the 4 options) and the determining of some key configurations.

The chapter initially details the methodology for model evaluation further, including the performance metrics for AFP selection and the data train test splitting strategy. After this some key configuration of the AFP will be determined for the different model classes. Once the key configurations are established, the model classes are ranked based on performance. After this step the combination of model classes that is expected to perform best will be selected. Finally the accuracy of this configuration will be demonstrated.

4.1 Selection process of the automated forecasting process

For the selection of the best performing AFP and corresponding configurations a grid search will be employed. Each configuration will be tested on different time-frames according to the rolling window train-test split elaborated in subsection 2.4.1. The time-frames of the rolling window splits are visible in table 4.1. The tuning and selection of the AFP will be based on a single performance metric, the interval score (outlined in equation 2.14). Using a single metric, as recommended by (Burkov, 2020), simplifies the specification tuning and AFP selection process. The interval score was selected as this metric because it effectively captures both the accuracy of the point forecast and the quality of the prediction interval. Accurate point forecasts naturally lead to reliable intervals, making the interval score a comprehensive measure of forecast performance. The prediction intervals constructed in this study are 95% confidence intervals. Two configurations will be optimized the training size and the block size of the bootstraps.

Training size will be optimized as older data might become less useful due to structural changes in the system (Hyndman & Athanasopoulos, 2018). For this reason a long training size might capture data that is no longer relevant, while a shorter training size might not be enough to capture the true underlying pattern. For this reason, finding an optimal size is important. The AFP's will be tested with a training size of 5, 10 and 15 years.

The block size, which is relevant within the Moving Block Bootstrap (MBB), is a configuration that determines the size of consecutive data points to be re-sampled. The optimal block size remains a subject of debate, as evidenced by various studies (G. S. H. Li & Maddala, 1996; *Carlstein1986*, n.d.; Kunsch, Ktnsch & Zurich, 1989; Hall & Horowitz, 1993; J. Li, 2021; Bergmeir et al., 2016). Bergmeir et al. (2016) recommended a 24-month block size for an exponential smoothing model to preserve any residual seasonality. Similarly, (J. Li, 2021) advise using large blocks in blocked bootstrapping for a VAR model. *Carlstein1986* (n.d.) observes that increasing the block size reduces bias but increases variance. Ambiguity remains due to the three different time-series models used (Holt-Winters, VAR, VECM). To address this, the research explores block sizes of 6, 12, 18, and 24 months to determine their impact on the accuracy of prediction intervals.

The difference in configuration in block size and training size will be tested on each different split. After which, for the forecasting horizons of 1, 2, 3 and 4 years the Interval score is calculated. This allows for selecting the optimal configuration through a grid search. While in this study only the training size, and block size is varied. This process can be extended for finding the optimal configuration for many other aspects e.g. method for data processing, incorporation of dummy variables, information criterion used. Due to the scope of this research this exploration will remain limited.

Table 4.1: Training and testing periods for different splits

Split	Train Period	Test Period
Split 1	1993-01-01 - 2007-12-01	2008-01-01 - 2011-12-01
Split 2	1994-01-01 - 2008-12-01	2009-01-01 - 2012-12-01
Split 3	1995-01-01 - 2009-12-01	2010-01-01 - 2013-12-01
Split 4	1996-01-01 - 2010-12-01	2011-01-01 - 2014-12-01
Split 5	1997-01-01 - 2011-12-01	2012-01-01 - 2015-12-01
Split 6	1998-01-01 - 2012-12-01	2013-01-01 - 2016-12-01
Split 7	1999-01-01 - 2013-12-01	2014-01-01 - 2017-12-01
Split 8	2000-01-01 - 2014-12-01	2015-01-01 - 2018-12-01
Split 9	2001-01-01 - 2015-12-01	2016-01-01 - 2019-12-01

4.2 Results: Holt-Winters automated forecasting process

Based on the results of the different configurations of the AFP, the different train-test splits and the various variables predicted, two tables are created. Table 4.2 delineates the optimal training duration and the ideal block size for forecasting prediction intervals across different variables. This table also presents the average interval score, which is calculated over all train-test splits and across different forecasting periods (1, 2, 3, and 4 years) for each variable. Table 4.3 shows the average interval score for each forecasting horizon. It becomes evident that the optimal training sizes and block sizes vary depending on the variable being forecasted. Notably, more volatile indices such as Steel, Copper, and Asphalt exhibit higher interval scores, indicating less precise intervals. As anticipated, forecasts extending over longer periods tend to be less accurate.

Regarding the configuration of the models, there is no single best approach that applies universally to all variables. Although longer training periods of 120 or 180 months are generally preferred, there remains some uncertainty in choosing between models. In terms of block size, the only consensus is that a shorter block size of 6 months has not been selected by any model.

Table 4.2: Holt-Winters: AFP results overview

Variable	Model Type	Train Size	Block Size	Average Interval Score
PPI: Steel	Holt-Winters	120	18	424.7
PPI: Concrete	Holt-Winters	180	12	78.1
construction wage	Holt-Winters	180	12	53.4
PPI: Copper	Holt-Winters	60	18	743.0
PPI: asphalt	Holt-Winters	120	24	275.6
PPI: sand	Holt-Winters	120	24	58.6

Table 4.3: Holt-Winters: results average interval score over the different forecasting horizons

Variable	12-Month Horizon	24-Month Horizon	36-Month Horizon	48-Month Horizon
PPI: Steel	189.8	403.8	552.8	584.7
PPI: Concrete	22.8	51.4	93.7	144.7
construction wage	20.7	36.9	62.8	93.2
PPI: Copper	405.3	659.4	878.0	1029.4
PPI: asphalt	204.5	263.9	299.8	334.1
PPI: sand	13.1	36.8	72.1	112.5

4.3 Results: VAR automated forecasting process

The AFP that utilized the VAR model had 2 key requirements that needed to be met before the VAR model can be utilized: the existence of Granger causality between the explained and explanatory variables, and the necessity for the data to achieve stationarity after applying natural logarithm transformations and differencing (detailed in section 3.3). From the results tables it is clear that these requirements where not met for the explained variables concrete and sand. Comparing the results with the previous Holt-Winters AFP, it is apparent that the interval scores for the VAR model are generally less favorable across all variables, with the exception of construction wage, which shows a slight improvement. Consistent with previous findings, a preference for longer training and block sizes is evident. Notably, the interval scores for steel and copper are particularly poor, suggesting that the VAR AFP has not successfully captured the underlying model for these variables.

Table 4.4: VAR: AFP results overview

Variable	Model Type	Train Size	Block Size	Average Interval Score
PPI: Steel	VAR	180	24	615.2
construction wage	VAR	120	6	50.0
PPI: Copper	VAR	180	24	1439.3
PPI: asphalt	VAR	180	24	383.0

Table 4.5: VAR: results average interval score over the different forecasting horizons

Variable	12-Month Horizon	24-Month Horizon	36-Month Horizon	48-Month Horizon
PPI: Steel	322.8	681.6	706.4	750.2
construction wage	15.5	38.7	61.5	84.2
PPI: Copper	569.4	1270.1	1768.7	2149.1
PPI: asphalt	76.6	259.9	476.1	719.4

4.4 Results: VECM automated forecasting process

The VECM model, an extension of the VAR model, shares similar prerequisites, such as stationarity (in VECM, this is manifested as having the same order of integration) and Granger Causality. Additionally, the VECM model requires a cointegrated relationship among the variables in the model. The explanatory variables that meet these criteria for model selection in the VECM AFP are the same as those in the VAR AFP. For the eligible variables, each appears to significantly outperform the preceding VAR model. In terms of optimal performance, both the block and training sizes tend to be longer. Notably, the VECM model for construction wage is particularly effective, yielding highly accurate prediction intervals.

Table 4.6: VECM: AFP results overview

Variable	Model Type	Train Size	Block Size	Average Interval Score
PPI: Steel	VECM	120	24	177.3
construction wage	VECM	120	12	16.7
PPI: Copper	VECM	120	18	650.8
PPI: asphalt	VECM	180	12	250.4

Table 4.7: VECM: results average interval score over the different forecasting horizons

Variable	**12-Month Horizon**	**24-Month Horizon**	**36-Month Horizon**	**48-Month Horizon**
PPI: Steel	76.4	227.6	205.1	199.9
construction wage	4.4	5.6	18.7	38.0
PPI: Copper	553.8	638.8	729.1	681.4
PPI: asphalt	40.8	83.8	173.3	703.7

4.5 Selection of the automated forecasting process

For the selection of the final configuration of the AFP, the four options listed initially in table 3.4 are compared. Before being able to make this comparison, the optimal training size and block size needs to be determined for each model class. In an ideal scenario, each variable would be processed using its best-performing configuration. However, from a practical standpoint, programming and integrating such a tailored approach into the tool is unfeasible given the scope of the research. For this reason the overall best performing configuration of each model class will be selected. Consequently, the optimal combination of training size and block size for the VECM, VAR and Holt-Winters AFP's is determined by evaluating the average interval score across all combinations of training sizes and block lengths. Interestingly, this analysis yielded the same optimal block size for all three model classes, being 24 months. The optimal training size, for the Holt-Winters and VECM model are both selected at 10 years, while the training size for the VAR model is optimal at 15 years.

Now that the configurations of the block size and training size have been determined, the four options for combinations are evaluated using this configuration. This comparison is visible in table 4.8. From the table it becomes evident that the AFP that only utilizes the Holt-Winters model performs best in terms of interval score. It can be noted that for the variables Concrete and sand the result are always the same since the Holt-Winters fallback model is always utilized. Additionally it is visible that incorporating the VAR model produces in the poorest results. Finally the interval scores for the copper and steel remain high, indicating that a model that captures the underlying system has not been found, for neither of the combinations. The rest of this research will utilize the AFP that only uses the Holt-Winters model.

Table 4.8: Comparison of average Interval Scores

Variable	**VECM, VAR, HW**	**VAR, HW**	**VECM, HW**	**HW**
PPI: Concrete	97.4	97.4	97.4	97.4
PPI: Copper	1071.7	1439.3	642.4	800.7
PPI: Steel	565.4	628.7	444.3	428.6
PPI: asphalt	692.3	509.3	692.3	275.6
PPI: sand	58.6	58.6	58.6	58.6
construction wage	61.5	71.2	55.1	57.9
Average of variables	**424.5**	**467.4**	**331.7**	**286.5**

4.6 Performance of the selected automated forecasting process

In evaluating the effectiveness of the Holt-Winters AFP, several key performance metrics are utilized. Mean Absolute Percentage Error (MAPE), defined in equation 2.12, measures the accuracy of the percentage change predictions. To identify any consistent bias in the forecasts, a bias metric (as detailed in equation 2.13) is applied, revealing whether the model tends to systematically overestimate or underestimate values. Additionally, interval coverage is examined to compare the aimed coverage with the actual coverage achieved by the model.

4.7 Overall performance

Table 4.9 presents a comprehensive summary of the performance metrics across various variables. These metrics are averaged over four different testing set lengths (1, 2, 3, and 4 years), with detailed results for each length available in Appendix B. The average MAPE of all variables remains below 10. It is evident from the table that the model's performance on copper is notably poorer compared to other variables, exhibiting a significant bias of 12, which indicates a systematic overestimation of steel prices. Asphalt also displays a bias, in this case, an underestimation of its price.

A key point of analysis in the model's performance is the interval coverage, which significantly diverges from the aimed 95% intervals. This divergence is particularly evident in the forecasting of construction wages. Despite demonstrating favorable interval scores, the coverage for construction wages is notably low. This pattern suggests that the forecasted intervals for construction wages are quite narrow, resulting in actual values frequently falling outside these predicted ranges. However, it's important to note that when actual values do deviate from these narrow intervals, the extent of deviation is minimal, which contributes to the favorable interval scores. This indicates that while the model is precise in its predictions for construction wages, there is a tendency for actual values to marginally exceed the forecasted range, highlighting an area for potential refinement in the model's predictive accuracy.

The forecasts for concrete, sand, and construction wages are relatively strong across most metrics. These results indicate a robust performance in these areas, highlighting the model's ability to accurately predict changes in these specific variables. However, the varying performance across different variables and the issues with interval coverage and bias in certain areas suggest areas for further refinement and improvement in the forecasting model.

Table 4.9: Summary of metrics for various variables

Variable	Average MAPE	Average Interval Score	Average Coverage	Average Bias
PPI: Concrete	2.9	97.4	59.8	0.2
PPI: Copper	20.7	800.7	46.2	12.0
PPI: Steel	15.0	436.7	60.1	-2.1
PPI: asphalt	5.5	275.6	71.2	-5.8
PPI: sand	1.5	58.6	65.9	0.8
construction wage	1.4	57.9	39.4	0.3
Average of variables	**7.8**	**287.8**	**57.1**	**0.9**

4.7.1 Performance of the different train-test splits

The investigation into the performance of different train-test splits is depicted in figure 4.1. This figure presents the starting date of the testing set on the horizontal axis, with each testing set spanning 12 months. On the vertical axis, both the interval score and the MAPE are displayed. A notable observation is the diminished performance for tests beginning in 2008 and 2009, with the year 2009, immediately following the financial crash, showing a significant spike in both MAPE and interval score across nearly all variables. Apart from this decline in 2009, the results demonstrate overall stability.

Further confirmation of this general stability is provided by Figure 4.2, which aggregates the interval scores and MAPEs across all results. This distribution indicates that the majority of observations are characterized by low interval scores and MAPEs, with only a small fraction deviating significantly. The pattern of this distribution appears to align with a Pareto distribution, suggesting a concentration of higher errors in a relatively small number of cases.

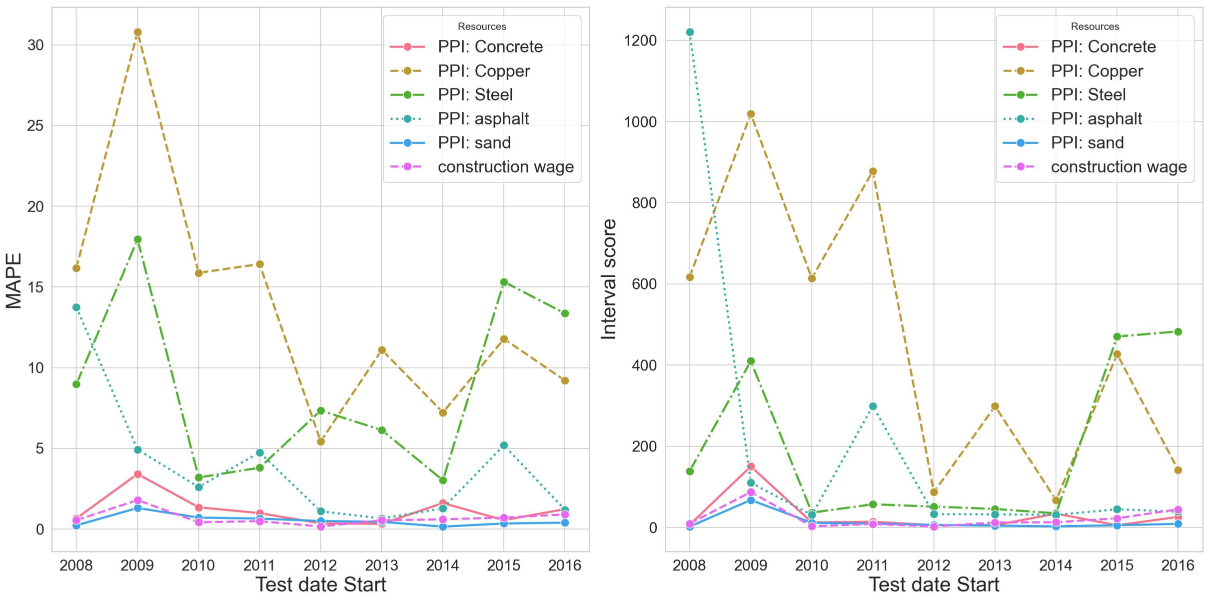


Figure 4.1: Performance of the different train-test splits (length of testing set is 12)

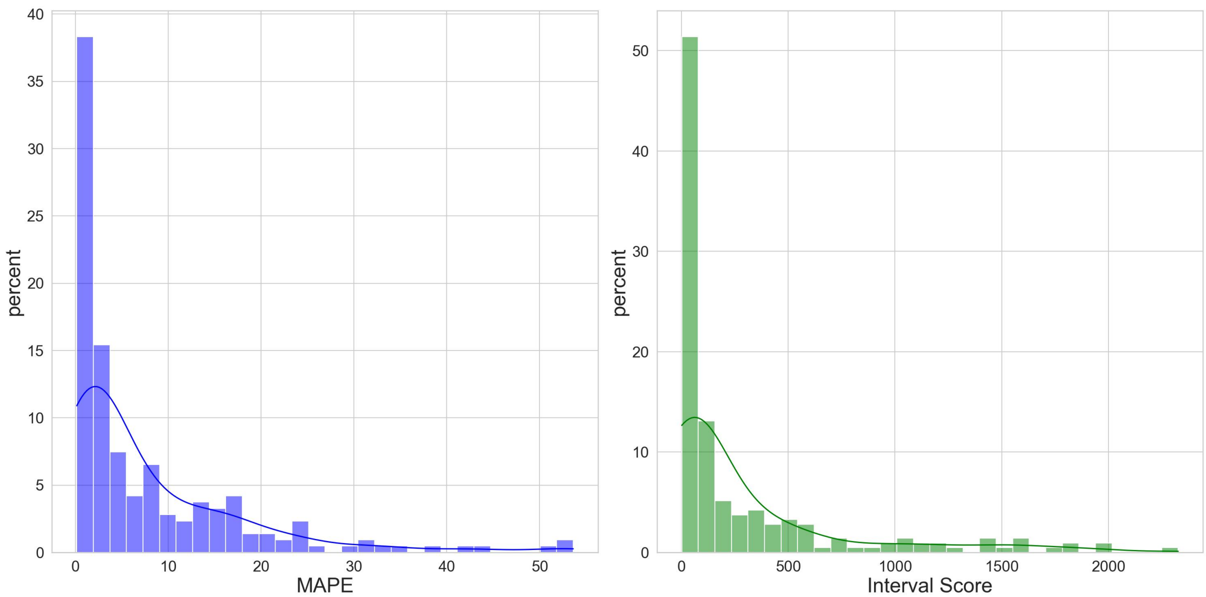


Figure 4.2: Distribution of the MAPE and interval score over all results

Chapter 5

Development of the tool: integrating change of resource price with the base estimate

This chapter focuses on integrating the selected Automated Forecasting Process (AFP) with the methodology for modeling the base costs, which will be introduced in this chapter. The initially proposed conceptual overview of the tool displayed in figure 1.1 will be expanded into a detailed final version of the tool's architecture. A demonstration of the tool is not covered in this chapter; instead, it will be explored in the following chapter 6.

Initially, the method for modeling project resource costs will be detailed. This will be followed by presenting the architecture of the final tool. The architecture will incorporate the AFP, previously demonstrated, along with the modeling of project resource costs. Lastly, the formulation of the Monte-Carlo simulation, including its mathematical process, will be detailed.

5.1 Modelling the base cost

The final cost escalation is dependent on both the base cost of that resource, and the change in the resource over a specific time frame (as can also be derived From formula 1.1). For this reason, capturing the base cost of a resource is an important aspect for the modeling of escalation. To adequately capture the base cost, the possible variation in construction cost needs to be taken into account. The proposed tool will incorporate the cost uncertainty by sampling from a Probability Density Function (PDF) that represents the potential deviations in base cost of the activities. By integrating this component, the tool captures the variation in realized monthly cost.

In order to model this cost variation an appropriate distribution needs to be selected. The conventional use of a normal distribution for this purpose is recognized and has been adopted in various studies, including those by (Touran et al., 2006; wei Xu & Moon, 2013). Nonetheless, the often skewed nature of project costs, characterized by a definitive lower bound and a more variable upper bound, necessitates the exploration of alternative distributions.

Several distributions have been investigated for better alignment with the skewness in project costs. These include the Triangular, PERT , and Lognormal distributions. Research indicates that the Triangular distribution tends to overestimate project costs, despite its ease of acquisition through project manager estimates on likely, pessimistic, and optimistic costs (Chau, 1995; Chou, 2011). In contrast, (Sonmez, 2004) highlights the PERT distribution as a best fit, whereas (Wall, 1997) argues for the Lognormal distribution based on his findings. However, (Wall, 1997) also points out that the impact of excluding correlations in cost data is more significant than the choice between the Lognormal and PERT distributions, suggesting a less pronounced difference between these options. For illustration purposes the triangular, PERT, lognormal and normal distribution are illustrated in figure 5.1

For the purpose of this study, the PERT distribution has been chosen, which in essence is a transformed beta distribution. This distribution offers both flexibility and ease of use, making it a suitable choice for adequately modeling project costs in Monte Carlo simulations. This project cost can be modeled by estimating the Optimistic (a), Most-likely (m), and Pessimistic (b) cost estimates for a project task. This is done by transforming these three values to a Beta distribution by calculating the α and β parameter as follows:

$$\alpha = 1 + 4 \times \frac{m-a}{b-a} \tag{5.1}$$

$$\beta = 1 + 4 \times \frac{b-m}{b-a} \tag{5.2}$$

Random costs can then be generated based on the transformed Beta distribution where the optimistic and pessimistic value are used for scaling:

$$\text{Random Costs} = a + (b-a) \times \text{Beta}(\alpha, \beta) \tag{5.3}$$

Where:

- a is the Optimistic cost estimate.
- m is the Most-likely cost estimate.
- b is the Pessimistic cost estimate.
- α and β are shape parameters for the Beta distribution.
- $\text{Beta}(\alpha, \beta)$ is the sample from the beta distribution.

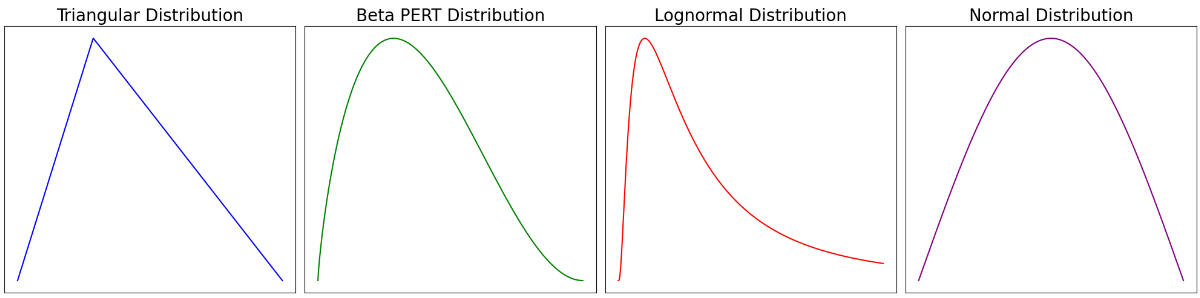


Figure 5.1: Illustrative plot of cost sampling distributions

5.2 Required project estimates

For the use of the tool a comprehensive input needs to be developed that allows for modeling the expected base cost of each resource. The design of the user input emphasized flexibility in the level of detail, allowing for estimates instead of insisting on precise values for all material costs. This led to the input format presented in table5.1. The developed input encompasses certain simplifications: it differentiates solely between material and labor costs, omitting equipment costs present in actual construction projects. Additionally, it utilizes a predetermined set of materials (outlined in 3.1.1), which might not cover every specific material required. Furthermore, it operates under the assumption that the cost of materials scales linearly to the total activity cost. This implies that if the cost of an activity is 10% higher, the cost of any material within that activity will also increase by 10%. The advantage of these simplifications lies in their elimination of the need for detailed information, thereby providing an approximation accurate enough to produce valuable output.

Beyond these simplifications designed for ease of input, there is a constraint. This limitation is that it only accommodates activities occurring in parallel, rather than activities in series. This simplification is due to the proof-of-concept objective for the tool. The specifics of each input category are elaborated further below:

- **ID:** unique ID of each of the activities
- **Object:** name of the activity
- **Duration:** a whole number that should tell the duration from the start of the activity until at what point the escalation should be calculated.
- **Optimistic, most likely and pessimistic value:** these are the inputs for creating the beta sampling distribution of that specific activity.
- **% cost due to labour, % cost due to n:** The percentage cost of that resource that contributes to the total activities cost of the initial estimate. This allows for calculating the cost of each resource dedicated to a specific activity.

Table 5.1: Project Cost Estimates and Contributions (including some example input)

id	activity	Duration (months)	Predecessor	Optimistic cost	Most-likely cost	Pessimistic cost	% cost due to Labor	% cost due to	% cost due to n
1	Sample activity 1	6		€600.000	€750	€1000	50%	30%	20%
...	Sample activity 2	...	..	...	...	...	...	...	...
n	Sample activity n	...	n-1	€...	€...	€...	...%	...%	...%

5.3 Architecture of the final tool

As a final step both the AFP for the indices and the distribution of the estimated base cost need to be integrated. In order to integrate both components a Monte Carlo simulation is utilized.

A Monte Carlo simulation is a computational technique that uses random sampling to model and analyze systems. The method involves running a large number of simulations with different random inputs to calculate a range of possible outcomes for a given scenario. After aggregating these outcomes, inference can be made on statistical properties (such as mean, variance) that help in understanding the likelihood of various results and in making informed decisions under uncertainty.

Integrating the AFP with the format of the project base cost results in figure 5.2, which is a concise version of the full architecture of the tool. The full architecture can be found in appendix C. It consists of two data inputs: one for generating a PDF for sampling the percentage change in the indices that represent the project resources, and another for sampling the base cost of the activities. The Monte Carlo Simulation will sample from both distributions. After the samples are pulled the needed calculations are done to translate the samples into the total cost and escalation of the various resources. Below a brief summary of the most important components are provided.

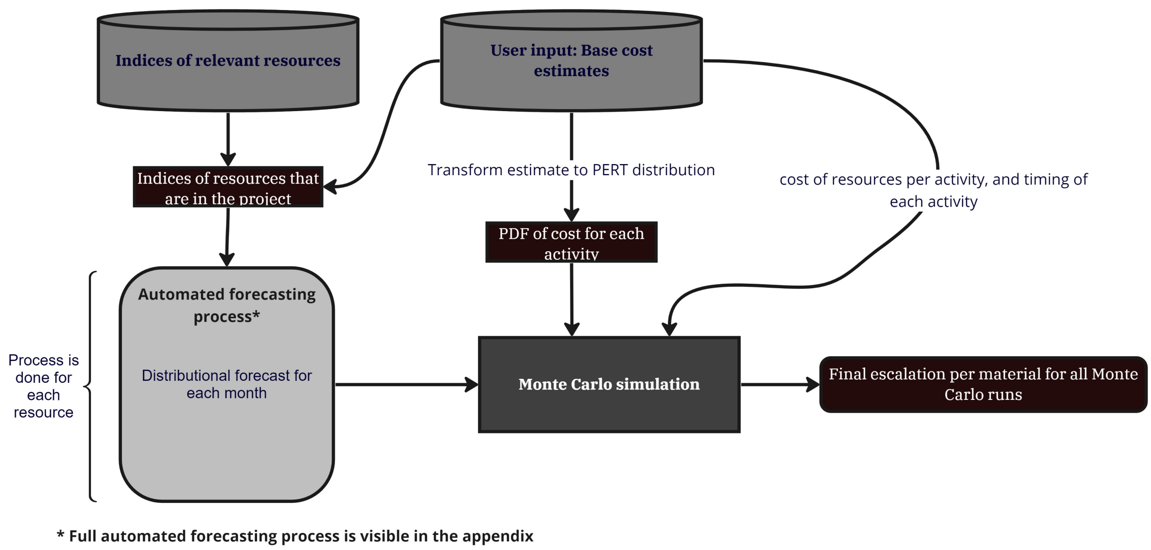


Figure 5.2: Tool architecture (concise), for complete version see the appendix

1. The **base cost estimates** is the user input which should accommodate different types of construction projects including the corresponding resources. It facilitates the calculation of uncertainty within the project, converting cost of the various activities into a PDF for sampling in the MCS.
2. The second data source consists of **historical data of the indices** needed for modeling the change in a projects resource prices. This data source are the indices selected in section 3.1.1.
3. The **PDF of base cost for each activity** is designed to capture the uncertainty of expenses of each object or activity, and with this capture the uncertainty in resource cost. This is achieved by converting the optimistic, most likely and pessimistic estimate of each object into the PDF of a PERT distribution.
4. The **Distributional forecast for each month** aims to model the anticipated monthly changes in the prices of resources, such as labor, steel, and concrete. This forecast is obtained by using the developed AFP, which uses the Holt-Winters model.
5. The **Monte Carlo Simulation** integrates the PDF project costs with the resource's PDF. An elaborate description of the calculation used in the Monte Carlo simulation is given in section 5.4.2.

It can be noted that the overall architecture and the 5 key components are in line with the Research objectives and the provided Conceptual overview of the tool .

5.4 The Monte Carlo simulation set-up

This section will elaborate further on the application of the Monte Carlo simulation. The Monte Carlo simulation allows for combining the uncertainty in base cost and uncertainty in the prediction of the change in resource price. Additionally some underlying calculations are done that account for the project timeline and material cost associated with each project activity.

First the method for sampling is outlined, after this the underlying calculations in the Monte Carlo simulation are detailed.

5.4.1 Sampling from forecasted time-series

In the context of future escalation sampling, two principal methodologies were explored at a high level. These methods were analyzed by examining how realistic the future paths will be and the consequent implications for predictive accuracy. Furthermore, the computational intensity of each approach, along with the associated time required for making predictions. Additionally, the overall complexity of the methodologies was examined to ensure a practical equilibrium between the expended effort and the derived utility.

The first method involves a step-wise prediction process: predicting a single time-step ahead, inclusive of upper and lower bounds, followed by sampling from this prediction for the subsequent time-step. This iterative procedure is repeated until the desired prediction horizon is achieved. The primary advantage of this approach is its ability to produce a plausible trajectory for the predicted time-series. However, a significant drawback is the expected computational intensity, which is intensified due to the chosen blocked bootstrapping technique for estimating the confidence interval.

This computational demand can be demonstrated with a simple numerical example. For this example the following properties are selected for the simulation: forecast horizon (H) of 12 months, wherein each month involves 500 bootstrap samples (B) for estimating the bounds, and 10,000 Monte Carlo simulation runs (M). The total number of model fittings (N_{fit}) is calculated as follows:

$$N_{\mathrm{fit}} = H \times B \times M = 12 \times 500 \times 10\,000 = 60 \times 10^6 \tag{5.4}$$

Considering the inherent computational demand of model fitting, this approach's nonlinear increase in computation time is sub optimal. The implementation complexity is also expected to be elevated due to the needed tight coupling between the Monte Carlo simulation and the time-series prediction model.

The second method, which was ultimately chosen, involves estimating the mean and standard deviation for the entire forecast horizon. In this approach, future values for the entire horizon are forecasted in a single computation. Subsequently, during the Monte Carlo simulation, the mean and standard deviation for the last month's forecast is sampled and the observation in between are interpolated linearly. This methodology effectively eliminates the H term from the equation reducing the frequency of model fittings, thereby significantly reducing computational time. A notable limitation, however, is the potential for less realistic future paths, as samples are only taken for the final observation. Nevertheless, the reduced complexity—stemming from a looser coupling between the time-series prediction and the Monte Carlo simulation—renders this method more advantageous and thus preferred.

5.4.2 Integration of time-series sample with the base cost samples

Now that the method of sampling from the escalation, and the method of sampling from the project cost has been discussed, both components can be brought together. The final simulation needs to account for the moment an activity is executed and for the specific resources used for that activity. Together this will give the project specific escalation.

The Monte Carlo simulation's methodology involves four steps for each resource in a single simulation run. These steps are essential for understanding the total cost escalation. Familiarity with the format of the project input, as outlined in Table 5.1, is recommended for effectively following the example.

Vector representations are used in the example to clarify the calculations. While the accompanying explanation aims to be accessible to a diverse audience. The resource escalation is calculated at the end of each activity, based on the assumption that the escalation index remains valid up to that point. In practice, the actual timing might vary, but for the purposes of the example, this does not matter.

The four steps below are replicated for each resource and every run of the Monte Carlo simulation. In the final tool, not only the escalation for each resource is tracked, but also the overall project cost and the individual cost for each resource.

STEP 1 generating a random sample of each activity's base cost (sampling from the PERT distribution) and the future escalation (sampling from the distributional forecast of each month). Additionally the percentage of resource cost per activity is retrieved. This will result in vector representations of the following format:
For the sample of the base cost:

$$\mathbf{c}_{\text{random cost}} = \begin{bmatrix} c_1 & c_2 & \cdots & c_{\text{N}} \end{bmatrix} \tag{5.5}$$

For the percentage of resource cost per activity, per resource:

$$\mathbf{p}_{\text{material cost \%}} = \begin{bmatrix} p_{1,i} & p_{2,i} & \cdots & p_{\text{N,i}} \end{bmatrix} \tag{5.6}$$

For the sample of project escalation for a specific resource the last months forecast value is sampled:

$$\mathbf{e}_{\text{prediction month T},i} = \begin{bmatrix} e_{\text{T}} \end{bmatrix} \tag{5.7}$$

This forecast is interpolated to get the predicted index of all months:

$$\mathbf{e}_{\text{predicted index},i} = \begin{bmatrix} e_0 & e_1 & \cdots & e_{\text{T}} \end{bmatrix} \tag{5.8}$$

Where:

- $\mathbf{e}_{\text{predicted index},i}$ represents the sampled index vector for resource i , with each element e_t corresponding to the resource index at month t.
- $\mathbf{c}_{\text{random cost}}$ is the vector representing the sampled cost of all activities. Each element c_n corresponds to the cost of the n^{th} activity.
- $\mathbf{p}_{\text{resource cost \%}}$ represents the percentage of resource cost per activity. Each element p_n corresponds to the percentage of the total cost that is attributed to the resource i for the n^{th} activity.
- N is the total number of activities/elements.
- T is the total number of months.
- i represent the different resources in the project

STEP 2 After selecting the samples, the index change between the start of the project and the moment of escalation, in this case at the end of an activity, is calculated. This calculation is represented by:

$$\Delta e_{n,i} = \frac{e_{t_n} - e_0}{e_0} \tag{5.9}$$

Where:

- $\Delta e_{n,i}$ denotes the percentage change in the index of the resource i from the start of the project to the end of activity n.
- t_n is the month number corresponding to the end of activity n,
- e_0 is the escalation index at the start of the project.

STEP 3 Calculate the escalated cost of the material for each activity by applying the percentage of material cost to the total cost of the activity, and then adjusting for escalation. This is represented by:

$$m_{n,i} = c_n \cdot p_{n,i} \cdot \Delta e_{n,i} \tag{5.10}$$

Where:

- $m_{n,i}$ represents the escalation of the resource i for activity n.
- c_n is the total cost of activity n, an element from $\mathbf{c}_{\text{random cost}}$.
- $p_{n,i}$ is the percentage of the cost attributed to a resource i for activity n, from $\mathbf{p}_{\text{material cost }\%}$.
- $\Delta e_{n,i}$ is the escalation factor for activity n.

STEP 4 Sum up the escalated resource costs for all activities to get the total escalated cost of the resource i. This is represented by:

$$E_{\text{total,i}} = \sum_{n=1}^{N} m_{n,i} \tag{5.11}$$

Where:

- $E_{\text{total,i}}$ represents the total escalation for that resource (i) for the entire project.
- m_n is the escalation of that resource for each activity n.

Chapter 6

Tool application: highway construction project

This chapter provides a demonstration of the developed tool, using data from a cost breakdown structure of a highway project as input. The primary aim of this demonstration is to showcase the functionality of the tool, and how to use it. It's important to note that the focus here is not on the validity of the input or the output.

The chapter will start by outlining the input required and the steps for using the tool. Following this, the various outputs of the tool will be examined using the highway project for demonstration. The process by which these outputs are derived from the initial Monte Carlo simulation will be detailed, and guidance will be provided on how to make decisions based on this output.

6.1 User data input: base cost estimates

This section will present both the interface designed for inputting project data and the specific project data utilized in the tool's demonstration.

6.1.1 Input interface

To enhance user accessibility, a hybrid data input method combining a web interface and an Excel file has been adopted. This approach leverages the familiarity of Excel for numerical data entry, supported by a web interface that facilitates the hosting of the model. The web interface enables users to download an Excel template for data input, which includes a tab with instructions for data input.

Users input their data in the Excel sheet, similar in format to the table shown in table 6.2. Once the Excel sheet is completed, it is uploaded via the web interface (illustrated in figure 6.1). A feature of this interface is that it allows for controlling the time over which the escalation of an activity is calculated. For instance if all materials for a steel frame are purchased at the project's start this can be reflected.

the user can select the moment until when the materials and labor of the corresponding activity are escalated. To elaborate, one might buy all materials for building the steel frame at the very start of the project, this can then be reflected in this input. The final input is then displayed on the web interface, as shown in figure 6.2. This is visualized, with an interactive chart, that shows the corresponding information of each activity (created with (Inc., 2024)). Subsequently, users can execute the escalation tool by clicking the designated button, triggering the generation of output (detailed in subsequent chapters). The web interface is developed using the Python package Streamlit (*Streamlit*, 2024), chosen for its simplicity in creating user interfaces for Python-based models.

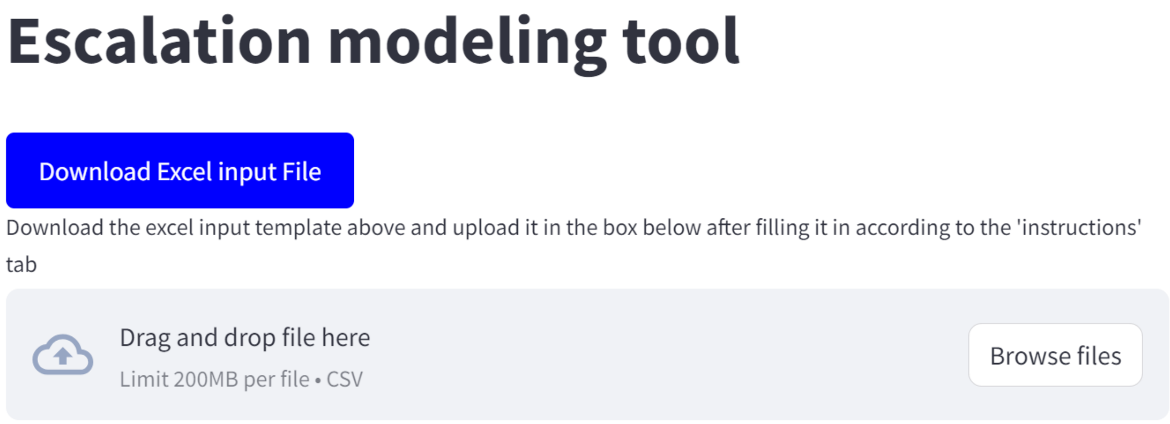


Figure 6.1: User interface: downloading and uploading project data Excel

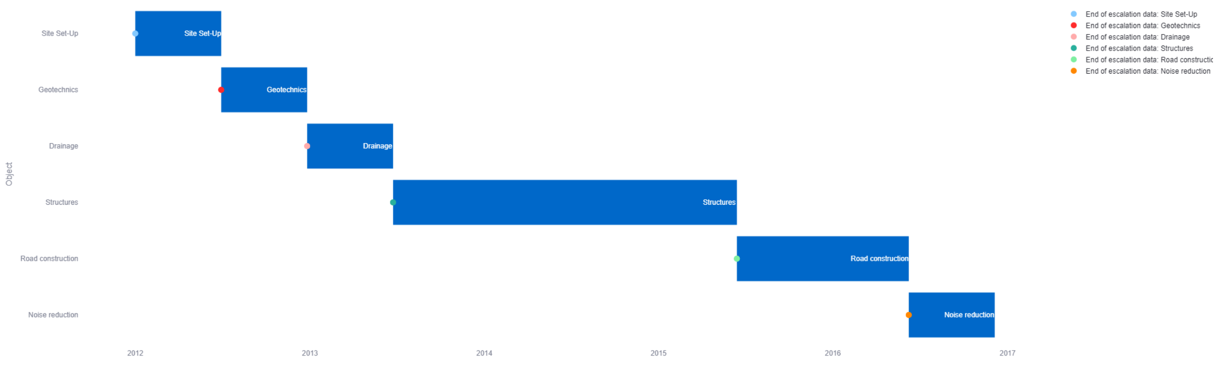


Figure 6.2: Timeline of uploaded project data, including markers for moment until when activity resources are escalated. In this example the escalation of the resources is calculated until the start of each activity.

6.1.2 User input: highway project

Using the previously described user-interface the input for the highway project can be filled into the Excel sheet. Although the actual Excel sheet comprises a single table, it has been divided for illustration purposes. Table 6.1 displays the optimistic, most likely, and pessimistic estimates of the project activities at current day value. These estimates are converted into a PERT distribution. The table also shows that all activities are sequenced in series, as indicated in the predecessors column. In reality there will be overlap in these, activities. For the highway project, all materials have been simplified to include only steel, concrete, sand, and asphalt (visible in table 6.2). Assuming that this will cover the large majority of materials. From the input it is visible that the project is expected to have a high labor intensity.

Table 6.1: Project cost estimates: highway project

ID	Object	Duration (Months)	Predecessor	Optimistic Cost	Most-Likely Cost	Pessimistic Cost
1	Site Set-Up	6	-	$30,384	$33,760	$37,136
2	Geotechnics	6	1	$38,250	$45,000	$54,000
3	Drainage	6	2	$13,500	$15,000	$17,250
4	Structures	24	3	$57,600	$72,000	$100,800
5	Road Construction	12	4	$73,600	$92,000	$115,000
6	Noise Reduction	6	5	$25,500	$30,000	$34,500

All numbers in thousands

Table 6.2: Cost per resource: highway project

id	Object	% Cost due to Labor	% Cost due to Steel	% Cost due to Concrete	% Cost due to Sand	% Cost due to Asphalt
1	Site Set-Up	100%	-	-	-	-
2	Geotechnics	80%	-	-	10%	-
3	Drainage	100%	-	-	-	-
4	Structures	70%	10%	20%	-	-
5	Road construction	65%	-	-	-	35%

6.2 Output: example highway project

After running the Monte-Carlo simulation, a large amount of outputs are generated. Each output represent a possible outcome, the more often a similar outcome is generated the more likely it is that an outcome as that will happen. For this reason it is relevant to look at the spread in the different outputs and not only consider the mean of all outputs.

The output will be split up in three aspects:

1. The total project cost
2. The total cost of escalation
3. The escalation of the different resources

The output, as common with Monte Carlo simulations, will be visually represented by a histogram. This way all outcomes and their corresponding frequency are visible. Additionally the outputs will be given numerically, by supplying the mean, standard deviation, Interquartile range and the Coefficient of Variation (CV). The CV ratio is the standard deviation divided by the mean. Care needs to be taken when interpreting as negative values, or values close to zero can be misleading.

6.2.1 Total project cost and total project escalation

The tool provides global output on the cost of the whole project: the total project cost and the total project escalation. The total project cost reflects the expected expenses for the entire project, factoring in escalation costs. Meanwhile, the total escalation represents the cumulative escalation across all resources. Notably, the histogram depicting the total project escalation shows a considerable spread. The mean escalation is estimated at around 5.8 million dollars, with a standard deviation of 1.5 million. For estimators aiming for a conservative approach, it may be prudent to consider the higher end of the interquartile range, which stands at 6.8 million dollars. This figure provides a more cautious estimate, accommodating potential uncertainties in the project's escalation costs.

Table 6.3: Descriptive statistics for total project cost and total sscalation

Statistic	Total Project Cost	Total Escalation
Mean	\$297,316	\$5,826
Standard Deviation	\$12,216	\$1,460
1st Quartile (25%)	\$288,679	\$4,823
3rd Quartile (75%)	\$305,822	\$6,797
Coefficient of Variation	0.041088	0.250571
All numbers in thousands		

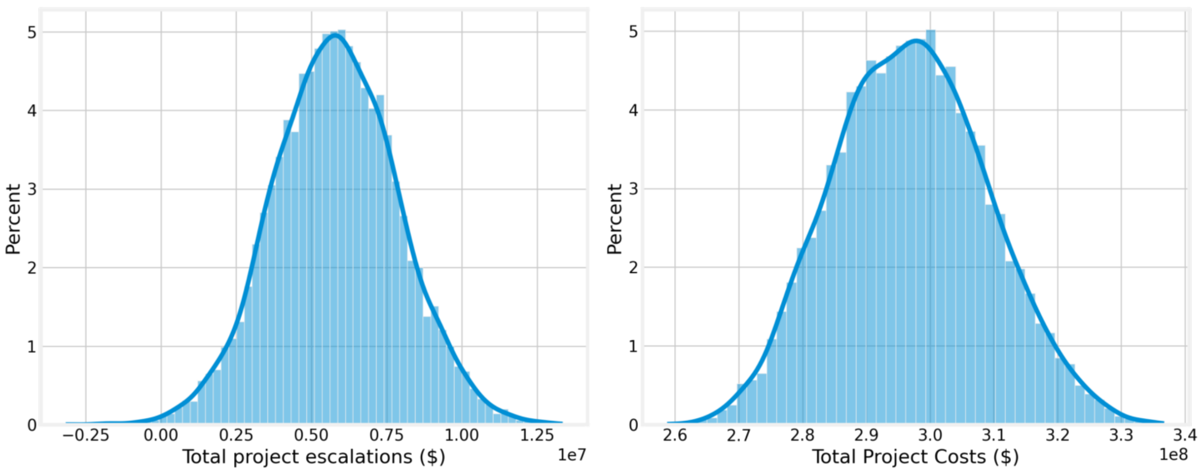


Figure 6.3: Histogram of total project cost and total project escalation

6.2.2 Escalation of various resources

The second output from the tool is the detailed escalation of the various resources modeled. This detailed breakdown is a significant added value of the tool, offering insights into material-specific escalations that generic indices cannot provide. For instance, the Coefficient of Variation (CV) ratio reveals that steel exhibits the highest volatility among the materials. This insight suggests that careful planning and timing in procuring steel can mitigate risks of price increases. A similar observation applies to concrete, where proactive procurement strategies could be beneficial. Although asphalt's CV ratio may not fully capture its volatility due to negative values, its large standard deviation is notable. For financial planning purposes, using the interquartile range to estimate future cash flows for these materials can provide a more reliable guide for budgeting and risk management.

Table 6.4: Descriptive statistics for escalation of various resources

Statistic	Labor	Steel	Concrete	Sand	Asphalt
Mean	$4,523.6	$30.7	$120.2	$29.4	$1,122.5
Standard Deviation	$1,374.0	$51.1	$93.5	$12.3	$431.5
1st Quartile (25%)	$3,590.0	-$3.5	$57.7	$21.2	$824.7
3rd Quartile (75%)	$5,426.7	$64.6	$180.9	$37.7	$1,406.9
Coefficient of Variation	0.303743	1.663532	0.777572	0.419484	0.384369

All numbers in thousands

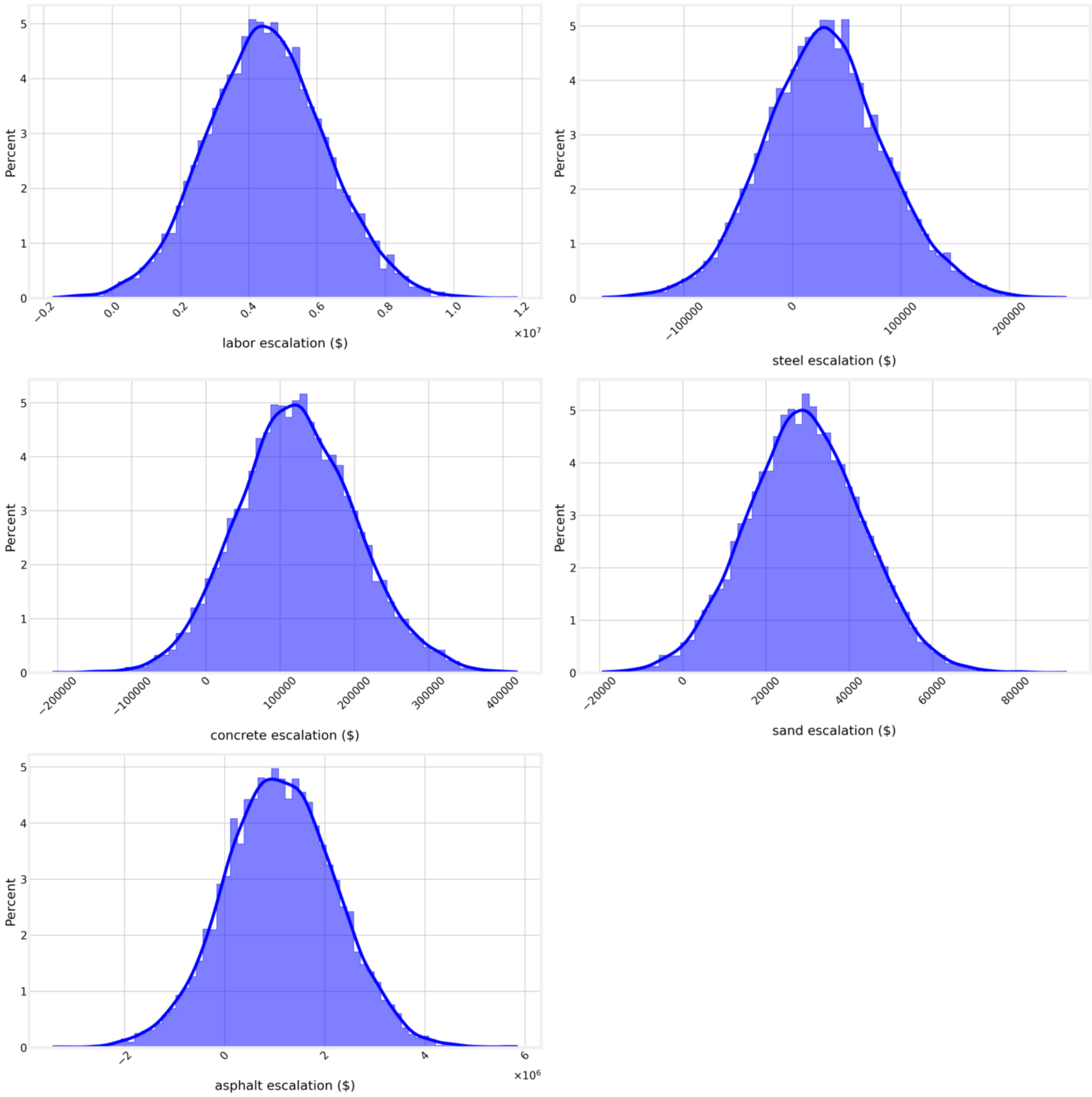


Figure 6.4: Escalation histogram of the project resources

Chapter 7

Discussion

7.1 Modeling resource price changes

This section of the discussion focuses on the initial 3 research questions, which relate to the forecasting of the future changes of resource prices. **Research Question 1** explores the necessity of identifying specific indices, such as those for labor costs, material prices, and macroeconomic factors, that are relevant in forecasting cost escalation in various construction projects. Moving forward, **Research Question 2** investigates the suitability of various time-series models within the context of this research. It examines models like ARIMA and Holt-Winters, among others, evaluating their effectiveness in forecasting the previously identified indices. Lastly, **Research Question 3** addresses the application of the chosen time-series model to forecast the selected indices, while also accounting for the inherent uncertainties in these forecasts.

In the following subsections, the findings related to these questions, along with the limitations encountered and recommendations for future research are discussed.

7.1.1 Findings

In the literature review, several time-series model classes were identified as suitable due to their reported high accuracy. A balance was sought between accuracy and explain-ability, with a preference for models that offered greater explain-ability in this context. This led to the selection of three potential models: Holt-Winters, vector autoregression (VAR), and vector error correction model (VECM). The gathered data encompassed multiple explanatory variables linked to the global economy and the construction industry. For material representation, Produces Price Index (PPI) data for various materials was compiled, along with the construction wage index. This combination of data and models enables the prediction of the change in future resource prices.

After selecting the indices and the three potential time-series models, forecasts need to be created with the selected models. The method of forecasting need to allow for forecasting various indices over multiple time-frames while also accounting for uncertainty of the forecasts. For this reason the aim was set to develop an Automated Forecasting Process (AFP), that allows for the whole sequence from data processing until creating a distributional forecast based on historic data.

In the development of distributional forecasts, a unique approach was adopted that merges Moving Block Bootstrap (MBB) with Seasonal and Trend decomposition using Loess (STL) in order to bootstrap new time-series that can later be forecasted. This method was specifically novel due to the implementation on multi-variate time-series. However, the resulting intervals were found to be too narrow. This outcome was anticipated, as the approach does not account for all sources of uncertainty. The implementation and underlying literature reviewed on the creation of these distributional forecast could be relevant for further research.

For each of the three time-series models, a suitable AFP was established. This process begins by evaluating the appropriateness of the specific model class, followed by the generation of potential candidate models. Subsequently, it selects the 'best' model from these candidates. Once the optimal model is chosen, a distributional forecast using the selected model is created. This resulted in three AFP's for each specific model. Since the VAR and VECM model where not always a suitable fit, these AFP's Needed to be combined with the Holt-Winters AFP as a fall-back mechanism to still allow for a forecast. The different combination of model specific AFP's with fallback models resulted in 4 possible options that combine the three models and associated AFP. The most elaborate version for the final AFP is VECM model that falls back on the VAR if possible and else on the Holt-Winters model. The simplest version is to always utilize the Holt-Winters model. And finally the combination of a VECM that falls back on Holt-Winters and a VAR model that falls back on holt-Winters model where made.

To find the most suitable set-up for the final AFP two steps where taken. First a grid search was employed to determine some key configurations within the AFP for the Holt-Winters, VAR and VECM. After the key configurations where determined the 4 options that combine the AFP's where compared. This result was that the AFP that only uses the Holt-Winters model performed best. This is the final AFP that will be integrated into the tool. The method utilized for configuring and finding the optimal set-up for an AFP could be valuable for similar applications.

Moving on to the results of the different model specific AFP's, the first finding, is the consistent difficulty in accurately forecasting the more volatile indices, such as copper and steel, using the VAR, VECM, and Holt-Winters models. This difficulty in predicting volatile indices is a well-documented challenge in the field of time-series forecasting (Cao & Ashuri, 2020). Furthermore, forecasts extending over a longer term, particularly those spanning four years, were found to be significantly less accurate, especially for these volatile indices. This resulted in substantial forecasting errors and unreliable prediction intervals. The inherent challenge of generating accurate long-term forecasts is a recognized limitation in many models, as noted in various studies (Wang & Ashuri, 2017; Ashuri et al., 2010; Cao & Ashuri, 2020; Faghih & Kashani, 2018).

While this difficulty is in line with most research, the study of Faghih and Kashani (2018) reported a MAPE of 0.57% when forecasting the steel price three years into the future. This is significantly more accurate than the results obtained from the VECM model in this study. The observed difference in model performance between that research and this study can be explained to several factors. The most evident reason is that the research of Faghih and Kashani (2018) only did a single train-test split, opposed to the 9 train-test splits done in this research. When a single train-test split is done, it is easy for data leakage to occur, which can finally result in over optimizing the model for that specific train test split. This method can result in poor out-of-sample forecasts (Cao & Ashuri, 2020). Another reason could be the fact that the VECM forecast created in this study's context were created through an AFP rather than through manual selection and finetuning. It is possible that the VECM model does not achieve its potential in accuracy and manual adjustments, or that a more elaborate automated process is needed to achieve this.

7.1.2 Limitations & recommendations

This chapter will focus on the limitations and recommendations arising from the aspects related to forecasting material price changes. As the aim of the tool is a proof of concept, the creation of the model includes limitations related to the scope. Specifically, the fact that extensive fine-tuning and prediction abilities during 'non-normal' times were outside of the scope leaving significant room for improvement. This, combined with the automatic modeling methodology applied showed that results are indicative but not conclusive. For example, the automatic modeling methodology might favor the Holt-Winters model, due to lack of sophistication at the VECM and VAR model.

In order to improve the accuracy of the models, it could be decided to apply different models within the framework supplied. For example, non-linear models are likely to be more accurate. Alternatively, improvements to the current models data processing could be made or additional configurations could be optimized through the grid search method.

Related to the data processing a robust form of anomaly detection and handling could be applied. A specific method that could be investigated is to use the in this research applied STL decomposition model provided for the purpose of detecting residuals that are far away of the mean. For example the residuals 2 or 3 standard deviations away could be trimmed down or replaced by the mean. As taking out residuals from the data will impact the final model, this needs be done with care. An adaquete method for handling anomalies can reduce the noise in the data, and with this make it more likely to create the true underlying model.

Another method related to the model input data, is to implement exogenous dummy variables in the model. Dummy variables are binary variables that have the value 1 in specific periods, or times. It is possible to apply a seasonal dummy variable for each month as done by wei Xu and Moon (2013). Or another application is to create dummy variables for 2008 financial crash or covid-19, similar to (Moon & Shin, 2018) who identified and implemented dummy variables where structural breaks where detected. This approach might also allow for modeling the escalation during non-normal times. This could possibly result in the creation of a more robust forecasting process.

Another limitation is related to the creation of the distributional forecast. Since the interval coverage where not in line with the aimed confidence intervals. While this was expected due to omitting sources of uncertainty, optimizing this process is needed before reliable utilizing the distributional forecasts. One potential method for improvement could be to scale the standard deviation identified in the distributional forecasts, or to include model uncertainty by considering variations in the potential models. Although exploring model variations was not feasible in this research due to computational constraints, investigating this aspect could be valuable for enhancing the quality of the distributional forecasts.

A second limitation in creating distributional forecasts arises from the practice of calculating the standard deviation of the forecasted values post-bootstrap. This standard deviation may not accurately represent the true distribution of the bootstraps, as the errors in the bootstraps could be biased in a particular direction. To address this, an alternative approach could involve directly storing the bootstrapped forecasts and subsequently sampling directly from these stored forecasts. This method might offer a more accurate reflection of the bootstrap distributions.

Finally the selection of the final model could be improved. This could be done by adding a selection step that looks at the out of sample forecast of the different candidate models and filters them based on this. Additionally different information criterion could be tested, and assessed on final out of sample accuracy. Within the selection of the final, the residual diagnostics could be improved by adding a check for Autocovariances between the residuals of the multivariate models, or by applying different methods for assessing if the residuals are white noise.

Addressing these limitations and exploring the recommended enhancements could significantly advance the tool's capabilities, making it more accurate and applicable in practice.

7.2 Modeling the base cost

This section of the discussion shifts focus to **Research Question 4**, which centers on modeling construction costs for various activities and resources, particularly under the lens of cost uncertainties. This question aims to explore the variability inherent in construction cost estimation, exploring methodologies for capturing these uncertainties. This question aims at estimating the base cost, which refers to the cost at the current day value.

7.2.1 Findings

The research has led to the creation of a tool designed for modeling construction costs, encompassing various activities and resources while taking into account the uncertainties associated with these costs. The uncertainty addressed is the potential deviations in the base cost for specific activities. To ensure practicality and adaptability for various project types, while offering flexibility in input detail, an Excel-based input method was developed. This method allows users to model different project activities and allocate the cost of resources within each activity as a percentage of the total activity cost.

The PERT distribution was identified as the most appropriate for cost modeling of each activity within the scope of this research. This choice was made after considering other distributions such as Lognormal, Normal, and Triangular. The PERT distribution was preferred due to its advantages, as highlighted in several studies. It is particularly adept at representing skewed data and is compatible with three-point estimates (minimum, most-likely, and optimistic values). This compatibility is relevant as it allows for an efficient translation of user input into a meaningful distribution, making it a practical choice for this research context.

7.2.2 Limitations & recommendations

While the developed method for modeling the base cost offers significant advantages, there are areas identified for further enhancement. One key area is the modeling of parallel activities, which would expand the tool's applicability. Additionally, introducing a variety of distributions for modeling the base cost of activities could enhance accuracy. Currently, the PERT distribution is employed for its general suitability, but this may not be optimal for all activities. Different distributions might more accurately represent certain activities. Another area for refinement is the integration of schedule considerations with cost modeling. Although the current model primarily focuses on cost, which is a crucial factor in escalation, incorporating the impact of schedule delays on final escalation costs could provide a more comprehensive understanding of project dynamics.

7.3 The overall tool

This section delves into **Research Question 5**, the final inquiry of this study, which seeks to integrate the forecasts derived from the time-series model with the corresponding construction costs. The focus here is on the development of an accessible tool that effectively combines these elements to provide clear insights into future cost escalation across various types of construction projects. This question is crucial as it aims to bridge the gap between theoretical forecasting models and practical cost estimation, ensuring that the uncertainties associated with both are comprehensively addressed and represented in an accessible format.

7.3.1 Findings

A key outcome of this research is the development methodology of a tool that integrates distributional forecasts, generated by the Automated Forecasting Process (AFP), with base cost estimates represented by the PERT distribution.This integration is achieved through a Monte Carlo simulation that pulls samples from the PERT distribution and the distributional forecast and combines them according to the associated material cost and the moment until when the user wants the material escalation to be calculated. This input is facilitated through a web-interface that hosts the Monte Carlo simulation, the AFP and all related calculations. Users can input their cost estimates in Excel format, and with a simple button press, the simulation is initiated. The output is presented as histograms displaying the escalation for each material, the total project cost, and the overall escalation, along with some key statistics. Overall, this tool offers a project-specific approach to modeling construction cost escalations while accounting for associated uncertainties.

7.3.2 Limitations & recommendations

The primary limitation of the tool stems from the lack of data on actual escalation cost in projects, which hinders the ability to validate the output of the tool. Future research should focus on tracking the real escalation in specific projects to better allow for comparing the modelled outcomes with the real-life outcomes.

A final limitation is that the tool's input and output formats have not undergone testing with a target group of practitioners. Implementing a testing phase could provide valuable insights into optimizing the tool's configuration, leading to a version that is more finely tailored to the needs and preferences of its intended users. This step is crucial for ensuring that the tool is not only theoretically sound but also practical and user-friendly in real-world applications.

Chapter 8

Conclusion

This research was aimed towards the modeling of cost escalation in construction projects. The central question in the research is: **How can cost escalation for various types of construction projects be predicted, accounting for uncertainties in final construction costs and the forecast?** This central questions was answered by developing a tool that allows for forecasting the future escalation while considering the projects specific resource characteristics and timeline.

The future change in material price is obtained by forecasting a Probability Density Function (PDF) of the expected future change of the projects resources for each month over the project timeline. The variation in material prices are depicted by their Produces Price Index (PPI), while changes in labor costs are indicated by the construction labor hourly rate index. To convert this forecast of the index to a monetary value (e.g. $), it is combined with the three point PERT estimate of the base cost. Where the base cost should reflect the current day value of the resources. The distributions for the change in resource price and base cost are combined through a Monte Carlo simulation that considers the timeline and cost of the individual resources. This results in a probabilistic output of the escalation of the individual resource, total project cost, and cost for the various activities.

The forecasting of the distribution, which represents monthly price changes, was achieved through a process that predicts the indices representing the resources. This process, termed the Automated Forecasting Process (AFP), automates data preparation and selects an appropriate forecasting model to generate a distributional forecast future index values. Three different AFPs were developed, each based on a distinct econometric model: Holt-Winters, the vector autoregression (VAR) model, and the vector error correction model (VECM). Given the specific data prerequisites for the VECM and VAR models, a strategy was devised to integrate these models with the Holt-Winters model. This integration ensures a fallback to the Holt-Winters model when the data requirements for VECM and VAR are not met, thereby guaranteeing consistent forecasting capability for the desired variable.

To evaluate these combinations, a rolling-window train-test split method was employed. The process of identifying the optimal AFP and its specific configurations was conducted through a grid search. Interestingly, the results of this search revealed that the best-performing AFP was the one solely utilizing the Holt-Winters model. The Holt-Winters Automated Forecasting

Procedure (AFP) showed reliable point estimates for less volatile indices, such as sand, labor, and concrete. However, it encountered difficulties in accurately forecasting more volatile indices, notably the prices of steel and copper. This highlights an area where further refinement is necessary. Additionally, the challenge of achieving prediction intervals with the desired coverage remained unresolved, primarily due to the omission of certain sources of uncertainty in the chosen methodology.

To enhance user accessibility and practical application, a web interface was developed for the tool. This interface facilitates the combination of current day value resource costs with projected future price change of the resources through a Monte Carlo simulation. It allows users to upload an Excel spreadsheet containing the relevant data, streamlining the process of cost forecasting and analysis. This development significantly improves the usability of the tool, making it more convenient for users to apply it to their specific forecasting needs.

The primary contribution of this research lies in its methodology for forecasting project-specific cost escalations using readily available data. This methodology is particularly relevant for researchers and practitioners interested in developing tools for automatic distributional forecasting for a range of time-series data. While this study primarily utilized indices from the United States, the tool's design, which incorporates automated processes, should allow for straightforward adaptation to different countries.

The developed proof of concept tool has several areas of possible improvements on both the modeling of the project cost and the accuracy of the distributional forecast of the change in resource prices. While the accuracy of the time-series model has been tested, the actual difference between the tools output and the real-world project escalation remain ambiguous. It is recommended for future research to focus on comparing the tool's predictions with real escalation data from completed projects. This comparison is essential for gaining a deeper understanding of project cost escalations. Furthermore, it is advisable to improve the accuracy of the AFP. This can be achieved by employing the grid-search method to test various configurations and by exploring different data preparation techniques. Progress in these two domains is key for allowing the tool's practical implementation in the industry.

Data Availability Statement

The data and python code used in this study are available from the author upon request.

References

(n.d.).

Akintoye, A., Bowen, P. & Hardcastle, C. (1998, 3). Macro-economic leading indicators of construction contract prices. *Construction Management and Economics*, *16*, 159-175. doi: 10.1080/014461998372466

Anvari, S., Tuna, S., Canci, M. & Turkay, M. (2016, 1). Automated box–jenkins forecasting tool with an application for passenger demand in urban rail systems. *Journal of Advanced Transportation*, *50*, 25-49. doi: 10.1002/atr.1332

Ashuri, B., Asce, M. & Lu, J. (2010). Time series analysis of enr construction cost index. *J. Constr. Eng. Manage.*, *136*, 1227-1237. doi: 10.1061/ASCECO.1943-7862.0000231

Ashuri, B. & Shahandashti, S. M. (2012). *Quantifying the relationship between construction cost index (cci) and macroeconomic factors in the united states.*

Bandara, K., Bergmeir, C. & Smyl, S. (2020, 2). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert Systems with Applications*, *140*, 112896. doi: 10.1016/j.eswa.2019.112896

Bergmeir, C., Hyndman, R. & Benítez, J. (2016, 4). Bagging exponential smoothing methods using stl decomposition and box–cox transformation. *International Journal of Forecasting*, *32*, 303-312. doi: 10.1016/j.ijforecast.2015.07.002

Bracher, J., Ray, E. L., Gneiting, T. & Reich, N. G. (2021, 2). Evaluating epidemic forecasts in an interval format. *PLOS Computational Biology*, *17*, e1008618. doi: 10.1371/journal.pcbi.1008618

Burkov, A. (2020). *Machine learning engineering*. True Positive Inc.

Cao, Y. & Ashuri, B. (2020, 7). Predicting the volatility of highway construction cost index using long short-term memory. *Journal of Management in Engineering*, *36*. doi: 10.1061/(asce)me.1943-5479.0000784

Chatfield, C. (1993, 4). Calculating interval forecasts. *Journal of Business Economic Statistics*, *11*, 121-135. doi: 10.1080/07350015.1993.10509938

Chatfield, C. (2001). Prediction intervals for time-series forecasting. *International Series in Operations Research Management Science*, *30*, 475-494.

Chau, K. W. (1995, 9). Monte carlo simulation of construction costs using subjective data. *Construction Management and Economics*, *13*, 369-383. doi: 10.1080/01446199500000042

Choi, C.-Y., Ryu, K. R. & Shahandashti, M. (2021, 2). Predicting city-level construction cost index using linear forecasting models. *Journal of Construction Engineering and Management*, *147*. doi: 10.1061/(asce)co.1943-7862.0001973

Chou, J.-S. (2011, 8). Cost simulation in an item-based project involving construction engineering and management. *International Journal of Project Management*, *29*, 706-717. doi: 10.1016/j.ijproman.2010.07.010

Clements, M. & Hendry, D. (1998). *Forecasting economic time series*. Cambridge University Press. doi: 10.1017/CBO9780511599286

Cleveland, R. B., Cleveland, W. S., McRae, J. E. & Terpenning, I. J. (1990). Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, *6*, 3-33.

Cook, T. R. & Doh, T. (2019). Assessing macroeconomic tail risks in a data-rich environment. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3494412

Dickey, D. A. & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366a), 427-431. Retrieved from https://doi.org/10.1080/01621459.1979.10482531 doi: 10.1080/01621459.1979.10482531

Dong, J., Chen, Y. & Guan, G. (2020). Cost index predictions for construction engineering based on lstm neural networks. *Advances in Civil Engineering*, *2020*. doi: 10.1155/2020/6518147

Duval, A. (2019, 4). Explainable artificial intelligence (xai).

Elfahham, Y. (2019, 6). Estimation and prediction of construction cost index using neural networks, time series, and regression. *Alexandria Engineering Journal*, *58*, 499-506. doi: 10.1016/j.aej.2019.05.002

Faghih, S. A. M., Gholipour, Y. & Kashani, H. (2021, 8). Time series analysis framework for forecasting the construction labor costs. *KSCE Journal of Civil Engineering*, *25*, 2809-2823. doi: 10.1007/s12205-021-1489-4

Faghih, S. A. M. & Kashani, H. (2018, 8). Forecasting construction material prices using vector error correction model. *Journal of Construction Engineering and Management*, *144*. doi: 10.1061/(asce)co.1943-7862.0001528

Fanchon, P. & Wendel, J. (1992, 2). Estimating var models under non-stationarity and cointegration: alternative approaches for forecasting cattle prices. *Applied Economics*, *24*, 207-217. doi: 10.1080/00036849200000119

Flyvbjerg, B. H. (2003). How common and how large are cost overruns in transport infrastructure projects? *Transport Rev*, *23*, 71-.

Gardner, E. S. (1985, 1). Exponential smoothing: The state of the art. *Journal of Forecasting*, *4*, 1-28. doi: 10.1002/for.3980040103

Ghysels, E. & Marcellino, M. (2018). *Applied economic forecasting using time series methods*. Oxford University Press.

Giraitis, L., Kapetanios, G. & Yates, T. (2018, 3). Inference on multivariate heteroscedastic time varying random coefficient models. *Journal of Time Series Analysis*, *39*, 129-149. doi: 10.1111/jtsa.12271

Gneiting, T. & Raftery, A. E. (2007, 3). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*, 359-378. doi: 10.1198/016214506000001437

Granger, C. W. J. (1969, 8). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*, 424. doi: 10.2307/1912791

Greene, W. H. (2012). *Econometric analysis* (7th ed.). Pearson Education Limited.

Grogan, T. (1992). Cost history-keeping track of a moving target. *ENR*, *228*, 42-47.

Hall, P. & Horowitz, J. L. (1993). *Corrections and blocking rules for the block bootstrap with dependent data.* Retrieved from https://EconPapers.repec.org/RePEc:uia:iowaec:93-11

Han, M., Zhang, R., Qiu, T., Xu, M. & Ren, W. (2019, 10). Multivariate chaotic time series prediction based on improved grey relational analysis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *49*, 2144-2154. doi: 10.1109/TSMC.2017.2758579

Holt, C. C. (2004, 1). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, *20*, 5-10. doi: 10.1016/j.ijforecast.2003.09.015

Hwang, S. (2011, 9). Time series models for forecasting construction costs using time series indexes. *Journal of Construction Engineering and Management*, *137*, 656-662. doi: 10.1061/(asce)co.1943-7862.0000350

Hwang, S., Park, M., Lee, H.-S. & Kim, H. (2012, 11). Automated time-series cost forecasting system for construction materials. *Journal of Construction Engineering and Management*, *138*, 1259-1269. doi: 10.1061/(ASCE)CO.1943-7862.0000536

Hyndman, R. J. & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd edition ed.). OTexts.

Hyndman, R. J., Koehler, A. B., Snyder, R. D. & Grose, S. (2002, 7). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, *18*, 439-454. doi: 10.1016/S0169-2070(01)00110-8

Inc., P. T. (2024). *Collaborative data science.* Author Retrieved from https://plot.ly

Jiang, H., Xu, Y. & Liu, C. (2013, 11). Construction price prediction using vector error correction models. *Journal of Construction Engineering and Management*, *139*. doi: 10.1061/(ASCE)CO.1943-7862.0000729

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, *12*(2), 231-254. Retrieved from https://www.sciencedirect.com/science/article/pii/0165188988900413 doi: https://doi.org/10.1016/0165-1889(88)90041-3

Kunsch, H. R., Ktnsch, H. R. & Zurich, E. (1989). *The jackknife and the bootstrap for general stationary observations* (Vol. 17).

Li, G. S. H. & Maddala. (1996, 1). Bootstrapping time series models. *Econometric Reviews*, *15*, 115-158. doi: 10.1080/07474939608800344

Li, J. (2021, 4). Block bootstrap prediction intervals for parsimonious first-order vector autoregression. *Journal of Forecasting*, *40*, 512-527. doi: 10.1002/for.2728

Liu, W.-H. (2007, 7). Forecasting the semiconductor industry cycles by bootstrap prediction intervals. *Applied Economics*, *39*, 1731-1742. doi: 10.1080/00036840600706995

Ljung, G. M. & Box, G. E. P. (1978, 8). On a measure of lack of fit in time series models. *Biometrika*, *65*, 297-303. doi: 10.1093/biomet/65.2.297

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer.

Makridakis, S. & Winkler, R. L. (1989). Sampling distributions of post-sample forecasting errors. *Applied Statistics*, *38*, 331. doi: 10.2307/2348063

McCormick, G. P. (1969, 1). Communications to the editor—exponential forecasting: Some new variations. *Management Science*, *15*, 311-320. doi: 10.1287/mnsc.15.5.311

Mir, M., Kabir, H. M., Nasirzadeh, F. & Khosravi, A. (2021, 7). Neural network-based interval forecasting of construction material prices. *Journal of Building Engineering*, *39*. doi: 10.1016/j.jobe.2021.102288

Moon, T. & Shin, D. H. (2018, 5). Forecasting construction cost index using interrupted time-series. *KSCE Journal of Civil Engineering*, *22*, 1626-1633. doi: 10.1007/s12205-017-0452-x

of Labor Statistics, U. B. (2023, 12). *Databases, tables calculators by subject.*

P., G. E. & Jenkins, G. M. B. (1984). *Time series analysis: Forecasting and control* (2nd Ed ed.). Holden-Day.

Ritvik, K. (2020). *ice cream interest[data set].*

Runeson, K. (1988, 3). Methodology and method for price-level forecasting in the building industry. *Construction Management and Economics*, *6*, 49-55. doi: 10.1080/01446198800000006

Scott, A. (2001). *Principles of forecasting* (Vol. 30; S. Armstrong, Ed.). Springer US. doi: 10.1007/978-0-306-47630-3

Seabold, S. & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python..

Sekma, N. C., Elleuch, A. & Dridi, N. (2016, 9). Automated forecasting approach minimizing prediction errors of cpu availability in distributed computing systems. *International Journal of Intelligent Systems and Applications*, *8*, 8-21. doi: 10.5815/ijisa.2016.09.02

Skitmore, R. (1987). *Construction market: The market effect.* The University of Salford.

Sonmez, R. (2004). Selection of a probability distribution function for construction cost estimation. *Association of Researchers in Construction Management*, *1*, 125-128.

Streamlit. (2024, 1).

Tang, B.-Q., Han, J., feng Guo, G., Chen, Y. & Zhang, S. (2019, 5). Building material prices forecasting based on least square support vector machine and improved particle swarm optimization. *Architectural Engineering and Design Management*, *15*, 196-212. doi: 10.1080/17452007.2018.1556577

Taylor, J. W. (2003, 10). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*, *19*, 715-725. doi: 10.1016/S0169-2070(03)00003-7

Taylor, R. G. & Bowen, P. A. (1987, 3). Building price-level forecasting: an examination of techniques and applications. *Construction Management and Economics*, *5*, 21-44. doi: 10.1080/01446198700000004

Touran, A., Asce, M. & Lopez, R. (2006). Modeling cost escalation in large infrastructure projects.
doi: 10.1061/ASCE0733-93642006132:8853

Vigen, T. (2023). *Spurious correlations.* Retrieved from https://www.tylervigen.com/spurious-correlations

Wall, D. M. (1997, 5). Distributions and correlations in monte carlo simulation. *Construction Management and Economics*, *15*, 241-258. doi: 10.1080/014461997372980

Wang, J. & Ashuri, B. (2017, 1). Predicting enr construction cost index using machine-learning algorithms. *International Journal of Construction Education and Research*, *13*, 47-63. doi: 10.1080/15578771.2016.1235063

wei Xu, J. & Moon, S. (2013, 1). Stochastic forecast of construction cost index using a cointegrated vector autoregression model. *Journal of Management in Engineering*, *29*, 10-18. doi: 10.1061/(ASCE)ME.1943-5479.0000112

Wilmot, C. G., Asce, M. & Mei, B. (2005). Neural network modeling of highway construction costs. *J. Constr. Eng. Manage.*, *131*, 765-771. doi: 10.1061/ASCE0733-93642005131:7765

Winters, P. R. (1960, 4). Forecasting sales by exponentially weighted moving averages. *Management Science*, *6*, 324-342. doi: 10.1287/mnsc.6.3.324

Wooldridge, J. M. (2009). *Introductory econometrics: A modern approach.* South-Western Cengage Learning.

Appendix A

Selected Explanatory Variables

Table A.1: Summary of Economic Indicators

Variable	**Explanation of Variable**
Consumer Price Index (CPI)	Measures average change over time in prices paid by consumers for a basket of goods and services.
Producer Price Index (PPI)	Measures average change over time in selling prices received by domestic producers for their output.
Gross Domestic Product (GDP)	Total value of goods and services produced over a specific time period within a country.
Money Supply (M2)	Includes cash, checking deposits, and near money. It's a measure of the money circulating in the economy.
Employment Rate	Proportion of the working-age population that is employed.
Total Construction Spending	Total dollar value of all construction work done in a period.
Labour Productivity	Measures output produced per unit of labor.
Interest Rate (Long-Term)	Interest rate on long-term debt instruments like bonds.
Housing Starts	Measure of the number of new residential construction projects that have begun over a specific period.
Iron Ore Price	Price for iron ore, a key raw material in steelmaking.
Building Permits	Official approvals issued for the construction or substantial alteration of a building.

Appendix B

Detailed results of the final Automated Forecasting Process configuration

Table B.1: Interval Score for Various Variables Over Different Forecasting Horizons

Variable	12-Month Horizon	24-Month Horizon	36-Month Horizon	48-Month Horizon
PPI: Concrete	28.9	66.3	117.8	176.9
PPI: Copper	461.0	770.7	955.6	1015.4
PPI: Steel	191.9	406.0	557.3	591.6
PPI: asphalt	204.5	263.9	299.8	334.1
PPI: sand	13.1	36.8	72.1	112.5
construction wage	22.6	40.5	68.2	100.2

Table B.2: Selected AFP: Mean Absolute Percentage Error (MAPE) for explained variables Over different forecasting horizons

Variable	12-Month Horizon	24-Month Horizon	36-Month Horizon	48-Month Horizon
PPI: Concrete	1.2	2.2	3.5	4.8
PPI: Copper	13.8	20.0	23.3	25.7
PPI: Steel	8.8	18.2	16.0	16.9
PPI: asphalt	3.9	5.0	6.1	7.0
PPI: sand	0.5	1.1	1.8	2.4
construction wage	0.7	1.1	1.7	2.3

Table B.3: Bias for Various Variables Over Different Forecasting Horizons

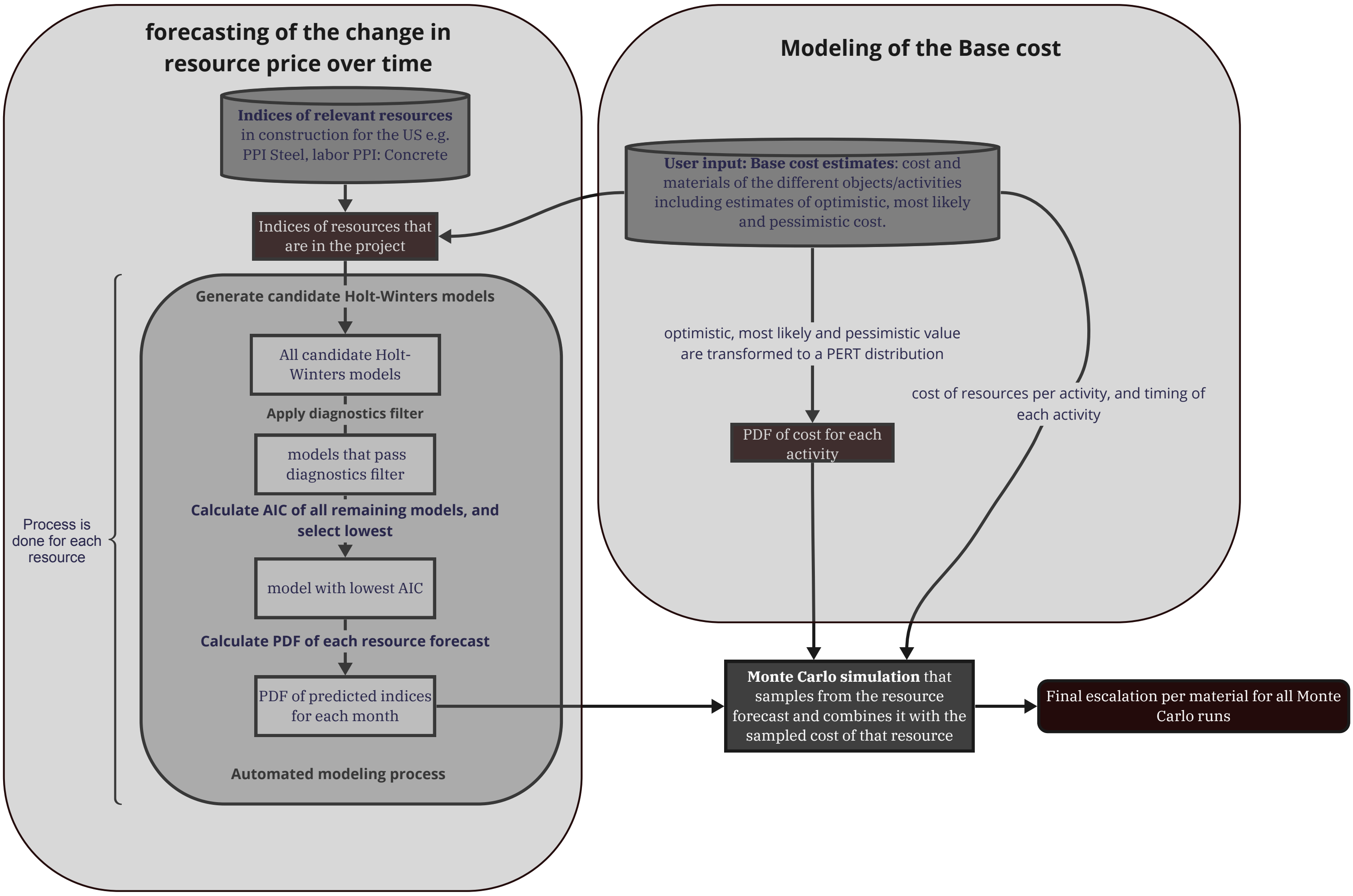
Variable	**12-Month Horizon**	**24-Month Horizon**	**36-Month Horizon**	**48-Month Horizon**
PPI: Concrete	0.6	0.6	0.3	-0.5
PPI: Copper	8.7	12.4	12.6	14.1
PPI: Steel	0.4	4.9	-6.5	-7.2
PPI: asphalt	-2.8	-4.7	-6.8	-9.0
PPI: sand	0.3	0.7	1.0	1.1
construction wage	0.1	0.3	0.4	0.4

Table B.4: Interval Coverage for Various Variables Over Different Forecasting Horizons

Variable	**12-Month Horizon**	**24-Month Horizon**	**36-Month Horizon**	**48-Month Horizon**
PPI: Concrete	70.1	67.1	55.0	46.9
PPI: Copper	58.1	44.4	41.4	40.8
PPI: Steel	73.5	60.4	53.0	53.6
PPI: asphalt	77.8	72.4	68.5	66.0
PPI: sand	74.4	67.6	61.9	59.6
construction wage	45.3	40.9	38.4	33.1

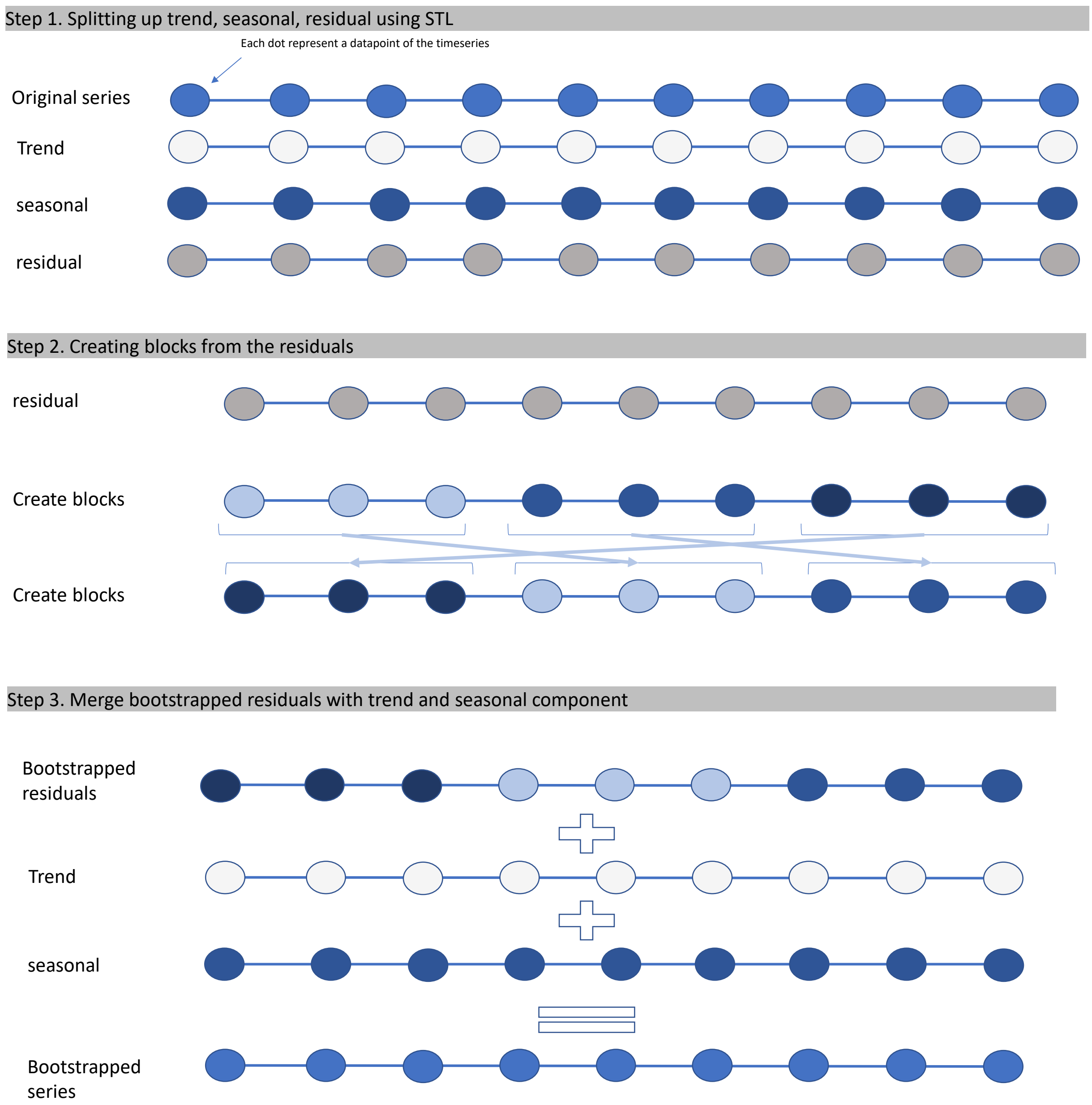
Appendix C

Final architecture of the model (full version)


forecasting of the change in resource price over time
Indices of relevant resources in construction for the US e.g. PPI Steel, labor PPI: Concrete
Indices of resources that are in the project
Generate candidate Holt-Winters models
All candidate Holt-Winters models
Apply diagnostics filter
models that pass diagnostics filter
Calculate AIC of all remaining models, and select lowest
model with lowest AIC
Calculate PDF of each resource forecast
PDF of predicted indices for each month
Automated modeling process
Process is done for each resource
Modeling of the Base cost
User input: Base cost estimates: cost and materials of the different objects/activities including estimates of optimistic, most likely and pessimistic cost.
optimistic, most likely and pessimistic value are transformed to a PERT distribution
PDF of cost for each activity
cost of resources per activity, and timing of each activity
Monte Carlo simulation that samples from the resource forecast and combines it with the sampled cost of that resource
Final escalation per material for all Monte Carlo runs

Appendix D

Moving Block Bootstrap visualization


Step 1. Splitting up trend, seasonal, residual using STL
Each dot represent a datapoint of the timeseries
Original series
Trend
seasonal
residual
Step 2. Creating blocks from the residuals
residual
Create blocks
Create blocks
Step 3. Merge bootstrapped residuals with trend and seasonal component
Bootstrapped residuals
Trend
seasonal
Bootstrapped series