Applying Bayesian Forecasting to Predict New Customers' Heating Oil Demand

Tsuginosuke Sakauchi
Marquette University

APPLYING BAYESIAN FORECASTING TO PREDICT NEW CUSTOMERS' HEATING OIL DEMAND

by

Tsuginosuke Sakauchi, B.S.

A Thesis Submitted to the Faculty of the
Graduate School, Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Master of Science

Milwaukee, Wisconsin

August 2011

ABSTRACT

APPLYING BAYESIAN FORECASTING TO PREDICT NEW CUSTOMERS' HEATING OIL DEMAND

Tsuginosuke Sakauchi, B.S.

Marquette University, 2011

This thesis presents a new forecasting technique that estimates energy demand by applying a Bayesian approach to forecasting. We introduce our Bayesian Heating Oil Forecaster (BHOF), which forecasts daily heating oil demand for individual customers who are enrolled in an automatic delivery service provided by a heating oil sales and distribution company. The existing forecasting method is based on linear regression, and its performance diminishes for new customers who lack historical delivery data. Bayesian methods, on the other hand, respond effectively in the start-up situation where no prior data history is available.

Our Bayesian Heating Oil Forecaster uses forecasters' past performances for existing customers to adjust the current forecast for target customers. We adapted a Bayesian approach to forecasting combined with domain knowledge and original ideas to develop our Bayesian Heating Oil Forecaster, which forecasts demand for target customers without relying on their historical deliveries.

Performance evaluation demonstrates that our Bayesian Heating Oil Forecaster shows increased performance over the existing forecasting method when the two techniques are combined. We used Root Mean Squared Error (RMSE) and Mean Absolute Percent Error (MAPE) to compare the performance of the two algorithms. Compared to the existing forecasting method alone, our Simple Average model, which combines the forecasts from the existing forecasting method and our Bayesian Heating Oil Forecaster, recorded an overall improvement of 2.4% in RMSE, 5.0% in MAPE Actual, and 2.8% in MAPE Capacity for company A and 0.3%, 7.1%, and 2.8% for company B.

ACKNOWLEDGMENTS

Tsuginosuke Sakauchi, B.S.

The completion of this thesis would have been impossible without the extensive encouragement and guidance I received from my mentors, colleagues, and my family. I would like to specifically express my sincere gratitude to Dr. Brown and Dr. Corliss. Dr. Brown offered profound technical insights from the energy forecasting domain in addition to financial support from his GasDay lab. Dr. Corliss patiently provided me with countless hours of advice, insights, guidance, and encouragement throughout the course of this work. Their extensive support and mentoring made this work possible.

I would like to express my appreciation to Dr. Adya and Dr. Povinelli for volunteering to offer their time, effort, and knowledge as my thesis committee members. Their valuable insights helped improve the quality of this work. I would also like to thank my colleagues, Anisha D'Silva, Samson Kiware, Yifan Li, Bo Pang, and Steve Vitullo for sharing their ideas and offering help and encouragement during my graduate studies. It was a pleasure working with them as graduate students in attaining common goals.

I dedicate this work to my parents, siblings, and mentors, who always supported and inspired me to achieve my academic goals.

TABLE OF CONTENTS

ACKNOWLEDGMENTS . . . . **i**
LIST OF TABLES . . . . **iv**
LIST OF FIGURES . . . . **v**
CHAPTER 1 Thesis Introduction **1**
1.1 Background of this Research Project 1
1.2 Problem Background 2
1.3 Types of Heating Oil Customers 5
1.4 Current Process 6
1.4.1 Linear Regression Model 6
1.4.2 Expert's Estimated $\mathcal{K}$-factor 10
1.4.3 Tank Size and $\mathcal{K}$-factor Relationship 10
1.5 Problem with the Current Process 11
1.6 Problem Statement 12
1.7 Assumptions 13
1.7.1 Availability of Historical Data 13
1.7.2 Demand Forecast and Actual Use 14
1.7.3 Stationarity of the $\mathcal{K}$-factor 15
1.7.4 Positive and Negative Errors 15
1.8 Evaluation 16
1.9 Organization of this Thesis 19
CHAPTER 2 Survey of Energy Forecasting Literature **20**
2.1 Existing Demand Forecasting Methods 20
2.1.1 Multiple Linear Regression 21
2.1.2 Artificial Neural Network 22
2.1.3 Ensemble and Combined Forecasts 23
2.2 Bayes' Theorem, Bayesian Probability, and Bayesian Inference 24
2.2.1 Discrete Bayesian Analysis 28
2.2.2 Continuous Bayes Inference 41
2.3 Existing Bayesian Forecasting Methods 45
2.3.1 Bayesian Networks 45
2.3.2 Bayesian Pooling / Empirical Bayes 55
2.3.3 Dynamic Linear Models 61
CHAPTER 3 Bayesian Heating Oil Forecaster **68**
3.1 Thought Experiment: Forecasting Demand Without Historical Data . 68
3.2 Overview of the Bayesian Heating Oil Forecaster 73
3.3 Computation Steps of our Bayesian Heating Oil Forecaster 76
3.3.1 Step 1: Compute initial belief 77
3.3.2 Step 2: Obtain $\mathcal{K}$-factor estimate from the expert 79

3.3.3 Step 3: Update belief using the expert likelihood 80
3.3.4 Step 4: Obtain posterior belief 85
3.3.5 Step 5: Obtain $\mathcal{K}$-factor estimate from the model 87
3.3.6 Step 6: Update belief using the model likelihood 87
3.3.7 Step 7: Obtain posterior belief 90
3.3.8 Step 8: Repeat steps 5 through 7 91
3.3.9 Step 9: Obtain $\mathcal{K}$-factor estimate 91
3.3.10 Step 10: Obtain estimated heating oil demand 92
3.4 Software Implementation 93

CHAPTER 4 Bayesian Heating Oil Forecaster Test Results 99

4.1 Evaluation Method 99
4.1.1 Backtesting Process 100
4.1.2 Evaluation Criteria 103
4.2 Models 105
4.3 Data Sets Used During the Test 106
4.4 Trimming 108
4.5 Chi-Square Goodness-of-Fit Test and Beta Probability Distributions . 109
4.6 Results 113

CHAPTER 5 Conclusions and Future Research 121

5.1 Conclusions 121
5.2 Recommendations 123
5.3 Future Research 124

BIBLIOGRAPHY 126

LIST OF TABLES

2.1 Joint probability distribution for the balls-in-an-urn example 32
2.2 Joint probability distribution for the balls-in-an-urn example with joint probabilities calculated 33
2.3 Posterior probability distribution after the first observation 34
2.4 Posterior probability distribution after the second observation 35
2.5 An example of a possible Prior(1) for the basketball example 37
2.6 Likelihood for the basketball example 39
2.7 Update for the basketball example 40

LIST OF FIGURES

1.1 Heating oil delivery company 2
1.2 Heating oil delivery seasons 3
1.3 Visual representation of the existing model 6
1.4 Heating Oil Consumption vs. Heating Degree Days 8
1.5 Visual representation of the existing model in its transient state 11
1.6 Visual representation of the existing model in its steady state 11
1.7 Steps of the evaluation method 17
2.1 How likely is a cause given the effect? 25
2.2 A model of how Bayes' Theorem updates forecaster knowledge 26
2.3 The balls-in-an-urn example 29
2.4 Relative strength of basketball teams example 36
2.5 Example Bayesian network modeling causes of wet grass 46
2.6 Example joint probability distribution 47
2.7 An example Bayesian Network that describes flight delays 50
2.8 An example Bayesian Network with probabilities 52
2.9 Overview of Bayesian pooling 56
2.10 Illustration of nonstationary time series with four pattern regimes 58
3.1 Experiment setup 69
3.2 Histogram of the latest $\mathcal{K}$-factor estimates for the existing customers . . 70
3.3 A simplified likelihood table 71
3.4 A model of how Bayes' Theorem updates forecaster knowledge 73
3.5 Observing the expert $\mathcal{K}$-factor estimate 74
3.6 Observing the model $\mathcal{K}$-factor estimate 74
3.7 Event timeline and algorithm behavior 75
3.8 Frequency distribution of the latest $\mathcal{K}$-factor estimates for the existing customers 78
3.9 Histogram of the latest $\mathcal{K}$-factor estimates for the existing customers . . 78
3.10 Empirical PDF (blue) and Beta PDF (red) 79
3.11 Portion of the Expert Likelihood Joint Frequency Distribution between November 14, 2007, and September 30, 2009 (175 observations) 81
3.12 Portion of the Expert Likelihood Joint Probability Distribution between November 14, 2007, and September 30, 2009 (175 observations) 81
3.13 Column 11 marginal distribution from the expert likelihood joint distribution 83
3.14 Histogram of the Marginal Frequency Distribution for the Expert Likelihood 84
3.15 Empirical PDF (blue) and Beta PDF (red) for the Expert Likelihood . . 84
3.16 Portion of the Model Likelihood Joint Frequency Distribution between November 14, 2007, and September 30, 2009 (24,121 observations) 88

3.17 Portion of the Model Likelihood Joint Probability Distribution between November 14, 2007, and September 30, 2009 (24,121 observations) 89
3.18 Histogram of the Marginal Frequency Distribution for the Model Likelihood 90
3.19 Empirical PDF (blue) and Beta PDF (red) for the Model Likelihood . . 90
3.20 Theoretical implementation of our Bayesian Heating Oil Forecaster . . . 94
3.21 Actual implementation of our Bayesian Heating Oil Forecaster 95
4.1 Ex-ante Forecast Training Set 100
4.2 Ex-post Forecast Training Set 100
4.3 Ex-post implementation of our Bayesian Heating Oil Forecaster 101
4.4 Number of Deliveries by Delivery Number 108
4.5 Comparison of the Fitness of Various Distributions for Company A . . . 111
4.6 Comparison of the Fitness of Various Distributions for Company B . . . 112
4.7 P-values of the Goodness-of-Fit Chi-Squared Test 113
4.8 RMSE 115
4.9 MAPE Actual 116
4.10 MAPE Capacity 117
4.11 RMSE and MAPE Percent Change 118
4.12 RMSE and MAPE Percent Change (Trim 10% Customer) 119
4.13 RMSE and MAPE Percent Change (Trim 10% Delivery) 119

CHAPTER 1

Thesis Introduction

This chapter introduces the context of this thesis and the problem being addressed. First, the background knowledge required to understand the scope of the problem is presented. Next, the project objectives and the evaluation criteria are discussed. Finally, the organization of the remainder of the thesis is outlined.

1.1 Background of this Research Project

This thesis is written as a part of Marquette University College of Engineering GasDay project. The GasDay project collaborates with natural gas distributers, called the Local Distribution Companies, to produce mathematical models that forecast natural gas demand. In addition, the project applies its prediction techniques to provide other services, including automatic detection of suspect natural gas meter readings and heating oil demand forecasting. Currently, the GasDay project relies heavily on techniques such as Multiple Linear Regression, Artificial Neural Network, and Dynamic Model Adaptation [48]. One of the disadvantages shared by these techniques is that the forecasting accuracy diminishes when there is a lack of historical data. The motivation of this thesis is to address this disadvantage by applying a forecasting technique that improves the prediction of heating oil demand even when there is a lack of historical data.

1.2 Problem Background

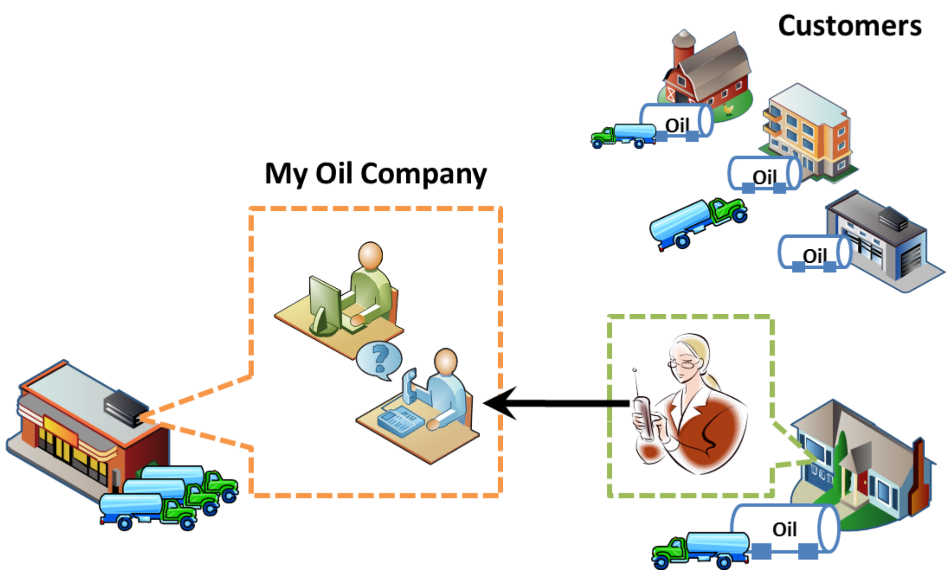

Figure 1.1: Heating oil delivery company

Figure 1.1 depicts a typical heating oil sales and distribution company. The company provides heating oil to residential and commercial customers using a fleet of delivery trucks. To reduce the number of unnecessary deliveries, the company estimates each of its customer's heating oil demand between deliveries. An estimate that is too large increases the operational cost because the company is delivering oil more frequently than necessary. An estimate that is too small risks allowing the customer to run out of fuel. This reduces the company's revenue since customers who run out of fuel typically switch suppliers. Therefore, an accurate estimate of each customer's heating oil demand is crucial for a heating oil supplier to minimize the operational cost without reducing the quality of service.

It is very common for a heating oil distribution company to have customers that lack sufficient historical data: hundreds of new customers sign up for the service every year. Customers are said to lack sufficient historical data when the data does not capture the behavior of the customers *throughout the year*. It is necessary to have historical data that extends throughout the year because of the manner in which the heating oil is delivered as described below.

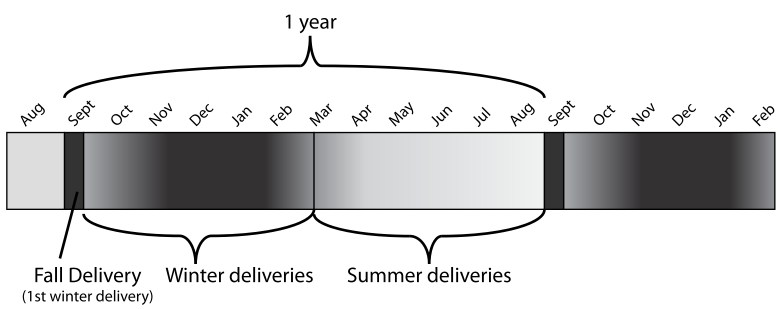

Figure 1.2: Heating oil delivery seasons

Figure 1.2 depicts a typical year with different seasons. Darker color represents high demand for oil, while lighter color represents low demand for oil. As shown in Figure 1.2, most of the heating oil *demand* is concentrated during the winter. In contrast, heating oil *delivery* takes place throughout the year.

Summer deliveries are different from winter deliveries because the heating oil demand tends to be very small. Hence, forecasting accuracy diminishes when the demand during the summer is forecast using the observations from the winter months.

By the beginning of each heating season (typically late September), the company delivers oil to all customers regardless of the estimated demand. This ensures that customers enter the heating season with a full tank of oil. This delivery is unusual since there is typically a gap of several months between this delivery and the previous delivery. Partly due to this large time interval, this delivery behaves differently from others.

To capture all of the customer's behavior throughout the year, at least one full winter, one full summer, and one fall delivery must be observed. Since some customers sign up during the heating season, at least one and a half years (18 months) worth of data must be collected to ensure that at least one fall delivery and one full winter is observed. In other words, new customers have insufficient historical data for high-quality forecasts during the first 18 months. For the purpose of this thesis, we define the period in which a new customer is considered to have insufficient historical data during the following deliveries:

- First 10 deliveries starting from the second delivery, and
- first 18 months counting from the date of the second delivery.

This thesis concerns only the period in which a customer has insufficient historical data, so the term "customer" refers to customers during this initial period, unless otherwise stated.

1.3 Types of Heating Oil Customers

In general, there are two types of residential and commercial heating oil customers.

Space heating customers only use heating oil to heat enclosed areas such as homes and storage facilities. These customers consume more heating oil as the temperature decreases. During the summer time when the temperature is high, these customers do not consume any heating oil.

Space and water heating customers use heating oil to heat rooms and to heat water. These customers behave similar to space heating customers during winter. However, these customers continue to consume a modest amount of heating oil to heat water during summer.

Both types of customers are considered in this thesis.

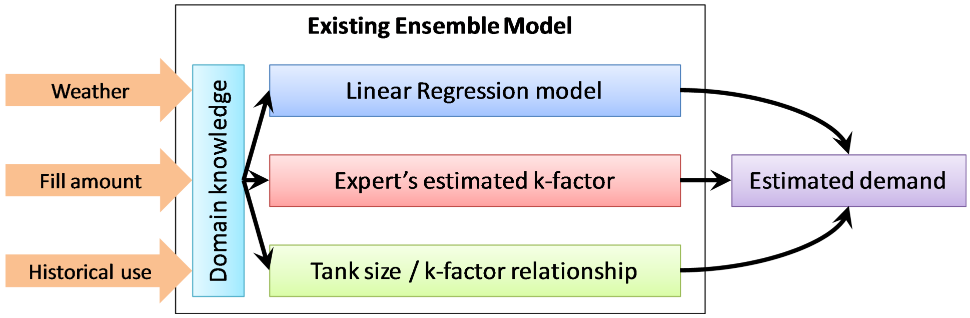

Figure 1.3: Visual representation of the existing model

1.4 Current Process

Currently, the heating oil demand is forecast using an ensemble model. As seen in Figure 1.3, the ensemble model consists of three components: the Linear Regression (LR) model, an expert's estimated $\mathcal{K}$-factor (measure of the response of use to variations in temperature), and the tank size and $\mathcal{K}$-factor relationship. Additionally, the ensemble model contains other enhancements based on domain knowledge. An estimated demand is calculated by combining the output from these components.

1.4.1 Linear Regression Model

The Linear Regression (LR) model takes advantage of the primarily linear relationship between the heating oil demand and heating degree days. This is demonstrated in Figure 1.4, which plots the Cumulative Heating Degree Day against the delivery amount. Since the LR model is a daily model, we divide the

Cumulative Heating Degree Days and the delivery amount by the number of days between deliveries. The plot clearly shows a linear trend where the delivery amount increases as the weather becomes colder and the cumulative heating degree day increases. The LR model is driven by weather inputs (actual and forecast wind and temperature data) with the following variables:

- $\widehat{s_k}$ is the estimated heating oil demand on the k^{th} day in gallons
- $\widehat{\beta_0}$ is the estimated baseload (BL) in gallons
- $\widehat{\beta_1}$ is the estimated heatload coefficient in gallons per Heating Degree Day
- $\widehat{\mathcal{K}}$ is the estimated $\mathcal{K}$-factor in Heating Degree Days per gallon
- $x_{1,k}$ or $\text{HDD}_{60,k}$ is the Heating Degree Day with reference temperature 60 °F on the k^{th} day

The model itself is expressed as

$$\widehat{s_k} = \widehat{\beta_0} + \widehat{\beta_1} x_{1,k} = \widehat{\beta_0} + (1/\widehat{\mathcal{K}})\text{HDD}_{60,k}. \tag{1.1}$$

Baseload describes the portion of the demand that is *not* affected by the daily average temperature. We expect that space heating does not occur when the temperature is high (i.e. during the summer months). Therefore, baseload for

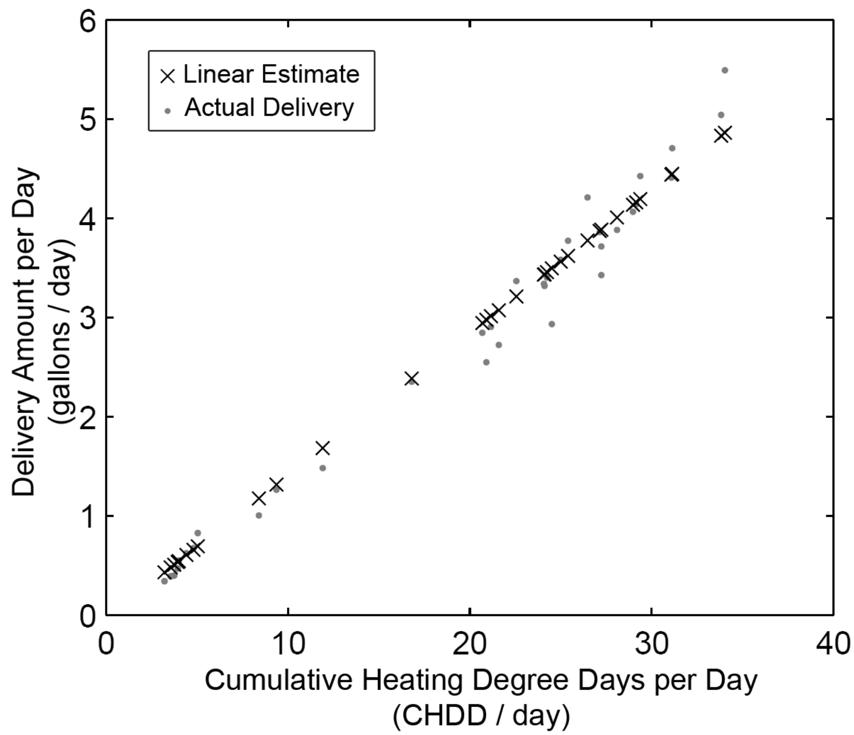

Figure 1.4: Heating Oil Consumption vs. Heating Degree Days

space heating customers theoretically should be zero. However, fitting a regression model often causes the baseload coefficient to be a small non-zero value for space heating customers. The baseload for space and water heating customers is typically a positive value because water heating occurs regardless of the temperature.

Heatload coefficient describes how a customer is "sensitive" to temperature change. Its unit is gallons per Heating Degree Day. A customer with a large heatload coefficient is said to be sensitive to the daily average temperature because a one degree increase in the Heating Degree Day greatly increases the estimated demand.

$\mathcal{K}$**-factor** is the number of Heating Degree Days required to consume one gallon of heating oil. Its unit is Heating Degree Days per gallon. One can think of this factor as a miles-per-gallon equivalent for a heating oil customer. A large MPG represents a fuel efficient car, and a large $\mathcal{K}$-factor represents a fuel efficient customer. $\mathcal{K}$-factor and the heatload coefficient are inversely related to each other.

Heating Degree Day (HDD) is defined as the reference temperature (T_{ref}) minus the average temperature on the k^{th} day (T_k). If the subtraction results in a negative number (i.e. if the average temperature is greater than the reference temperature), then the HDD is set to 0:

$$\text{HDD}_{T_{ref},k} = \max(0, T_{ref} - T_k).$$

The concept of Heating Degree Day was introduced because of the non-linear relationship between the daily average temperature and the heating oil demand. In other words, Heating Degree Day mathematically expresses the fact that customers no longer consume heating oil when the temperature is warmer than, say, 65 °F.

1.4.2 Expert's Estimated $\mathcal{K}$-factor

A domain expert at the heating oil company provides his own $\mathcal{K}$-factor estimate for each delivery. This piece of information is especially valuable during the initial deliveries when domain knowledge compensates for the lack of historical data. The weight of this component is reduced as the number of deliveries (and the amount of historical data available) increase.

1.4.3 Tank Size and $\mathcal{K}$-factor Relationship

This component takes advantage of the domain knowledge that customers have large fuel tanks because they tend to consume more fuel. This suggests that the $\mathcal{K}$-factor is inversely proportional to the tank size. Hence we fit a simple linear regression model to the existing customers' tank sizes and $\mathcal{K}$-factor estimates. The slope and intercept parameter estimates are used to estimate the target customer's $\mathcal{K}$-factor based on the customer's tank size. Similar to the expert's estimated $\mathcal{K}$-factor, the weight of this component is reduced as the number of delivery increases.

1.5 Problem with the Current Process

The operation and performance of the ensemble model changes depending on the amount of available historical data. **Transient** (Figure 1.5) refers to the start-up period when there is limited historical data. **Steady-state** (Figure 1.6) refers to a time period when there is enough historical data to perform reliable forecasting.

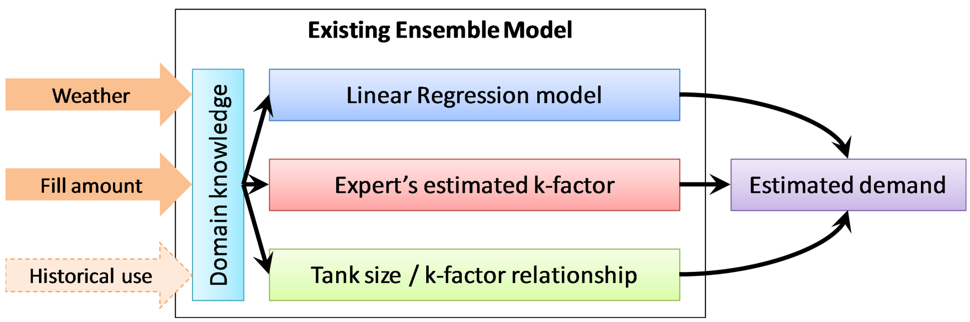

Figure 1.5: Visual representation of the existing model in its transient state

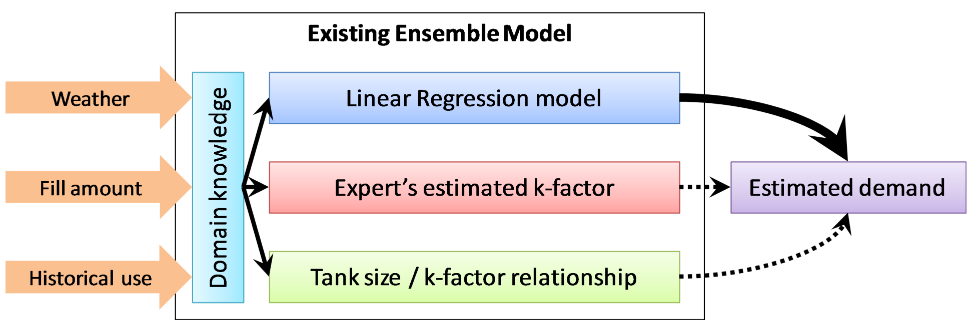

Figure 1.6: Visual representation of the existing model in its steady state

During the transient period, the ensemble model combines all three components to compensate for the lack of data. In the steady-state, the ensemble model only uses the LR model. In general, the model's performance improves as the amount of available historical data increases. Hence, this method performs well with existing customers who reached steady-state by accumulating large numbers of past deliveries. Since the current method heavily relies on historical data, the method's forecasting accuracy diminishes for new customers who lack historical data.

1.6 Problem Statement

This thesis addresses the following business and mathematical problem:

- **Business statement:** To lower the operating costs for a heating oil sales and distribution company by improving new customers' heating oil demand forecast for initial deliveries.
- **Mathematical statement:** Develop a Bayesian forecasting method that reduces the error between the new customers' forecast and actual heating oil demand during initial deliveries.

The preceding sections provided an overview of the problem and the context of this project. The next two sections focus on the details of the project and the solution: proposed solution, assumptions, and evaluation methods.

1.7 Assumptions

This section outlines the key assumptions of this research project. The summary of each assumption and the reasons why it is necessary are outlined below.

1.7.1 Availability of Historical Data

Target customer is a customer whose future heating oil consumption is being forecast. **Existing customer** is a customer whose past forecast and delivery amount are known to the forecaster at the time of the forecast. This project focuses on cases where there are little to no historical data available for the **target customers**. This project, however, assumes that sufficient historical data for **existing customers** are available at the time of the forecast. This distinction is important since this research is *not* about forecasting demand without *any* historical data, but about *forecasting demand of the target customer using historical data from other existing customers.*

It should be noted that this approach is similar to *surrogate modeling*, which mimics or forecasts the behavior of an original system by constructing a surrogate system using samples taken from the original system [12]. Hence, a surrogate method can be used to predict the behavior of a surrogate system (target customers) by taking samples (historical data) from the original system ("donor" customers) [7].

1.7.2 Demand Forecast and Actual Use

Since the actual use cannot be measured directly, the actual use is assumed to be the amount delivered during a delivery. This assumption holds well if the tank is always filled to its capacity. However, there are cases when the amount delivered does not equal the actual use. For example, the tank sometimes is not fully refilled because the delivery truck ran out of oil or the shutoff valve prematurely triggered. These special cases are handled by an underfill-overfill detection mechanism.

When the tank is not fully refilled, the amount delivered is likely to be significantly less than the estimated demand. Hence, this is known as an underfill condition. Typically, the company schedules another delivery to finish filling the tank. The amount delivered during this followup delivery is likely to be significantly more than the estimated demand. Hence, this is known as an overfill condition. The detection mechanism detects this condition by checking previous deliveries for each customer. If a customer has an underfill delivery immediately followed by an overfill delivery, then the mechanism replaces the two deliveries with a single artificial delivery computed by adding the delivery amount and heating degree days from the two deliveries.

1.7.3 Stationarity of the $\mathcal{K}$-factor

$\mathcal{K}$-factor describes how fuel efficient a customer is. Unless there is a major change in the behavior of a customer (long vacation away from home, major house renovations that improved insulation, installation of a new furnace, newborn infant in the house, aging parents visiting, etc.), $\mathcal{K}$-factor should remain relatively constant. This is important for our Bayesian Heating Oil Forecaster because the proposed method is trying to unbias the estimates to match the true $\mathcal{K}$-factor. If the customer's true $\mathcal{K}$-factor is changing frequently, then the adjustment becomes non-trivial. Although the $\mathcal{K}$-factor for a customer may change over a long period of time (i.e., several years), it is very unlikely to change significantly over a short period of time (i.e., during the first few deliveries). *Hence, for the scope of the thesis, the short-term $\mathcal{K}$-factor is assumed to be stationary.*

1.7.4 Positive and Negative Errors

In many forecasting applications, positive and negative errors carry different meanings, consequences, and associated costs. In this thesis, errors are defined as Estimated demand − Actual demand $= \widehat{s_k} - s_k$. Hence, a positive error is reported when the estimated demand is larger than the actual demand. A negative error is reported when the estimated demand is smaller than the actual demand. In the area of heating oil forecasting, positive errors increase the number of unnecessary deliveries because tanks are considered to have less oil than they actually contain. This can increase the overall operation cost for the company. Negative errors might result in customers running out of oil because the tank is estimated to have more oil than it actually contains. Customers who run out of fuel typically switch supplier, which results in lost revenue for the heating oil company. *Since costs associated with loss of customers are much higher than expected increases in operational costs, the negative errors are less desirable than positive errors.*

The assumptions discussed in this section are applicable at a conceptual level. Mathematical assumptions that apply to the estimation process are discussed in later chapters. The next section discusses the evaluation method used to determine the effectiveness of our Bayesian Heating Oil Forecaster.

1.8 Evaluation

This section briefly outlines the evaluation method and defines the criteria of an acceptable solution. This project is successful if our Bayesian Heating Oil Forecaster consistently produces better forecasts compared to the existing method for the same set of initial customers.

The comparison of the current and proposed methods is performed using a backtesting system. The backtesting system performs two sets of ex-post forecasts and compares the forecasting error of the existing forecasting method against the error of our Bayesian Heating Oil Forecaster. The comparison involves the following five steps as shown in Figure 1.7:

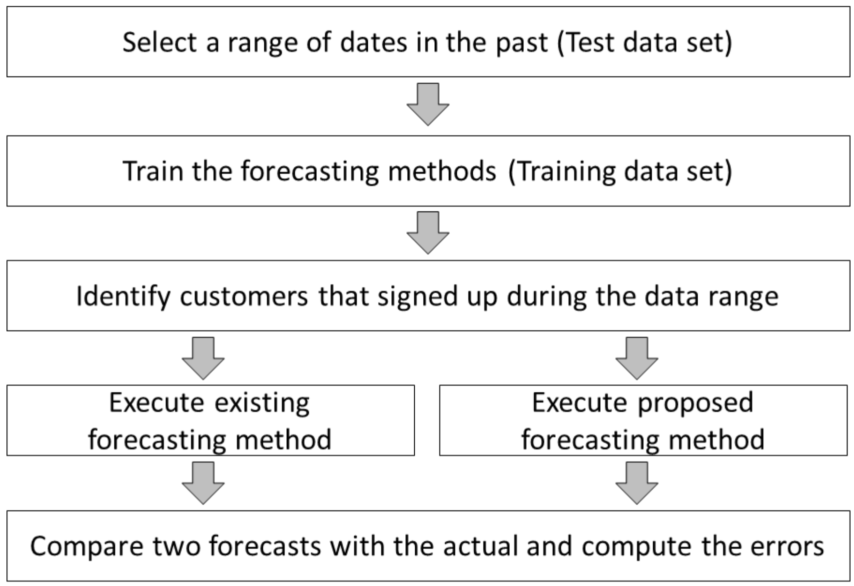

Figure 1.7: Steps of the evaluation method

1. The tester specifies a range of dates to use as a test data set. Simulated forecasts that occur during this date range are used to evaluate and compare the existing forecasting method and our Bayesian Heating Oil Forecaster.

2. The backtesting system identifies new customers that signed up during the dates specified.

3. The backtesting system trains the existing forecasting method and our Bayesian Heating Oil Forecaster using training data set that are available up to the beginning of the test data set.

4. The backtesting system performs ex-post forecast using the existing forecasting method for the customers identified in the previous step. This produces a set of forecasts from the existing method for each of the new customers.

5. The backtesting system performs ex-post forecast using our Bayesian Heating Oil Forecaster for the same set of customers. This produces a set of forecasts from the proposed method for each of the new customers.

6. The backtesting system compares the two sets of forecasts with the actual delivery amount, computes the errors between them, and reports the result. In general, the method with a smaller forecasting error is considered the better method. The error is measured in Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE), as well as a weighted error measure which assigns a larger weight to negative errors (potential loss of customers) than positive errors (potential increase in operational costs).

A more thorough discussion of the evaluation process and the backtesting system can be found in Chapter 3. The next section briefly reviews the structure of the remainder of this thesis.

1.9 Organization of this Thesis

This thesis consists of five chapters. Chapter 1 introduced the project background, current process, and the details of the problem being addressed. Chapter 2 provides an overview of the Bayesian forecasting techniques applied to various real-world data sets. Chapter 3 introduces our Bayesian Heating Oil Forecaster and how it applies to our test data set. Chapter 4 compares the results of our Bayesian Heating Oil Forecaster with those of the existing forecasting method. Finally, Chapter 5 offers conclusions and opportunities for further research.

CHAPTER 2

Survey of Energy Forecasting Literature

Chapter 1 introduced and outlined the problem of forecasting new customers' heating oil demand. Chapter 2 provides an overview of existing demand forecasting techniques, such as Multiple Linear Regression, Artificial Neural Network, and Ensemble forecasting. Examples of Bayesian forecasting techniques, such as Bayesian Network and Dynamic Linear Model, are discussed. This chapter also contains an overview of the mathematical concepts such as regression analysis and Bayes' Theorem. These are fundamental concepts used in our Bayesian Heating Oil Forecaster presented in Chapter 3.

2.1 Existing Demand Forecasting Methods

Multiple Linear Regression, Artificial Neural Network, and Ensemble forecasting are three forecasting methods that have been applied successfully to demand forecasting, namely natural gas daily demand forecasting [48]. This section presents an overview of these methods.

2.1.1 Multiple Linear Regression

A multiple linear regression model expresses the dependent variable as a function of one or more independent variables assuming a linear relationship [6; 48]. Suppose we want to forecast a daily demand S on a k^{th} day in the future, using m independent variables, $x_{k,j}$, where $j = 1, \ldots, m$. Then the estimated daily demand on the k^{th} day is

$$s_k \approx \widehat{s_k} = \beta_0 + \sum_{j=1}^{m} \beta_j x_{k,j} \ ,$$

where β_js are parameters that describe how independent variables are related to the estimated daily demand. The independent variable $x_{k,1}$ may represent Heating Degree Days, while β_0 is the baseload, and β_1 is the heatload coefficient.

Multiple linear regression extrapolates very predictably, adapting well to situations where the inputs are different from past observations. However, multiple linear regression performs poorly when the linearity assumption does not hold. Since past observations are used to estimate the parameters, a multiple linear regression model requires historical data. Generally, the more historical data is available, the better the parameter estimates [48].

A more thorough discussion of the Multiple Linear Regression technique can be found in introductory textbooks such as *Forecasting, Time Series, and*

Regression: An Applied Approach [6] and *Introduction to Linear Regression Analysis* [32].

2.1.2 Artificial Neural Network

Another tool commonly used for estimation and forecasting is an Artificial Neural Network (ANN). An ANN maps an unknown nonlinear relationship between the inputs and the output. This mapping is accomplished through a training process during which the ANN learns from past observations. Because an ANN handles nonlinear relationships, multiple related factors, such as temperature, wind speed, and prior day temperatures can be used as inputs [48].

An ANN excels when the inputs are similar to, but not the same as, the training data. However, an ANN does not perform as well in cases where the inputs are beyond the domain of the training knowledge. For example, the accuracy of an ANN diminishes when it forecasts natural gas demand for the coldest day on record. Since an ANN must be trained using past observations to expand the domain of the training knowledge, it is not suitable for situations where there is little historical data [48].

A more thorough discussion of Artificial Neural Networks can be found in introductory textbooks such as *An Introduction to Neural Network* [25] and

Gateway to Memory: An Introduction to Neural Network Modeling of the Hippocampus and Learning [23].

2.1.3 Ensemble and Combined Forecasts

Ensemble forecasting combines multiple forecasts produced by different forecast methods to obtain a single forecast with variance smaller than the variance of any of the components. Various factors influence dependent variables, and factors that are captured by any one of the forecast methods might be incomplete and limited. However, multiple forecast methods can better capture these factors when combined together. The combined forecast tends to reduce the effects of faulty assumptions, bias, or mistakes in data [2; 48]. As a result, combined forecasts almost unanimously increases forecast accuracy, regardless of the nature of the forecast [9]. Even simple averaging, the most simple combination method, is shown to improve the performance of the forecast [2]. In general, forecasts are combined by taking an weighted average of multiple independent forecasts, or according to a set of rules. Weights are calculated according to a repeatable rule, such as equal weighting, domain knowledge, and past forecast accuracy. Other methods include voting, simulation, combiner, stacked generalization, principle component analysis, singular value decomposition, and artificial neural networks [15]. As specific examples of existing ensemble forecasting techniques, Dhillon cites Fan et al. [17], whose work introduces and compares combiner and stacked generalization, which are meta-learning techniques that improves the performance of a single classifier by combining multiple classifiers. Araújo and New [1] apply ensemble forecasting frameworks, such as the bounding box, consensus, and probabilistic techniques, to improve the robustness of bioclimatic modeling.

Readers who are interested in additional materials should also refer to an annotated bibliography by Clemen [9]. Clemen offers a brief overview, historical development, and an extensive list of over 200 applied and theoretical articles covering various combined forecasting techniques.

This concludes the brief overview of the existing forecasting techniques used in energy demand forecasting. The following section discusses the Bayesian approach to probability and forecasting.

2.2 Bayes' Theorem, Bayesian Probability, and Bayesian Inference

Various Bayesian techniques discussed in the remainder of this thesis, including our Bayesian Heating Oil Forecaster, take advantage of the Bayesian approach to forecasting. This section provides an introduction to Bayes' Theorem to gain a better understanding of the Bayesian approach to forecasting, and how various Bayesian forecasting techniques are implemented. Materials and discussions contained in this and later sections are drawn from textbooks on Bayesian forecasting such as *Introduction to Bayesian Statistics* [5] and *Statistics: A Bayesian Perspective* [4]. Both are introductory statistics textbooks that extensively use Bayesian inference. The latter book is recommended especially for readers interested in a solid review of probability theory. *Introduction to Bayesian Statistics* [5] is for upper level undergraduate students with a background in calculus and probability theory. It offers in-depth discussions of Bayesian probability and statistics.

Bayes' Theorem was proposed by Reverend Thomas Bayes in the 18th century and was later extended by Laplace in the 19th century [36; 47]. From a statistical inference perspective, the theorem is significant because it allows one to infer the probability of a cause when its effect is observed [36]. In other words, Bayes' Theorem helps answer questions such as "I have a stiff neck (effect). How likely am I to have a meningitis (cause)?", see Figure 2.1.

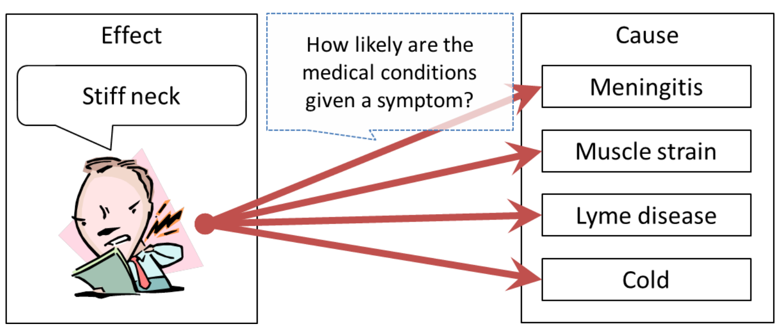

Figure 2.1: How likely is a cause given the effect? [33]

Bayes' Theorem can also be viewed as a thought process. It dictates the way in which the probabilities change in the light of evidence [4]. In other words, Bayes' Theorem describes mathematically the process by which forecasters update their knowledge in response to an observed event, as suggested by Figure 2.2.

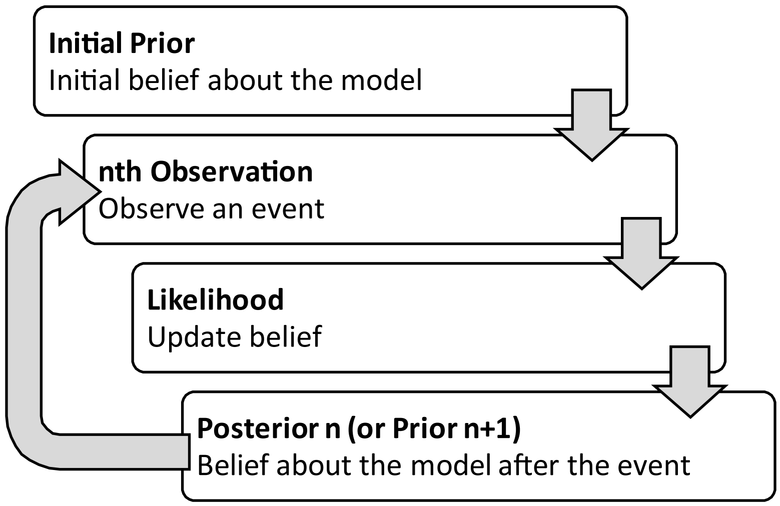

Figure 2.2: A model of how Bayes' Theorem updates forecaster knowledge

The knowledge of the forecaster is represented mathematically using probability distributions. The update process can be described using three distinct probability distributions:

Prior represents our knowledge before we observe evidence. The prior probability of an event A is expressed as $P(A)$.

Likelihood represents a factor that is used to update our prior knowledge. The likelihood for an event A and an evidence B is expressed in terms of a conditional probability $P(B|A)$.

Posterior represents our knowledge after we observe evidence. The posterior probability of an event A given the evidence B is expressed in terms of a conditional probability $P(A|B)$.

In summary, Bayes' Theorem says

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

It states that the posterior is proportional to the product of the prior and the likelihood. In other words, we can obtain our posterior knowledge by 1) multiplying our prior and the likelihood and 2) scaling the product.

Figure 2.2 graphically represents how the forecasters update their knowledge using the prior, likelihood, and posterior distributions. As seen in the diagram, the update process is iterative: The current posterior becomes the prior of the next step. The process iterates when a new event is observed.

The following two sections further describe Bayes' Theorem using simple examples. The first section describes the theorem using discrete probability distributions. The second section describes the theorem using continuous probability distributions.

2.2.1 Discrete Bayesian Analysis

This section applies Bayes' Theorem using two separate examples. The first example is a very simple balls-in-an-urn example drawn from Bolstad [5]. This example illustrates how the prior, likelihood, and posterior distributions interact to update the forecaster's knowledge about a model. The second example involves forecasting the relative strength of two basketball teams. The basketball example, drawn from Berry [4], illustrates how to apply Bayes' Theorem to perform forecasts. Later, the second example is extended to illustrate the difference between discrete and continuous Bayesian forecasting.

Example: Balls-in-an-urn

Suppose there is an urn with five balls inside. The balls are colored either red or blue, but we cannot see the contents of the urn. The objective is to estimate the number of red balls in the urn by drawing a ball out of the urn one by one without replacement. Since we are interested in the number of red balls, let the random variable X be the number of red balls in the urn. If we draw a ball from the urn, the color of the ball is either red or blue. To represent this mathematically, let the random variable $Y = 1$ if the draw is red, and $Y = 0$ if the draw is blue. Figure 2.3 summarizes this setup.

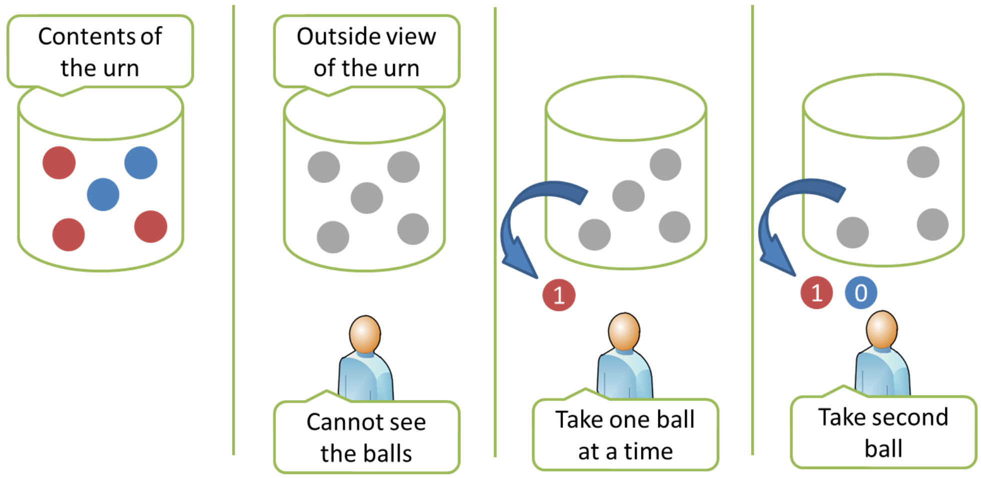

Figure 2.3: The balls-in-an-urn example

Prior and posterior beliefs

As stated above, our objective is to estimate the number of red balls in the urn. Hence, our belief is our estimate of the number of red balls in the urn. Our prior belief is our estimate of the number of red balls in the urn *before* we draw a ball. Our posterior belief is our estimate of the number of red balls in the urn *after* we draw a ball. Note that our prior and posterior beliefs change as we continue to draw the balls out of the urn. For example, our first prior belief (denoted Prior(1)) is our estimate of the number of red balls in the urn **before we draw the first ball** out of the urn. Our first posterior belief (denoted Posterior(1)) is our estimate of the number of red balls in the urn **after we draw the first ball** out of the urn. Posterior(1) is also our Prior(2) because Posterior(1) is our estimate *before* we draw the second ball. If we define n to be the observation number, then this relationship can be summarized as

$$\text{Posterior}(n) = \text{Prior}(n+1) \text{ for } n \geq 1.$$

First prior belief

Although Prior(n) for $n \geq 2$ are computed iteratively, we have to present our own estimate for Prior(1). Initially, we know that the total number of balls in the urn is five, but we have no idea how many of them are red. What we know for sure is that the number of red balls can be only 0, 1, 2, 3, 4, or 5. In this case, we might assume that all possible outcomes are equally likely. Translating this prior knowledge into probability gives

$$P(X=0) = P(X=1) = ... = P(X=5) = 1/6, \text{ and}$$

$$P(X<0) = P(X>5) = 0.$$

Likelihood

Likelihood is the probability of observing an evidence given the truth. The "evidence" is the color of the ball we draw from the urn. The "truth" is the actual number of red balls in the urn. In other words, it describes how "likely" it is to draw a ball with a certain color if the number of red balls in the urn is either 0, 1, 2, 3, 4, or 5. For instance, $P(Y=1|X=2)$ represents the likelihood (probability) of drawing a red ball from the urn if the number of red balls in the urn is 2. Since there are 5 balls in the urn, the likelihood of drawing a red ball from the urn when there are 2 red balls in the urn is 2 out of 5. Using the notation for conditional probability, this can be written as

$$P(\text{ draw red ball } | \text{ number of red ball in the urn } = 2) = P(Y = 1|X = 2) = 2/5.$$

The likelihood changes as the observation (the color of the ball drawn) changes. For example, $P(Y = 0|X = 2)$ represents the likelihood (probability) of drawing a blue ball from the urn if the number of red balls in the urn is 2. Since there are a total of 5 balls, if there are 2 red balls, then the remaining 3 would be blue. Hence,

$$P(\text{ draw blue ball } | \text{ number of red ball in the urn } = 2) = P(Y = 0|X = 2) = 3/5.$$

Update Using Joint Probability

In this example, we have two different random variables, $X =$ number of red balls in the urn, and $Y =$ color of the ball. The probability that $X = x_i$ and $Y = y_i$ occur simultaneously is called the joint probability,

$$f(x_i, y_i) = P(X = x_i, Y = y_i).$$

Using this notation, the probability that *the number of red balls in the urn = 2* and *draw a red ball* occurring simultaneously is expressed as $f(2,1) = P(X = 2, Y = 1)$. Since we have a total of 5 balls and 2 colors, there are 10 possible joint probabilities. The 10 joint probabilities together form a **joint probability distribution** of the random variables X and Y. A joint probability distribution represents the probability of all possible combinations of the joint random variables, and can be expressed in a table form as shown in Table 2.1.

		Color of ball drawn (Y)	
		0 (Blue)	1 (Red)
No of red balls in urn (X)	0	$f(0,0)$	$f(0,1)$
	1	$f(1,0)$	$f(1,1)$
	2	$f(2,0)$	$f(2,1)$
	3	$f(3,0)$	$f(3,1)$
	4	$f(4,0)$	$f(4,1)$
	5	$f(5,0)$	$f(5,1)$

Table 2.1: Joint probability distribution for the balls-in-an-urn example

Individual joint probability can be computed using the following relationship:

$$f(x_i, y_i) = g(x_i) \times f(y_i|x_i), \text{ and} \tag{2.1}$$

$$P(X = x_i \wedge Y = y_i) = P(X = x_i) \times P(Y = y_i|X = x_i). \tag{2.2}$$

Since we update our prior belief by multiplying our prior belief and an appropriate likelihood, calculating the joint probability is equivalent to updating our prior belief using an appropriate likelihood. The joint probabilities for the case when the first ball picked is red can be computed as follows:

$$f(0,1) = P(X = 0) \times P(Y = 1|X = 0) = 1/6 \times 0/5 = 0$$

$$f(1,1) = P(X = 1) \times P(Y = 1|X = 1) = 1/6 \times 1/5 = 1/30$$

$$f(2,1) = P(X = 2) \times P(Y = 1|X = 2) = 1/6 \times 2/5 = 2/30$$

$$f(3,1) = P(X = 3) \times P(Y = 1|X = 3) = 1/6 \times 3/5 = 3/30$$

$$f(4,1) = P(X = 4) \times P(Y = 1|X = 4) = 1/6 \times 4/5 = 4/30$$

$$f(5,1) = P(X = 5) \times P(Y = 1|X = 5) = 1/6 \times 5/5 = 5/30.$$

If we repeat the calculation for the case when the ball is blue, then we can obtain a full joint probability distribution as shown in Table 2.2.

		Color of ball drawn (Y)	
		0 (Blue)	1 (Red)
No of red balls in urn (X)	0	$\frac{5}{30}$	$\frac{0}{30}$
	1	$\frac{4}{30}$	$\frac{1}{30}$
	2	$\frac{3}{30}$	$\frac{2}{30}$
	3	$\frac{2}{30}$	$\frac{3}{30}$
	4	$\frac{1}{30}$	$\frac{4}{30}$
	5	$\frac{0}{30}$	$\frac{5}{30}$

Table 2.2: Joint probability distribution for the balls-in-an-urn example with joint probabilities calculated

If the first ball was red, then the column in the joint probability distribution with $Y = 1$ is our posterior knowledge, except that the sum of the products of the priors and the likelihoods equal to $1/2$. Since our knowledge must be expressed in terms of probability, the sum must equal to 1. This can be accomplished by dividing (scaling) the products by the sum of the products.

Repeat

As we repeat the drawings, we also repeat the calculations. For the second draw, the posterior we obtained after the first draw becomes our new prior. The update process continues as we draw more balls from the urn. The actual calculations are shown in Tables 2.3 and 2.4. The posterior probability in Table 2.3 replicates the above calculation and shows the case when the first ball drawn is red. The posterior probability in Table 2.3 is when the second ball drawn is blue.

x_i (No. of red)	Prior	Likelihood	Prior $\times$ Likelihood	Posterior
0	$\frac{1}{6}$	$\frac{0}{5}$	$\frac{1}{6} \times \frac{0}{5} = 0$	$0/\frac{1}{2} = 0$
1	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{6} \times \frac{1}{5} = \frac{1}{30}$	$\frac{1}{30}/\frac{1}{2} = \frac{1}{15}$
2	$\frac{1}{6}$	$\frac{2}{5}$	$\frac{1}{6} \times \frac{2}{5} = \frac{2}{30}$	$\frac{2}{30}/\frac{1}{2} = \frac{2}{15}$
3	$\frac{1}{6}$	$\frac{3}{5}$	$\frac{1}{6} \times \frac{3}{5} = \frac{3}{30}$	$\frac{3}{30}/\frac{1}{2} = \frac{3}{15}$
4	$\frac{1}{6}$	$\frac{4}{5}$	$\frac{1}{6} \times \frac{4}{5} = \frac{4}{30}$	$\frac{4}{30}/\frac{1}{2} = \frac{4}{15}$
5	$\frac{1}{6}$	$\frac{5}{5}$	$\frac{1}{6} \times \frac{5}{5} = \frac{5}{30}$	$\frac{5}{30}/\frac{1}{2} = \frac{5}{15}$
Sum	1		$\frac{1}{2}$	1

Table 2.3: Posterior probability distribution after the first observation

This example introduced the concept and the relationships of prior belief,

x_i (No. of red)	Prior	Likelihood	Prior × Likelihood	Posterior
0	0			0
1	$\frac{1}{15}$	$\frac{4}{4}$	$\frac{1}{15} \times \frac{4}{4} = \frac{1}{15}$	$\frac{1}{15}/\frac{1}{3} = \frac{2}{10}$
2	$\frac{2}{15}$	$\frac{3}{4}$	$\frac{2}{15} \times \frac{3}{4} = \frac{1}{10}$	$\frac{1}{10}/\frac{1}{3} = \frac{3}{10}$
3	$\frac{3}{15}$	$\frac{2}{4}$	$\frac{3}{15} \times \frac{2}{4} = \frac{1}{10}$	$\frac{1}{10}/\frac{1}{3} = \frac{3}{10}$
4	$\frac{4}{15}$	$\frac{1}{4}$	$\frac{4}{15} \times \frac{1}{4} = \frac{1}{15}$	$\frac{1}{15}/\frac{1}{3} = \frac{2}{10}$
5	$\frac{5}{15}$	0	$\frac{5}{15} \times 0 = 0$	$0/\frac{1}{3} = 0$
Sum	1		$\frac{1}{3}$	1

Table 2.4: Posterior probability distribution after the second observation

likelihood, and posterior belief. The update process illustrates how Bayesian inference is applied to estimate the probability of an unknown and unobservable quantity (number of red balls in the urn) in light of evidence (the color of the ball that is drawn from the urn). The next example focuses more on how to apply Bayesian inference in the context of forecasting.

Example: Relative strength of two basketball teams

This example is adapted from a similar example presented by Berry [4]. Consider two basketball teams: MU and UC. The two teams belong to the same conference, and have several games each season. Our objective is to estimate the relative strength of MU, and *forecast the probability of MU winning the next game* against UC. A relative strength of 0 means MU can never win over UC. A relative strength of 1 means MU can always win over UC. If the relative strength is 0.8, then

MU is expected to win over UC for 80 percent of the time. Suppose the season just started so that the two teams have not met this season. Since we are interested in the relative strength of MU, let the random variable X be the relative strength of MU. To simplify the problem, let us also assume that the teams will either win or lose and will never end a game with a tie. To represent this mathematically, let the random variable $Y = 1$ if MU wins, and $Y = 0$ if UC wins. Figure 2.4 summarizes this setup.

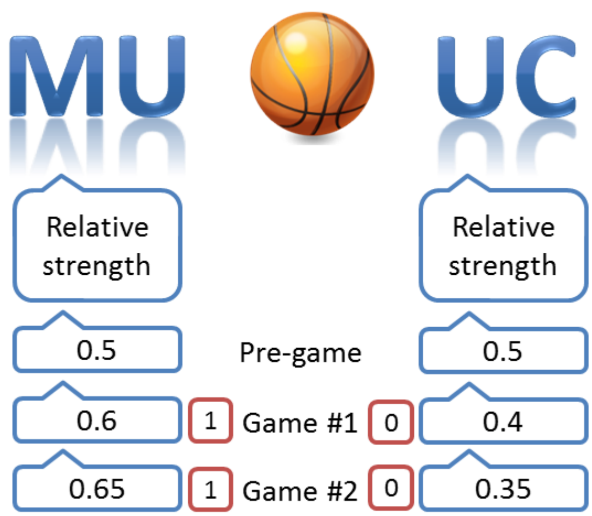

Figure 2.4: Relative strength of basketball teams example

Prior and posterior beliefs

Our objective in this example is to estimate the relative strength of MU by observing games between MU and UC so that we can forecast the winner of the next game. Hence, our belief is our estimate of the relative strength of MU. Using the same notation presented in the balls-in-an-urn example, our prior belief, Prior(n), is our estimate of the relative strength of MU *before* we observe the n^{th} game of the current season. Our posterior belief, Posterior(n), is our estimate of the relative strength of MU *after* we observe the n^{th} game.

First prior belief

Next, we have to present our own estimate for Prior(1). Initially we know that the relative strength can range between 0 and 1. To simplify the example, we discretize the range by 0.1 increments so that the only possible relative strengths are 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0. Additionally, say we are informed that MU was stronger than UC during last year's season, but both teams won at least once during the same period. It seems reasonable to assume that relative strengths of 0.0 and 1.0 are unlikely, and relative strengths greater than 0.5 are more probable than relative strengths less than 0.5. Translating this prior knowledge into probability might look like Table 2.5:

Strength of MU	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	Sum
Probability	0%	2%	3%	5%	8%	12%	22%	26%	17%	5%	0%	100%
Strength × Probability	0.00	0.00	0.01	0.02	0.03	0.06	0.13	0.18	0.14	0.05	0.00	0.61

Table 2.5: An example of a possible Prior(1) for the basketball example

Notice that unlike our first example, we have assigned unequal probabilities to different strengths. These probabilities can be our best guess, since the update process adjusts these estimates based on future observations. Our objective is to forecast the winner of the next game. This can be accomplished by computing the predicted relative strength of MU. The predicted relative strength of MU is computed by multiplying each of the possible relative strengths of MU by its probability and adding the products. The figure shows that the predicted relative strength of MU is 0.61. This matches our expectation since we assumed that MU might continue to be stronger than UC during the current season.

Likelihood

Likelihood is the probability of observing an evidence given the truth. The "evidence" is the result of the game. The "truth" is the actual relative strength of MU. In other words, it describes how "likely" it is for MU to win if the actual relative strength of MU is either 0, 0.1, 0.2, ..., 0.9, or 1.0. For instance, $P(Y = 1|X = 0.2)$ represents the likelihood (probability) of MU winning the game if the relative strength of MU is 0.2. If the relative strength of MU is 0.2, then the likelihood of MU winning the game is also 0.2. Using the notation for conditional probability, this can be written as

$$\begin{aligned}\text{Likelihood} &= P(\text{ MU wins} \mid \text{actual relative strength of MU } = 0.2)\\ &= P(Y = 1|X = 0.2)\\ &= 0.2.\end{aligned}$$

The likelihood changes when the outcome is different. For example, the likelihood of

MU losing the game if its actual relative strength is 0.2 is

$$\begin{aligned}\text{Likelihood} &= P(\text{ MU loses } \mid \text{ actual relative strength of MU } = 0.2)\\ &= P(Y=0|X=0.2)\\ &= 1-0.2\\ &= 0.8.\end{aligned}$$

Table 2.6 shows two different likelihoods for all possible relative strengths (0.0, 0.1, ..., 1.0; columns in the figure) and for all possible outcomes (MU wins, MU loses; rows in the figure).

Strength of MU	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Likelihood of MU winning	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Likelihood of MU losing	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0

Table 2.6: Likelihood for the basketball example

Update

Suppose MU won the first game against UC. We update our prior belief by multiplying our prior belief and the likelihood:

$$P(X = 0.0) \times P(Y = 1|X = 0.0) = 0.0 \times 0.0 = 0$$

$$P(X = 0.1) \times P(Y = 1|X = 0.1) = 0.02 \times 0.1 = 0.002$$

$$...$$

$$P(X = 0.9) \times P(Y = 1|X = 0.9) = 0.05 \times 0.9 = 0.045$$

$$P(X = 1.0) \times P(Y = 1|X = 1.0) = 0.0 \times 1.0 = 0.$$

Table 2.7 shows the update calculation for all possible relative strengths (0.0, 0.1, ..., 1.0; columns in the figure). The posterior row is computed by dividing (scaling) the products by the sum.

Model	1	2	3	4	5	6	7	8	9	10	11	Sum
Strength	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	-
Prior	0%	2%	3%	5%	8%	12%	22%	26%	17%	5%	0%	100%
Likelihood (MU wins)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	-
Prior × Likelihood	0.000	0.002	0.006	0.015	0.032	0.060	0.132	0.182	0.136	0.045	0.000	0.61
Posterior	0.0%	0.3%	1.0%	2.5%	5.2%	9.8%	21.6%	29.8%	22.3%	7.4%	0.0%	100%
Strength × Posterior	0.000	0.000	0.002	0.007	0.021	0.049	0.130	0.209	0.178	0.066	0.000	0.66

Table 2.7: Update for the basketball example

The posterior belief is our estimate of the relative strength of MU after the first game. The predicted relative strength of MU after the first game is computed by multiplying each of the possible relative strengths of MU by its probability and adding the products. Table 2.7 shows that the predicted relative strength of MU is

0.66, which is larger than the initial estimate of 0.61. This matches our expectation since MU just won the first game.

This example illustrates the application of Bayesian inference to estimate the probability of an unknown and unobservable quantity (relative strength of MU) in light of evidence (results of the game). If we generalize this to energy (heating oil) demand forecasting, Bayesian inference can be used to estimate the probability of an unknown and unobservable quantity ($\mathcal{K}$-factor) in light of evidence ($\mathcal{K}$-factor that is observed between deliveries). Once the $\mathcal{K}$-factor is known, the heating oil demand can be computed using a regression model.

Since the $\mathcal{K}$-factor is a continuous quantity, the next section discusses the difference between the discrete and continuous approaches to Bayesian inference.

2.2.2 Continuous Bayes Inference

The logical steps of computing the continuous Bayesian inference is identical to its discrete counterpart: we start with a prior belief, observe an event, update our belief, and compute the posterior belief. What differs between the two are the use of continuous random variables and probability distributions.

The balls-in-an-urn example is a discrete example since the quantity (number of balls; $1, 2, 3, 4, 5$) as well as the possible outcomes (red/blue) are discrete. The basketball example has a continuous quantity (relative strength; ranging from (0...1)) and discrete outcomes (win/lose). Any continuous probability distribution can be used to describe the continues random variable, including but are not limited to uniform, beta, gamma, normal, and empirical distributions [5].

Empirical Distribution

An empirical distribution is a probability distribution that is generated directly from the observed (sample) data. It represents the estimated probability of a certain observation occurring in the population. A histogram is a scaled version of the empirical probability density function. An empirical PDF is computed by scaling the histogram: $\frac{\text{count}}{\text{sample size}\times\text{bin width}}$.

Beta Distribution

A beta distribution frequently is used in the context of Bayesian estimation because it drastically simplifies the update process [4; 5]. A beta distribution is parameterized by two parameters, often denoted by a and b. The distribution itself is sometimes denoted as $\beta(a, b)$. The probability function of the beta distribution

$\beta(a,b)$ is

$$\begin{aligned} P(x) &= f(x; a, b) \\ &= \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \\ &= \frac{(a+b-1)!}{(a-1)!(b-1)!} x^{a-1}(1-x)^{b-1}. \end{aligned}$$

Beta distributions have the following properties that help simplify the update process [4; 5]:

- The product of two beta distributions is a beta distribution, and
- Multiplication of two beta distributions can be accomplished by adding their parameter values.

The above properties are demonstrated below:

$$\begin{aligned} f(x; a_1, b_1) \times f(x; a_2, b_2) &= \frac{1}{B(a_1, b_1)B(a_2, b_2)} x^{(a_1-1)}(1-x)^{(b_1-1)} x^{(a_2-1)}(1-x)^{(b_2-1)} \\ &= \frac{1}{B(a_1 + a_2 - 2, b_1 + b_2 - 2)} x^{(a_1+a_2-2)}(1-x)^{(b_1+b_2-2)} \\ &= f(x; a_1 + a_2 - 2, b_1 + b_2 - 2). \end{aligned}$$

The expected value of a beta distribution is computed from the parameter values,

$$E(X) = \frac{a}{a+b}. \tag{2.3}$$

Maximum Likelihood Estimation

We use distribution fitting techniques, such as maximum likelihood estimation (MLE), to fit a continuous distribution to a set of data. Maximum likelihood estimation is a statistical technique that identifies a probability distribution that makes the observed data most likely. In other words, it maximizes the likelihood $P($ observed data $|$ parameters $)$ for a set of probability distribution parameters and observed data. Since each probability distribution is different, the maximum likelihood estimation for each distribution is also different.

The maximum likelihood estimates for the Beta distribution are computed numerically based on the equation given by Johnson, Kotz, and Balakrishnan [29]. Others, such as Beckman and Tietjen [3] have developed a numerical technique in which the maximum likelihood estimates for the Beta parameters are computed.

Readers who are interested in an introduction to maximum likelihood estimates may read an article by Myung for a quick introduction [35]. Moore [34] uses a Gaussian distribution to step through the calculation process of maximum likelihood estimation. The NIST handbook also has an entry about likelihood estimation for Beta distributions [19].

2.3 Existing Bayesian Forecasting Methods

Bayesian forecasting methods have unique advantages over traditional forecasting methods. One advantage is their effectiveness during the initial transient period when little or no prior data is available [28]. This section provides an overview of Bayesian forecasting techniques that have been applied in areas including engineering, business, meteorology, and energy. This should help us see how the Bayesian forecasting algorithm, presented in Chapter 3, is related to other techniques that are already in use.

2.3.1 Bayesian Networks

A Bayesian Network is a probabilistic graphical model that often drastically reduces the computational complexity of the original problem [33]. The network is a graphical representation of the probabilistic relationships among many variables with cause-effect relationships [36]. The nodes in the network represent random variables, and edges represent dependence among the variables. A network as a whole represents a joint probability distribution over a set of random variables. In other words, the network represents all possible combinations of the joint random variables and their probabilities. It is a directed acyclic graph: each node is guaranteed not to be its own child or its own parent (Figure 2.5).

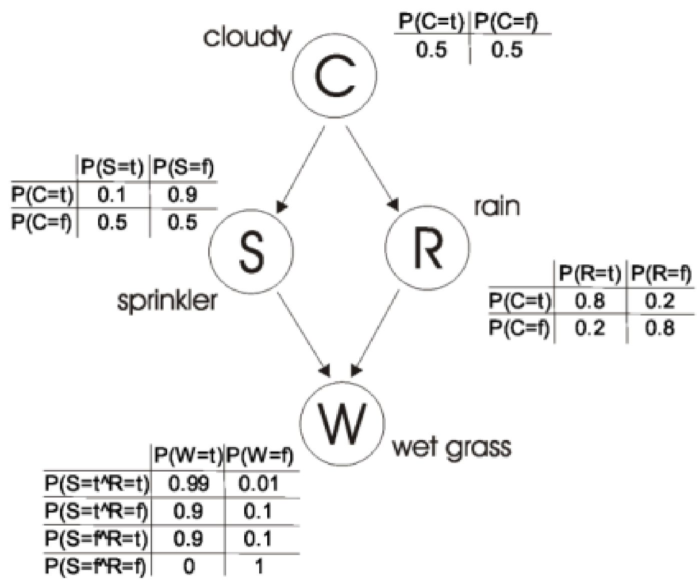

Figure 2.5: Example Bayesian network modeling causes of wet grass [45]

Each variable is only dependent on its parents, which means that they are independent from all other non-parent variables. This independence frequently enables the model to reduce the number of parameters compared to the model that does not account for such independence. It also drastically simplifies the joint probability distribution. Simplifying the joint probability distributions reduces the cost of computing the posterior probabilities. For cases where the joint probability distribution is large, the calculation becomes impractical without simplifying the joint probability distribution using a Bayesian Network [36; 43].

To illustrate how a Bayesian Network can be constructed and used, let us consider the following simple example adopted from a lecture by Moore [33]. You are at a small regional airport interested in estimating the probability of delays under various conditions. Suppose we have the following five events:

- S: It is sunny.
- M: The airline is Delta. (If not, then it is United.)
- R: The airplane is Boeing. (If not, it is Airbus.)
- L: The airplane arrives late.
- T: The airplane leaves on time.

5 variables

S	M	R	L	T	Probability
0	0	0	0	0	0.02
0	0	0	0	1	0.03
0	0	0	1	0	0.02
...					
0	1	0	1	1	0.03
0	1	1	0	0	0.04
0	1	1	0	1	0.04
...					
1	1	1	1	0	0.02
1	1	1	1	1	0.01
				Sum	1.00

$2^5 = 32$ rows

Figure 2.6: Example joint probability distribution [33]

Figure 2.6 is an example of a joint probability distribution that expresses the uncertainty involved in this problem. The joint probability distribution can be used to calculate various probabilities such as

- the probability that the airplane leaves on time, when it is raining ($S = 0$), the airline is Delta ($M = 1$), the airplane is Boeing ($R = 1$), and the airplane arrives on time ($L = 0$):

$$P(T|\neg S \wedge M \wedge R \wedge \neg L);$$

- the probability that the airplane leaves on time, when the airline is United and it is sunny:

$$P(T|\neg M \wedge S);$$

- the probability that the airplane arrives late, when the airplane is Airbus:

$$P(L|\neg R).$$

Specifying the entire joint probability distribution with five binary random variables requires 32 different probabilities. The following example uses the Bayesian Network to reduce the number of probabilities required to calculate the joint probability distribution from 32 to 10.

A Bayesian Network requires knowledge of the cause-effect relationships among the five variables. For this example, the following assumptions are made:

- Weather condition does not depend on and does not influence which airline is flying the aircraft.

- Weather condition does not depend on and does not influence the manufacturer of the aircraft.
- Once we know which airline is flying the aircraft, then whether it arrives late does not affect the manufacturer of the aircraft.
- Regardless of the airline, flights are frequently delayed due to bad weather.
- United is more likely to arrive late than Delta.
- United is more likely to use Boeing aircraft than Delta.
- Airplanes are more likely to leave on time if the airplanes arrived on time.

The first assumption describes the independence between weather condition and the airline. This is specified by the statement $P(S|M) = P(S)$. Similarly, the second assumption describes the independence between the weather condition and the aircraft manufacturer. This is specified by the statement $P(S|R) = P(S)$. The third assumption is a conditional independence between airplane manufacturer and the lateness of the flight given the airline. In other words, L and R are conditionally independent given M. This is specified by the statement $P(L|M, R) = P(L|M)$ and $P(R|M, L) = P(R|M)$. The fourth assumption indicates that weather condition influences the lateness of the flight. The fifth assumption indicates that the airline influences the lateness of the flight. Similarly, the sixth assumption indicates that the airline influences which manufacturer built the aircraft. The last assumption indicates that arriving late influences the probability of leaving on time. Expressing these assumptions using a Bayesian Network, we obtain Figure 2.7.

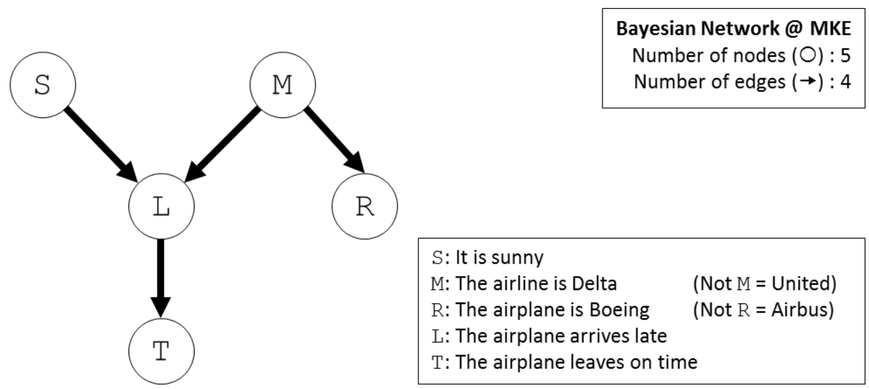

Figure 2.7: An example Bayesian Network that describes flight delays [33]

We used known relationships among variables to construct the Bayesian Network. The next step is to assign probabilities that describe each node. In general, the table for node A must list $P(a|\text{Parent values})$ for each possible combination of parent values. For example, node L is dependent on parents S and M. Hence, possible combinations are:

- $P(L|M \land S) = P(\text{Airplane arrives late}|\text{Airline is Delta} \land \text{Sunny})$
- $P(L|M \land \neg S) = P(\text{Airplane arrives late}|\text{Airline is Delta} \land \text{Rainy})$
- $P(L|\neg M \land S) = P(\text{Airplane arrives late}|\text{Airline is United} \land \text{Sunny})$
- $P(L|\neg M \land \neg S) = P(\text{Airplane arrives late}|\text{Airline is United} \land \text{Rainy})$

The probabilities for these values come from domain knowledge, observations, or experiments. Suppose we reviewed the on-time performance log published by the airport and empirically determined the probabilities:

- When the airline was *Delta* and it was *sunny*, the flight arrived late 5% of the time = $P(L|M \wedge S) = 0.05$.
- When the airline was *Delta* and it was *rainy*, the flight arrived late 10% of the time = $P(L|M \wedge \neg S) = 0.1$.
- When the airline was *United* and it was *sunny*, the flight arrived late 10% of the time = $P(L|\neg M \wedge S) = 0.1$.
- When the airline was *United* and it was *rainy*, the flight arrived late 20% of the time = $P(L|\neg M \wedge \neg S) = 0.2$.

Hence,

- $P(\neg L|M \wedge S) = 1 - 0.05 = 0.95$,
- $P(\neg L|M \wedge \neg S) = 1 - 0.1 = 0.9$,
- $P(\neg L|\neg M \wedge S) = 1 - 0.1 = 0.9$, and
- $P(\neg L|\neg M \wedge \neg S) = 1 - 0.2 = 0.8$.

Repeating this for all nodes in the graph, we obtain Figure 2.8 with 10 probabilities:

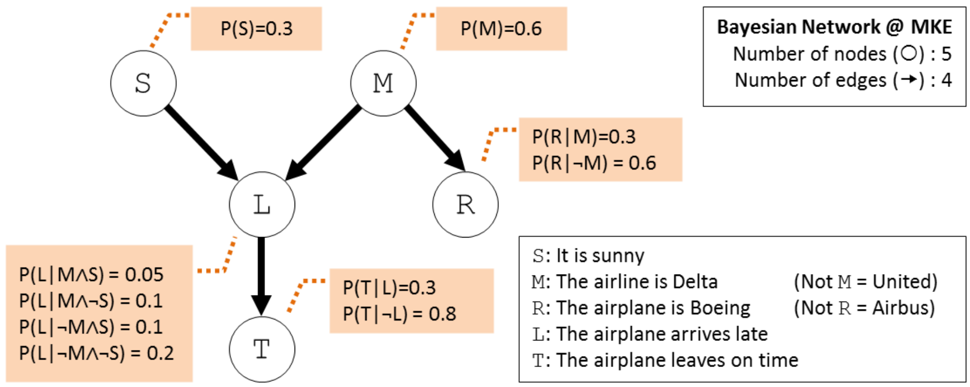

Figure 2.8: An example Bayesian Network with probabilities [33]

The Bayesian Network in Figure 2.8 only contains 10 probabilities. However, the full joint probability distribution is expressed using 32 probabilities (Figure 2.6). Using boolean arithmetic, we can compute any entry in the joint probability distribution using the 10 probabilities from the Bayesian Network. For example, to compute a table entry $P(S \wedge \neg M \wedge L \wedge \neg R \wedge T)$, we use Equations 2.1 and 2.2:

$$P(T \wedge \neg R \wedge L \wedge \neg M \wedge S)$$

$$= P(T|\neg R \wedge L \wedge \neg M \wedge S) \times P(\neg R \wedge L \wedge \neg M \wedge S).$$

Since the network indicates that L is the only immediate parent of T,

$$P(T|\neg R \wedge L \wedge \neg M \wedge S) \times P(\neg R \wedge L \wedge \neg M \wedge S)$$

$$= P(T|L) \times P(\neg R \wedge L \wedge \neg M \wedge S)$$

$$= P(T|L) \times P(\neg R|L \wedge \neg M \wedge S) \times P(L \wedge \neg M \wedge S).$$

Since the network indicates that M is the only immediate parent of R,

$$P(T|L) \times P(\neg R|L \wedge \neg M \wedge S) \times P(L \wedge \neg M \wedge S)$$

$$= P(T|L) \times P(\neg R|\neg M) \times P(L \wedge \neg M \wedge S)$$

$$= P(T|L) \times P(\neg R|\neg M) \times P(L|\neg M \wedge S) \times P(\neg M \wedge S)$$

$$= P(T|L) \times P(\neg R|\neg M) \times P(L|\neg M \wedge S) \times P(\neg M|S) \times P(S)$$

$$= P(T|L) \times P(\neg R|\neg M) \times P(L|\neg M \wedge S) \times P(\neg M) \times P(S).$$

Inserting probabilities from the network yields

$$P(T \wedge \neg R \wedge L \wedge \neg M \wedge S)$$

$$= P(T|L) \times P(\neg R|\neg M) \times P(L|\neg M \wedge S) \times P(\neg M) \times P(S)$$

$$= (0.3) \times (1 - 0.6) \times (0.1) \times (1 - 0.6) \times (0.3)$$

$$= 0.00144.$$

Following the same process, any entry in the joint probability distribution can be calculated using the Bayesian Network.

Bayesian Networks have been applied to a variety of time-series forecasting scenarios. For instance, Cofiño et al. [10] applied Bayesian Networks to forecast meteorological time series; rainfall in the Iberian peninsula. The researchers chose Bayesian Networks because existing techniques, such as regression, hidden Markov models, and neural networks, rely on past evidence collected from each of the individual weather stations. Bayesian Networks, on the other hand, are capable of modeling both temporal and spatial dependencies among weather stations. Zhang et al. [51] applied Bayesian Networks to forecast short-term time series; traffic flow in Beijing road links recorded every 15 minutes. The authors observed that existing time series models, such as ARIMA, seasonal ARIMA, Kalman filter, neural networks, non-parametric, simulation, local regression, ATHENA, and KARIMA, do not incorporate information from adjacent road links. Using the intuition that vehicles travel from one road link to another, the researchers use Bayesian Networks to model the temporal and spatial dependencies among the interconnected road links.

A common theme between both groups is that the Bayesian Networks leverage on the temporal and spatial relationships that were unaccounted by the existing methods. Heating oil forecasting faces similar issues with existing methods. Existing methods rely heavily on historical data and do not fully incorporate all of the information that is available. In addition to the historical data, Bayesian Networks take advantage of the temporal and spatial relationships among the observations. Unfortunately, such relationships are very weak among heating oil customers. Unlike weather or traffic flow, neither demand nor heatload sensitivity ($\mathcal{K}$-factor) "travel" between individual customers over time across geographic regions. Lacking clear temporal and spatial relationships, Bayesian Networks do not seem to apply well for forecasting heating oil demand. The next section discusses another Bayesian forecasting technique called Bayesian pooling, which also takes into account information that is overlooked by the existing methods.

2.3.2 Bayesian Pooling / Empirical Bayes

Bayesian pooling (also known as Bayesian shrinkage, empirical Bayes, or Stein estimation) is a forecasting technique that is designed to adapt rapidly to pattern changes. This rapid and accurate adaptation is accomplished by incorporating analogous time series when forecasting a single target time series [2]. Analogous time series are time series that are closely related (correlated) to each other. A set of analogous time series that follow a similar pattern is called the equivalence group. Analogous time series are incorporated to a forecast by combining local and group models. A **local model** is estimated for the target time series being forecast. A **group model** is estimated using the equivalence group's pooled data. The two models are combined using weights. The weights are inversely proportional to the variance of the parameter estimates. In other words, estimates that are more precise are emphasized, and estimates that are less precise are deemphasized. A summary of the calculation steps is:

1. Select an equivalence group and extract time series data,

2. Scale each time series,
3. Construct local and group models,
4. Combine two models using Bayesian shrinkage weights,
5. Forecast using the combined (pooled) model, and
6. Rescale the forecasts to match the raw data.

Figure 2.9 is a graphical summary of the calculation steps. First, an equivalence group is selected from a set of analogous time series. Time series in the equivalence groups must be scaled so that the magnitude is standardized. An equivalence group that is scaled is called pooled data. The target time series is used to construct the local model, while the pooled data is used to construct the group model.

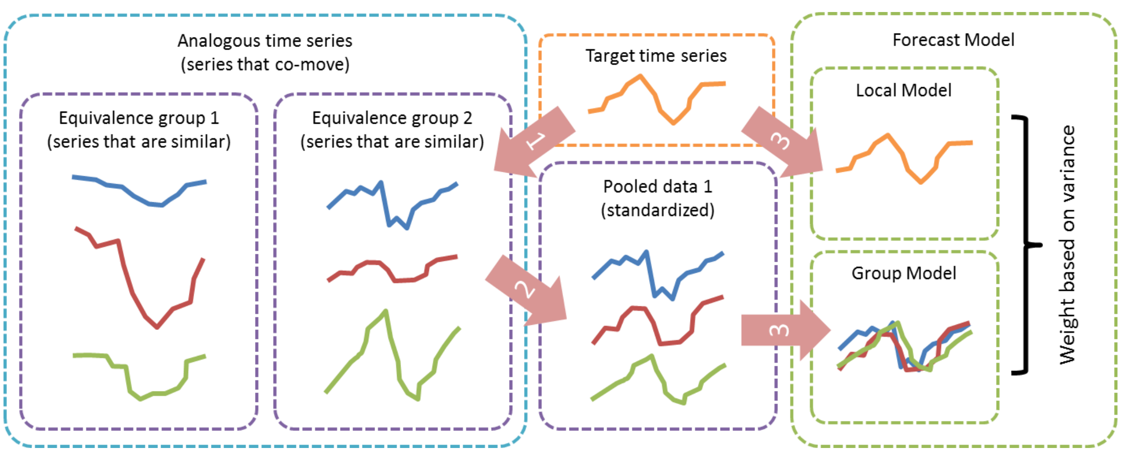

Figure 2.9: Overview of Bayesian pooling [2]

Examples of analogous time series include a set of time series that describe the sales of similar products in the same geographic area, or a single product sold in different geographic areas. In econometrics, Garcia-Ferrer [21] used output growth rate, real stock returns, and growth rate of real money supply from nine countries between 1954 and 1981 as analogous time series. In energy forecasting, temperature and energy consumption are analogous time series.

When a pattern change (Figure 2.10) occurs, the parameter estimates of the local model become imprecise. The parameter estimates of the group model, on the other hand, remain precise if the analogous time series in the equivalence group continue to co-move with the target time series. Using weights that are inversely proportional to the variance, the parameter estimates of the group model is given a larger weight. This, in turn, improves the precision of the combined model. Hence, Bayesian pooling is most useful when the target time series is highly volatile or is characterized by multiple distinct time-based patterns such as the ones shown in Figure 2.10 [2].

An example illustrates how to combine nonseasonal univariate local and group models. If we let

- i be the target series index (i.e. target series $i = 1$ might refer to output growth rate in the United States, while target series $i = 2$ might refer to output growth rate in England),

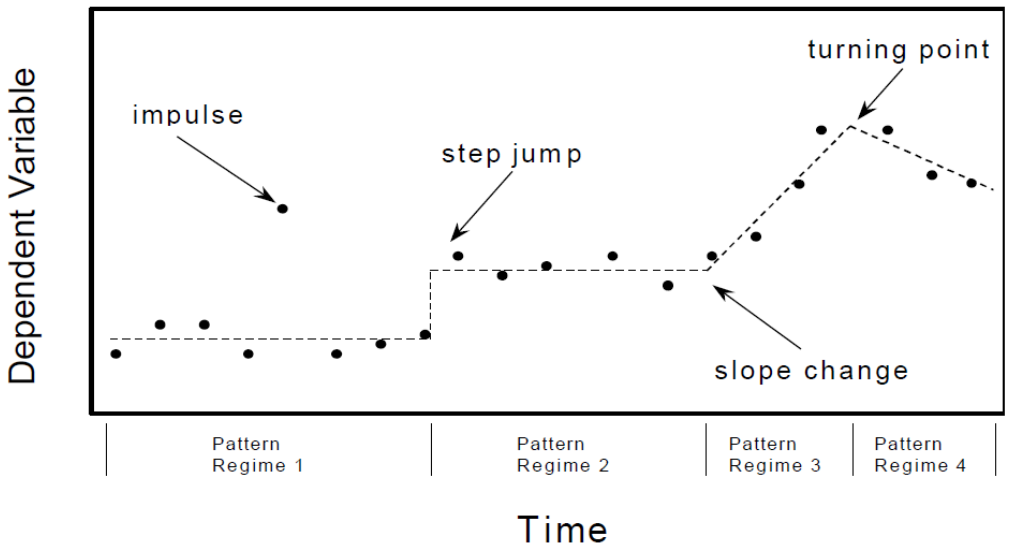

Figure 2.10: Illustration of nonstationary time series with four pattern regimes [2]

- k be a forecast origin (the most recent historical period for which data is used to build a forecasting model. For example, if the model was trained using data up to 1973, then $k = 1973$),
- $\widehat{l}_{i,k}$ be a combined final intercept term estimate for target series i at forecast origin k,
- $\widehat{s}_{i,k}$ be a combined final slope term estimate for target series i at forecast origin k,
- $l_{i,k}$ be an estimated intercept term from the local model for target series i at forecast origin k,
- $s_{i,k}$ be an estimated slope term from the local model for target series i at forecast origin k,
- $\bar{x_k}$ be a sample mean of group pooled data at forecast origin t,

- $\bar{\delta}_k$ be a sample mean of group first differences at forecast origin k,
- u_1 be a shrinkage weight that is inversely proportional to the estimated variance of l, the estimated intercept term from the local model,
- u_2 be a shrinkage weight that is inversely proportional to the estimated variance of $\bar{x}_k$,
- w_1 be a shrinkage weight that is inversely proportional to the estimated variance of s, the estimated slope term from the local model, and
- w_2 be a shrinkage weight that is inversely proportional to the estimated variance of $\bar{\delta}_k$.

Then the local and group model parameters can be combined as follows:

$$\widehat{l}_{i,k} = u_1 l_{i,k} + u_2 \bar{x}_k, \text{ and}$$

$$\widehat{s}_{i,k} = w_1 s_{i,k} + w_2 \bar{\delta}_k.$$

Using the intercept and slope calculated for the combined model, the n-step ahead forecast is:

$$\widehat{y}_{k+n} = \widehat{l}_{i,k} + n\widehat{s}_{i,k}.$$

The Bayesian pooling technique frequently is used in financial and economic forecasting since business cycles and processes often cause pattern changes in time series. Duncan et al. [16] applied a Bayesian pooling technique refereed to as

C-MSKF (Multi-State Kalman Filter with Conditionally Independent Hierarchical method). The researchers forecast income tax revenue for each of 40 school districts in Allegheny County, Pennsylvania, based on fifteen years of data. The study uses a Dynamic Linear Model to perform the actual forecast, but uses a Bayesian pooling technique to identify and group districts that behave similarly based on the sensitivity of their revenue collections to economic cycles.

The researchers chose a Bayesian pooling technique as their forecast method because existing methods did not incorporate all of the information that is available. This reasoning is very similar to the theoretical justifications for using Bayesian Networks: both methods try to incorporate information that previously was unaccounted for. The two methods differ in the sense that Bayesian Networks rely on the spatial and temporal relationships among observations. Bayesian pooling, on the other hand, relies on analogous time series that move together. These time series are not required to have a cause-and-effect or spatial-temporal relationship among them. Hence, Bayesian pooling is most useful when numerous time series are available with parallel observations that co-move (i.e. economic and business indicators) [16]. Aside from temperature, individual customer's heating oil demand does not strongly co-move with other time series. Additionally, it is unlikely to observe a clear change in heating oil demand patterns during the first few initial deliveries. Hence, Bayesian pooling techniques do not seem to help improve short-term heating oil forecasts on a per-customer basis. The next section discusses a technique called a Dynamic Linear Model, which is also a technique that accounts for nonstationary behaviors in time series.

2.3.3 Dynamic Linear Models

A Dynamic Linear Model (DLM) is a structure that is used to model time series with nonstationary components [41]. DLM is a sequential parametric model consisting of two equations that describe how the parameters change over time as a result of systematic effects and random shocks. The **observation equation** specifies the stochastic relationship between the independent and dependent variables using parameters at time t. The **system equation** describes how the parameters change stochastically over time [28].

A Dynamic Linear Model is a framework that one can use to model complex time series that are difficult to model otherwise. Complex time series are easier to model with this framework because the framework allows the forecaster to describe how parameters change over time. The framework also allows the forecaster to include explicitly both systematic and random effects that cause the parameters to change. By explicitly accounting for the random effects, the framework is able to express parameters using probability. In contrast, traditional forecasting techniques, such as Multiple Linear Regression, do not explicitly describe how parameters change over time or express parameters using probability. Due to these flexibilities, DLM can be applied to various forecasting problems that saw limited success with traditional methods. DLM is a framework in a sense that many traditional forecasting models can be expressed as special cases of DLM.

Let

- y_t be a $(m \times 1)$ vector of observations (dependent variable) at time t,
- θ_t is a $(n \times 1)$ vector of parameters at time t,
- F_t be a $(m \times n)$ matrix of independent variables at time t,
- G is a $(n \times n)$ system matrix that is known,
- v_t be a $(m \times 1)$ random normal vector with zero means and variances known at time t, and
- w_t be a $(n \times 1)$ random normal vector with zero means and variances known at time t.

Then a DLM consists of observation and system equations

$$y_t = F_t\theta_t + v_t, v_t \sim N(0, V_t), \text{ and} \tag{2.4}$$

$$\theta_t = G\theta_{t-1} + w_t, w_t \sim N(0, W_t), \tag{2.5}$$

where Equation 2.4 is the observation equation, and Equation 2.5 is the system equation. Of these components, G and w_t are the most significant. The matrix G defines how parameters change over time, and w_t is the component that adds randomness when parameters change over time, allowing the parameters to be expressed in terms of probability.

As it was mentioned earlier, traditional forecasting models can be expressed as special cases of DLM. For instance, a DLM with the following components express a simple regression model. Set $G = I$ and $W_t = 0$ so that the system equation is $\theta_t = \theta_{t-1} + w_t, w_t \sim N(0, 0)$. Since w_t has zero mean and zero variance, w_t is no longer a random variable. Instead, it is a constant with a value of zero. Hence, the system equation is $\theta_t = \theta_{t-1}$, which simply states that the parameters do not change over time.

A Dynamic Linear Model is frequently referred to as the Bayesian forecasting technique due to its use of the Kalman filter, which recursively computes the parameter distributions. Recall that Bayes' Theorem describes how we obtain posterior knowledge by updating our prior knowledge. Similarly, if the prior distribution of the parameters has a normal distribution with mean m_0 and variance C_0, then updating the prior distribution using past observation values (y^t and F^t) yields a posterior distribution of the parameters at time t. This posterior distribution is normally distributed with mean m_t and variance C_t. Hence, if we let

- θ_0 be a prior probability distribution,
- $(\theta_t|y^t, F^t)$ be a posterior probability distribution,
- y^t be a sequence of values from y_1 to y_t $(y_1...y_t)$, and
- F^t be a sequence of values from F_1 to F_t $(F_1...F_t)$,

then the prior and the posterior are expressed as

$$\theta_0 \sim N(m_0, C_0), \text{ and}$$

$$(\theta_t|y^t, F^t) \sim N(m_t, C_t).$$

The values of m_t and C_t can be obtained recursively as follows. If we let

$$\widehat{y} = F_t G m_{t-1},$$

$$e = y_t - \widehat{y},$$

$$R = G C_{t-1} G^T + W_t,$$

$$\widehat{Y} = F_t R F_t^T + V_t, \text{ and}$$

$$A = R F_t^T \widehat{Y}^{-1},$$

then

$$m_t = Gm_{t-1} + Ae, \text{ and}$$

$$C_t = R - A\widehat{Y}A^T.$$

A more through discussion of DLM is available in textbooks such as *Time Series: Modeling, Computation, and Inference* [41], or in journal articles such as *Bayesian Forecasting* [28].

One of the earliest applications of DLM can be seen by the work of Green and Harrison [24]. The researchers applied DLM to forecast the sales of ladies' dresses sold by a mail order company between August 1970 and December 1970. The motivation of the study is to use a Bayesian approach to forecast the sales of a new product in the absence of a sales history. Additionally, the prior estimates are obtained by expressing the experiences of the staff members in terms of possibilities and probabilities. Johnston and Harrison [30] applied DLM to forecast the demand of alcoholic beverages in the United Kingdom between 1977 and 1980 using historical sales data between 1970 and 1976. During the summer of 1976, an unusually hot and dry summer significantly increased the demand, while the introduction of Excise Duty at the end of 1976 depressed demand for alcoholic beverages. The authors applied DLM because traditional forecasting techniques, such as Linear Growth Seasonal Model, performed poorly given these unusual events. A DLM was applied to incorporate previously unaccounted information to the overall model. Pezzulli et al. [38] forecasted electricity peak demand daily trajectory during the winter season in Central England and Wales using DLM. The study used daily peak demand fluxes for winters between 1986 and 2003. The model uses three components: a calendar component that accounts for the day of the week and winter cycles; an economic component that accounts for industrial activities measured by the Service Sector Index; and a weather component that accounts for temperature, wind, and solar radiation.

Green and Harrison [24] demonstrate that the Bayesian approach to forecasting can be applied to cases where there is a lack of prior historical sales data. This is critical because this project aims to forecast heating oil demand for new customers without prior historical demand data. Additionally, this project must incorporate an estimate from the outside expert as a part of the forecast. The study also incorporated the forecasting process used by the expert (staff members) into the overall model. Johnston and Harrison [30] and Pezzulli et al. [38] both demonstrate that the technique can be applied successfully to forecast demand that is dependent on weather conditions. A successful application of the Bayesian methods in these studies suggest that Bayesian methods also can be applied to forecast heating oil demand.

This concludes the overview of the key concepts and existing forecasting techniques. Chapter 3 discusses the details of our Bayesian Heating Oil Forecaster and its evaluation method.

CHAPTER 3

Bayesian Heating Oil Forecaster

This chapter introduces the implementation of our Bayesian Heating Oil Forecaster by stepping through the estimation process using a hypothetical new customer. An overview of software architecture and implementation constraints are also covered. The major challenge for our Bayesian Heating Oil Forecaster is to improve the forecasts without knowing the historical behavior of the target customer being forecast. The next section provides a high-level overview of how our Bayesian Heating Oil Forecaster generates its estimate.

3.1 Thought Experiment: Forecasting Demand Without Historical Data

This section contains a thought experiment that steps through the logical process of generating a forecast without relying on the historical behavior of a target customer. The aim of this exercise is to illustrate how our Bayesian Heating Oil Forecaster is generating a forecast without using the historical data of the target customer.

Consider an example illustrated in Figure 3.1. Suppose there is an operator

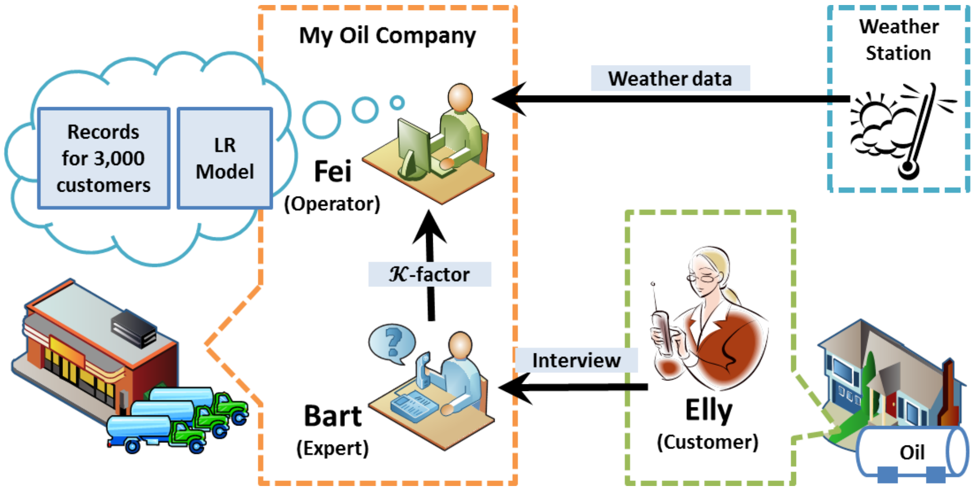

Figure 3.1: Experiment setup

named Fei who works for a heating oil distribution company, My Oil Company (MOC). The supervisor notifies Fei that Elly, a new customer, has just signed up for the delivery service offered by the company. The supervisor asks Fei to estimate the next delivery date for Elly. Fei has no information about Elly. We can assume that Fei has access to the company database, which contains historical delivery records and the latest delivery estimates for over 3,000 existing customers. Fei has access to past and future (10-day forecast) weather data from a near-by weather station. We can also assume that Fei knows the forecasting model that is used by the company, a simple regression model as shown in Equation 3.1 that relates temperature to the estimated demand ($\widehat{s_k}$) using a heatload factor (β_1) called the $\mathcal{K}$-factor ($\mathcal{K}$),

$$\widehat{s_k} = \beta_1 x_{1,k} = (1/\widehat{\mathcal{K}})\text{HDD}_{60,k}. \tag{3.1}$$

Since Fei can calculate the $\text{HDD}_{60,k}$ using data from the weather station, Fei can estimate Elly's demand if he can estimate the $\mathcal{K}$-factor and fit the model. Hence, the objective becomes estimating Elly's $\mathcal{K}$-factor. Given the contents of the database, Fei can estimate Elly's most likely $\mathcal{K}$-factor by calculating a range of $\mathcal{K}$-factors for a typical customer. Using the database, Fei can generate a customer $\mathcal{K}$-factor histogram such as the one shown in Figure 3.2. The histogram shows the number of existing customers with their latest $\mathcal{K}$-factor estimates. For example, the histogram indicates that most of the existing customers have $\mathcal{K}$-factors between 3 and 8, and over 70 existing customers have latest $\mathcal{K}$-factors of around 6. Having such a histogram enables Fei to identify what $\mathcal{K}$-factor values are the most common. Similarly, Fei can compute the average $\mathcal{K}$-factor for all existing customers and treat that as the $\mathcal{K}$-factor estimate for Elly. Hence, Fei has gained some knowledge of what the $\mathcal{K}$-factor for Elly might be without knowing Elly's historical behavior.

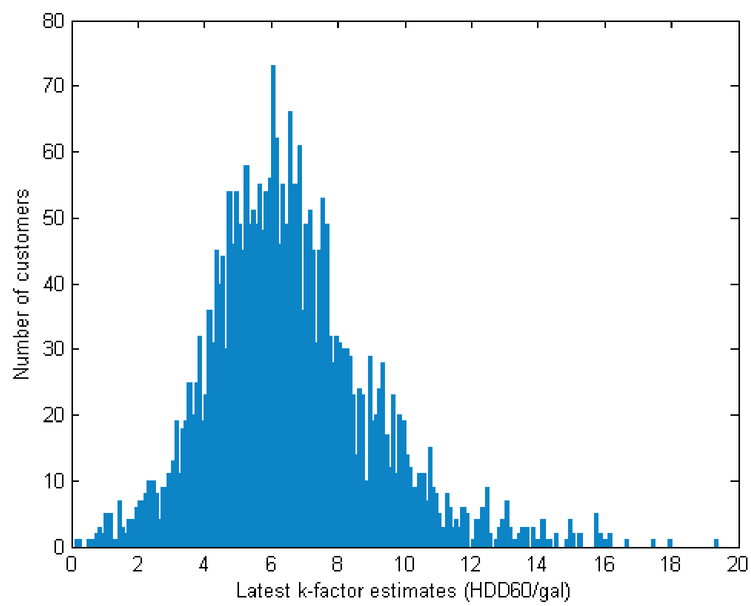

Figure 3.2: Histogram of the latest $\mathcal{K}$-factor estimates for the existing customers

Besides Fei, there is an expert, Bart, in the company. Bart contacts new customers, interviews them, and estimates their $\mathcal{K}$-factors. Bart has been an expert at the company for over 30 years, and his estimates usually are good. Additionally, the company database stores Bart's past estimates. Hence, when Bart estimates the $\mathcal{K}$-factor, Fei can look at the database to see how Bart has performed in the past. For example, the historical data demonstrates that when Bart estimated the $\mathcal{K}$-factor to be between 5.0 and 5.5, the actual $\mathcal{K}$-factor was also between 5.0 and 5.5 for 10 percent of the time. Similarly, for the same estimated $\mathcal{K}$-factor, the actual $\mathcal{K}$-factor was between 6.0 and 6.5 for 12 percent of the time. Repeating this process, Fei can create a probability table that looks like Figure 3.3. This table indicates the **likelihood** of the actual $\mathcal{K}$-factor when Bart provided a certain $\mathcal{K}$-factor estimate. Using the table shown in Figure 3.3, it can be seen that Bart tends to underestimate the $\mathcal{K}$-factor for the customers.

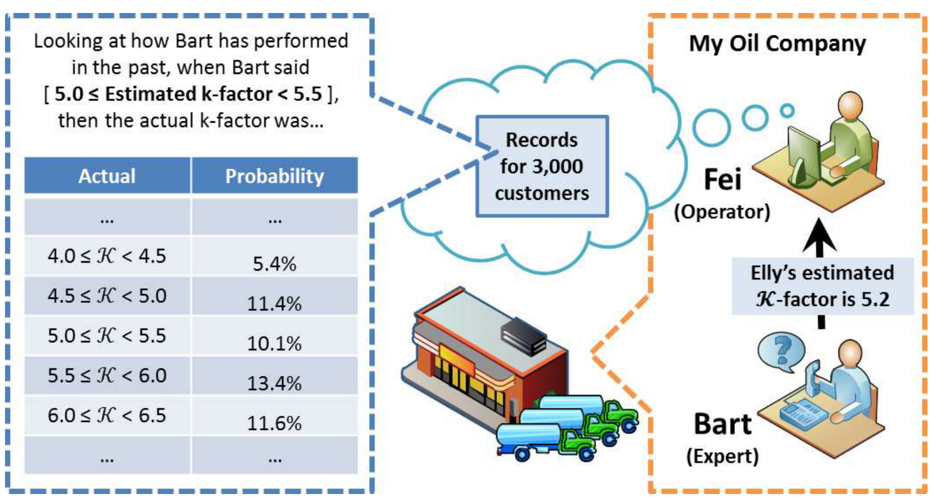

Figure 3.3: A simplified likelihood table

Before asking Bart for his estimate, Fei estimated that the $\mathcal{K}$-factor for Elly is the average $\mathcal{K}$-factor for all existing customers. Assume that the average was 6.2. Then Bart estimated Elly's $\mathcal{K}$-factor to be 5.2. Looking at Bart's past performance, Fei found that Bart is likely to underestimate. Based on the above, Fei concludes that the $\mathcal{K}$-factor for Elly is close to 6.2 but unlikely to be smaller than 5.2. Again, the main point is that Fei is able to forecast and refine the $\mathcal{K}$-factor estimate without requiring Elly's historical data.

In summary, Fei estimated Elly's $\mathcal{K}$-factor by first looking at the existing customer's latest $\mathcal{K}$-factor estimates. Fei used an average of the existing customer's $\mathcal{K}$-factors and used it as his estimate for Elly's $\mathcal{K}$-factor. Next, Fei obtained a $\mathcal{K}$-factor estimate from Bart, the company expert. Finally, Fei refined his $\mathcal{K}$-factor estimate for Elly without Elly's historical estimates by combining his estimate with Bart's estimate. Our Bayesian Heating Oil Forecaster follows a very similar process. Instead of using a single number (i.e. average of the existing customers), our Bayesian Heating Oil Forecaster uses the distribution of $\mathcal{K}$-factors to capture more information. The algorithm ultimately represents the belief about the customer's $\mathcal{K}$-factor using a probability distribution. The belief is updated according to Bayes' Theorem as outlined in Chapter 2.

3.2 Overview of the Bayesian Heating Oil Forecaster

Bayesian forecasting is an iterative process that revises the belief about an unobservable quantity. This process is depicted in a 4-step flowchart originally shown in Figure 2.2 (also reproduced below as Figure 3.4).

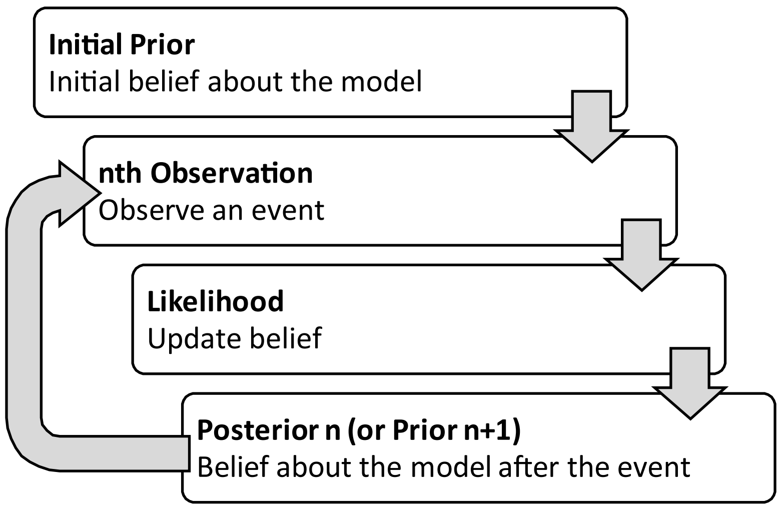

Figure 3.4: A model of how Bayes' Theorem updates forecaster knowledge

Our Bayesian Heating Oil Forecaster follows the steps shown in Figure 3.4 with a slight difference. Instead of observing a same kind of event and updating the belief based on the same likelihood, our Bayesian Heating Oil Forecaster observers two different kinds of events: the expert's $\mathcal{K}$-factor estimate and the estimate generated by the existing forecasting method. In the beginning when no delivery information is available, the algorithm relies on the subjective $\mathcal{K}$-factor estimate provided by the expert (Figure 3.5). In subsequent steps when delivery information becomes available, the algorithm relies on the $\mathcal{K}$-factor estimates provided by the existing forecasting method, as suggested by Figure 3.6.

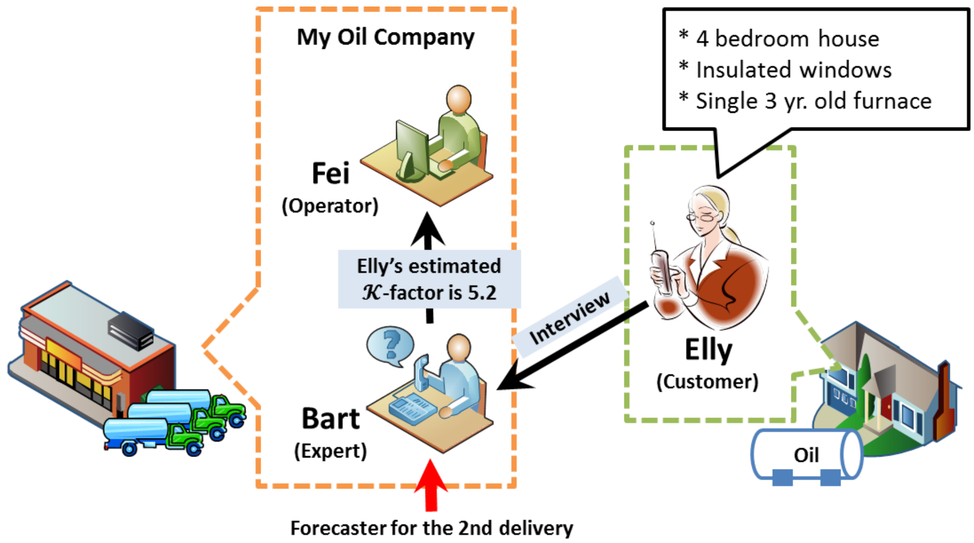

Figure 3.5: Observing the expert $\mathcal{K}$-factor estimate

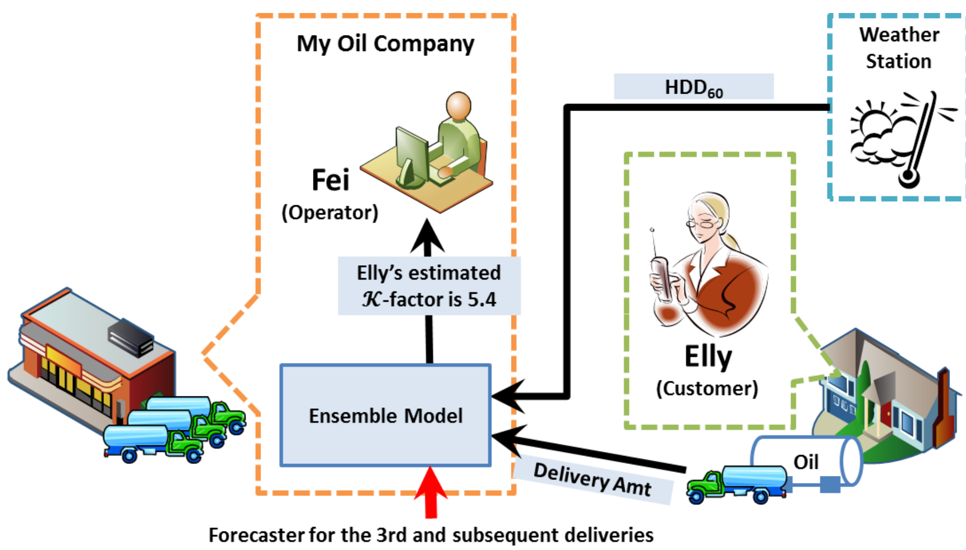

Figure 3.6: Observing the model $\mathcal{K}$-factor estimate

Hence, instead of updating the belief using a single likelihood, our Bayesian Heating Oil Forecaster relies on two different likelihoods. One likelihood is computed using the subjective $\mathcal{K}$-factor estimates provided by the expert. We call this likelihood the **expert likelihood**. Another likelihood is computed using the $\mathcal{K}$-factor estimates provided by the existing forecasting method. We call this likelihood the **model likelihood**. A summary of what our Bayesian Heating Oil Forecaster does under different conditions can be seen in Figure 3.7. Before the second delivery is made, our Bayesian Heating Oil Forecaster uses the $\mathcal{K}$-factor estimate provided by the expert to perform the forecast. After the second delivery is made, our Bayesian Heating Oil Forecaster uses the $\mathcal{K}$-factor estimate provided by the existing forecasting method.

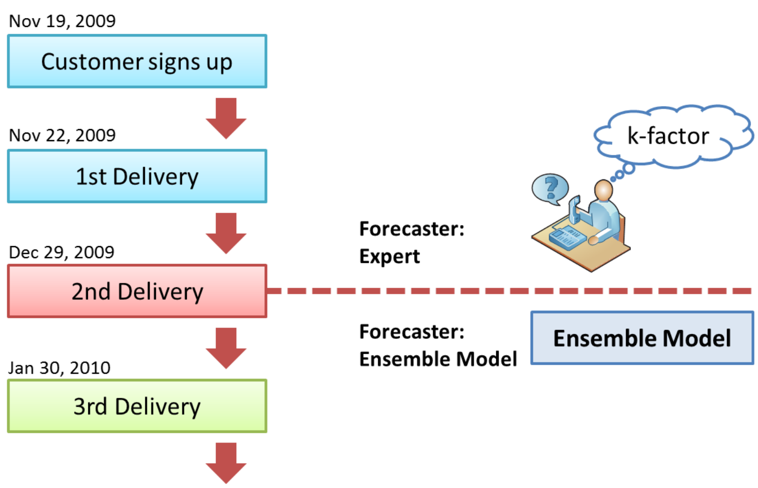

Figure 3.7: Event timeline and algorithm behavior

Next, the individual computation steps of our Bayesian Heating Oil Forecaster are examined in detail.

3.3 Computation Steps of our Bayesian Heating Oil Forecaster

Our Bayesian Heating Oil Forecaster consists of the following 10 steps:

Bayesian Heating Oil Forecaster Algorithm

1. Compute the initial belief about the $\mathcal{K}$-factor of the target customer;
2. Observe the $\mathcal{K}$-factor estimate from the expert;
3. Update the initial belief using the expert likelihood;
4. Obtain the posterior belief about the $\mathcal{K}$-factor of the target customer;
5. Observe the $\mathcal{K}$-factor estimate from the existing forecasting method;
6. Update the initial belief using the model likelihood;
7. Obtain the posterior belief about the $\mathcal{K}$-factor of the target customer;
8. Repeat steps 5 through 7 for every $\mathcal{K}$-factor estimate generated by the existing forecasting method;
9. Compute a $\mathcal{K}$-factor estimate from the posterior belief; and
10. Obtain the estimated heating oil demand by evaluating the regression model using the $\mathcal{K}$-factor estimate.

This section discusses the details of each of the computation steps, using actual training data (with scaled $\mathcal{K}$-factors) from a heating oil sales and distribution company that are collected between November 14, 2007, and September 30, 2009.

3.3.1 Step 1: Compute initial belief

Input: Existing customers' latest $\mathcal{K}$-factor estimates

Output: Beta distribution parameters for the first prior

The first step of the Bayesian Heating Oil Forecaster Algorithm is to compute the prior probability distribution of the $\mathcal{K}$-factor estimate. This probability distribution must be generated without knowledge about the target customer. We rely on the most recent $\mathcal{K}$-factor estimates of the existing customers.

First, we obtain a probability distribution estimate using the latest $\mathcal{K}$-factor estimates of the existing customers. Figure 3.8 shows a histogram of the latest $\mathcal{K}$-factor estimates collected during the training period for about 3,000 existing customers. The figure indicates that there are 54 existing customers whose latest $\mathcal{K}$-factor estimates are $[4.9, 5.0)$. Similarly, there are 58 existing customers whose latest $\mathcal{K}$-factor estimate fall between the range of $[5.2, 5.3)$. Figure 3.2 (also reproduced as Figure 3.9) is a histogram of the latest $\mathcal{K}$-factor estimates.

Our Bayesian Heating Oil Forecaster requires the probability distribution to be expressed as a beta distribution. The update process multiplies prior and likelihood distributions, and multiplying two empirical distributions is far more computationally intensive compared to multiplying two beta distributions. A maximum likelihood estimation technique is used to obtain a pair of beta

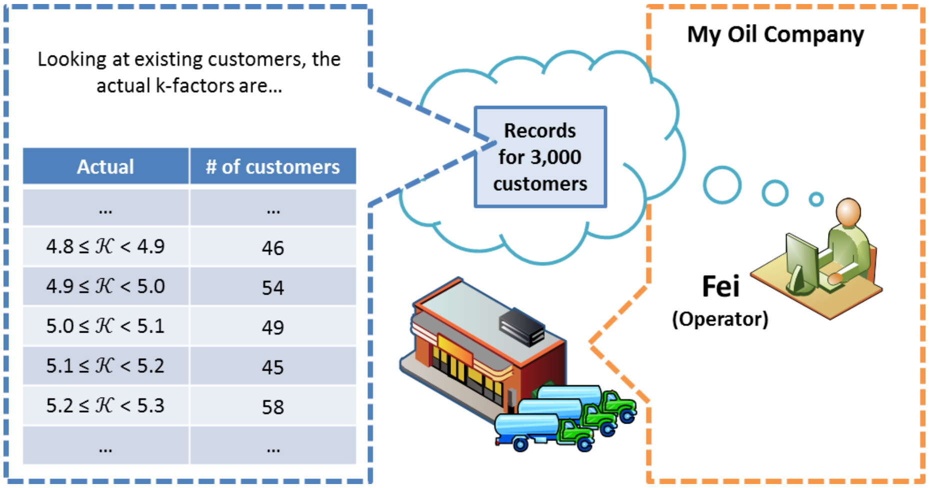

Figure 3.8: Frequency distribution of the latest $\mathcal{K}$-factor estimates for the existing customers

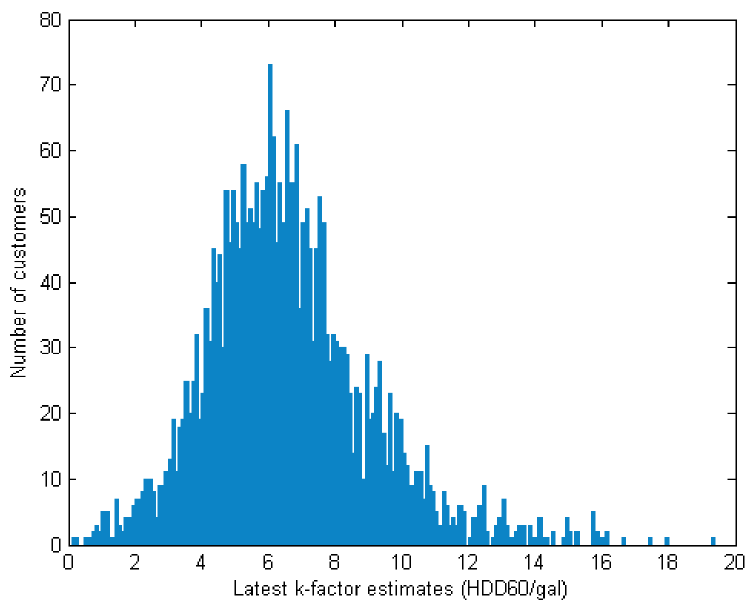

Figure 3.9: Histogram of the latest $\mathcal{K}$-factor estimates for the existing customers

parameters that best fit the empirical distribution. (Details of beta distribution and maximum likelihood estimation are discussed in Section 2.2.2.) Figure 3.10 shows a beta distribution that best fits the empirical distribution, which is

$$\text{Prior}(1) = \beta(\widehat{a}_1^{Pri}, \widehat{b}_1^{Pri}) = \beta(4.4245, 8.9541).$$

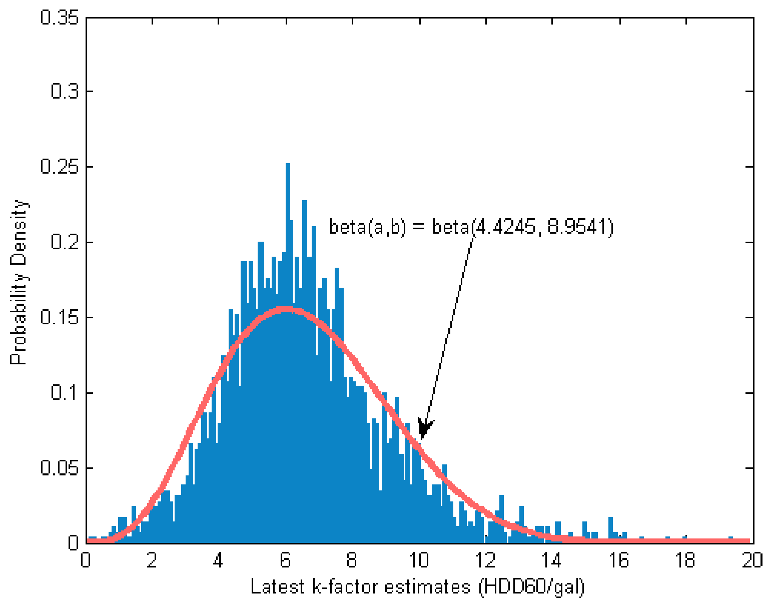

Figure 3.10: Empirical PDF (blue) and Beta PDF (red)

3.3.2 Step 2: Obtain $\mathcal{K}$-factor estimate from the expert

Output: Expert's $\mathcal{K}$-factor estimate for the target customer

The second step is to obtain a $\mathcal{K}$-factor estimate from the expert. The estimate can come from any source, including subjective estimates from a human expert in the field and a systematic process of estimating the initial $\mathcal{K}$-factor value. Experience-based estimation, as shown in Figure 3.5, is acceptable. The only requirement is that the historical record of the past estimates are available at the time of this forecast. Suppose the expert estimated the target customer's $\mathcal{K}$-factor to be 5.2.

3.3.3 Step 3: Update belief using the expert likelihood

Input: Expert's historical performance (Expert's estimate vs. Latest estimate)

Output: Beta distribution parameters for the expert likelihood

The third step involves constructing a joint probability distribution for the expert likelihood. A joint frequency distribution (Figure 3.11) and a joint probability distribution (Figure 3.12) summarizes the past performance of the forecaster by comparing the past $\mathcal{K}$-factor forecasts with the latest $\mathcal{K}$-factor. The figures only show a portion of the full distribution. The full distribution contains data from 175 customers whose $\mathcal{K}$-factor estimates range from 0 to 20. For example, the distribution shown in Figure 3.11 has a cell in the 8th row and 11th column (labeled 3.5 and 5.0 respectively) that contains a value of 4. The 11th column represents a case where the expert estimated the $\mathcal{K}$-factor to be in the range $[5.0, 5.5)$. The 8th row represents a case where the latest $\mathcal{K}$-factor estimates were in the range $[3.5, 4.0)$. Hence, row 8 column 11 represents a case where the expert overestimated the $\mathcal{K}$-factor by about 1.5. Since the cell contains the value 4, the expert overestimated the $\mathcal{K}$-factor by 1.5 for 4 out of 175 times. As time elapses and we accumulate additional customers, the number of initial $\mathcal{K}$-factors increases, which in turn increases the observations that are available in the distribution.

The distribution is constructed by going over the historical performance of the forecaster. For each existing customer, we look at the initial $\mathcal{K}$-factor estimate

Observed k-factor (Expert's k-factor estimate)

Latest k-factor (Estimated true k-factor)

	Col	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Row	k	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0
1	0.0															
2	0.5					1										
3	1.0															
4	1.5				1											
5	2.0								1			1				
6	2.5					1	1					2				
7	3.0							1				2				
8	3.5							2				4				
9	4.0									3	1	7				1
10	4.5											16	1			
11	5.0											13				
12	5.5									1		16				
13	6.0								1			15			1	
14	6.5							1				7			1	1
15	7.0											13				1
16	7.5									1		3				
17	8.0											5				

Unit: Number of observations

Figure 3.11: Portion of the Expert Likelihood Joint Frequency Distribution between November 14, 2007, and September 30, 2009 (175 observations)

Observed k-factor (Expert's k-factor estimate)

Latest k-factor (Estimated true k-factor)

	Col	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Row	k	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0
1	0.0															
2	0.5					0.57										
3	1.0															
4	1.5				0.57											
5	2.0								0.57			0.57				
6	2.5					0.57	0.57					1.14				
7	3.0							0.57				1.14				
8	3.5							1.14				2.29				
9	4.0									1.71	0.57	4.00				0.57
10	4.5											9.14	0.57			
11	5.0											7.43				
12	5.5									0.57		9.14				
13	6.0								0.57			8.57			0.57	
14	6.5							0.57				4.00			0.57	0.57
15	7.0											7.43				0.57
16	7.5									0.57		1.71				
17	8.0											2.86				

Unit: Percentage (shown without the % symbol to conserve space)

Figure 3.12: Portion of the Expert Likelihood Joint Probability Distribution between November 14, 2007, and September 30, 2009 (175 observations)

produced by the forecaster and the latest $\mathcal{K}$-factor estimate produced by the existing forecasting method. If the initial $\mathcal{K}$-factor estimated by the forecaster for an existing customer is 6.0, and the latest $\mathcal{K}$-factor estimated by the existing forecasting method is 4.0, then we add a value one to the cell in row 9 column 13. We repeat this process for all known existing customers until the joint frequency distribution is fully populated.

Our Bayesian Heating Oil Forecaster requires the likelihood to be expressed in terms of probability. If we divide each number in the joint frequency distribution with the sum of all entries in the joint frequency distribution, we obtain the joint probability distribution of the expert likelihood as shown in Figure 3.12. For example, if we divide row 8 column 11 by the total number of entries, then the resulting number ($4/175 \approx 2.29\%$) is a probability of the expert predicting the $\mathcal{K}$-factor to be in the range $[5.0, 5.5)$ and the latest $\mathcal{K}$-factor happens to be in the range $[3.5, 4.0)$. This probability, however, cannot be used as the likelihood. The desired likelihood expresses the probability of the expert overestimating or underestimating its forecast, given the initial estimate provided by the expert. This is analogous to selecting a single column in the joint distribution that matches the estimate provided by the expert. Hence, instead of dividing the number in row 8 column 11 by the sum of all entries in the joint distribution, we divide the number by the sum of all entries in column 11. The resulting number is the probability of the expert overestimating by 1.5 when the expert predicts a $\mathcal{K}$-factor between $[5.0, 5.5)$.

The likelihood is a probability distribution rather than a single probability.

Hence, the likelihood distribution is obtained by dividing each element in the column by the sum of all elements in the same column. The result is a marginal (as opposed to joint) probability distribution, which in our case is a discrete set of probabilities as seen in Figure 3.13 (The column is transposed for display purposes). Cells before the 11th element are the probabilities that the forecaster is overestimating the $\mathcal{K}$-factor. The 11th cell is the probability that the forecaster is correctly estimating the $\mathcal{K}$-factor. Cells after the 11th element are the probabilities that the forecaster is underestimating the $\mathcal{K}$-factor.

Row	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5
Freqency	0	0	0	0	1	2	2	4	7	16	13	16	15	7	13	3	5	3	3	4
Probability (%)	0.0	0.0	0.0	0.0	0.8	1.6	1.6	3.1	5.4	12.4	10.1	12.4	11.6	5.4	10.1	2.3	3.9	2.3	2.3	3.1

Row	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	
	10.0	10.5	11.0	11.5	12.0	12.5	13.0	13.5	14.0	14.5	15.0	15.5	16.0	16.5	17.0	17.5	18.0	18.5	19.0	19.5	Sum
Freqency	3	3	1	0	4	0	0	1	1	0	0	1	0	1	0	0	0	0	0	0	129
Probability (%)	2.3	2.3	0.8	0.0	3.1	0.0	0.0	0.8	0.8	0.0	0.0	0.8	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	100

Figure 3.13: Column 11 marginal distribution from the expert likelihood joint distribution

What we have so far is an empirical marginal probability distribution for the likelihood. For the ease of calculation, it is desirable that the probability distribution is expressed as a beta distribution. We can estimate the beta parameters that best fit our empirical marginal probability distribution using a maximum likelihood estimation technique. Figure 3.14 is the histogram representation of Figure 3.13. Figure 3.15 shows a beta distribution that best fits the empirical distribution, which is Likelihood(1) $= \beta(\widehat{a}_1^L, \widehat{b}_1^L) = \beta(4.5656, 8.8003)$.

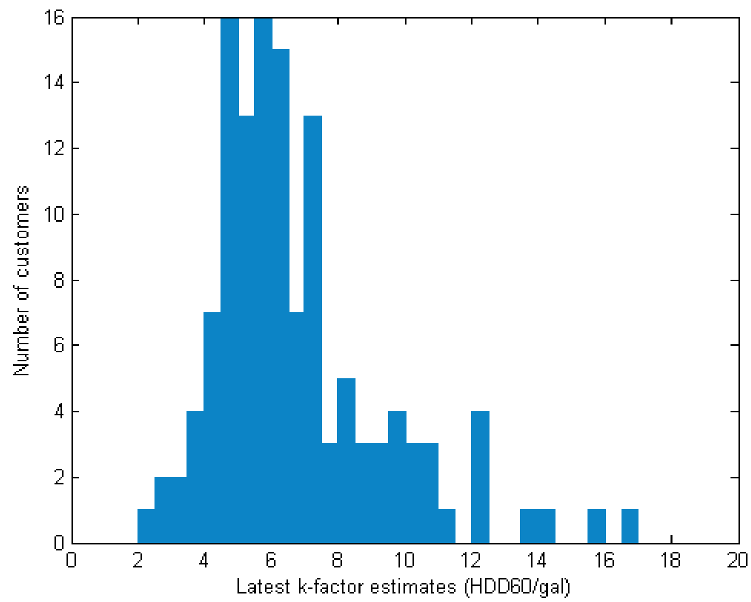

Figure 3.14: Histogram of the Marginal Frequency Distribution for the Expert Likelihood

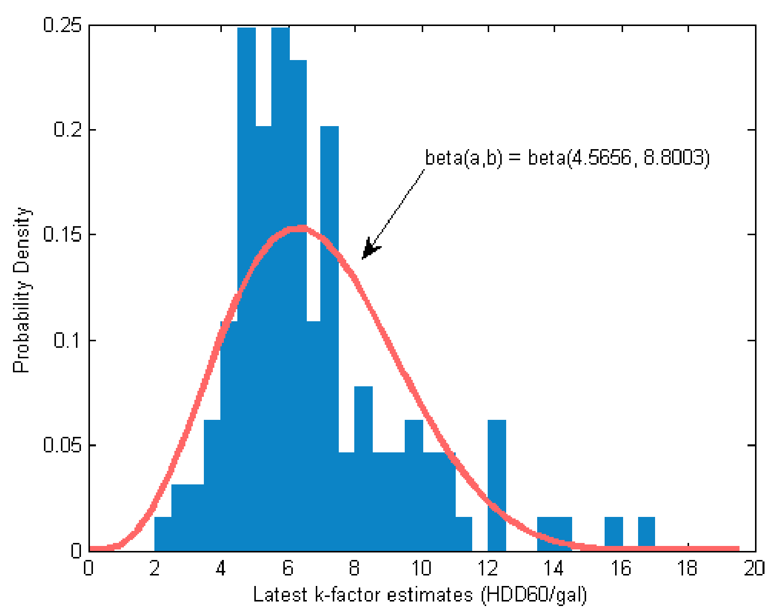

Figure 3.15: Empirical PDF (blue) and Beta PDF (red) for the Expert Likelihood

In summary, Step 3 consists of the following tasks:

1. Construct the joint distributions for the expert likelihood using past estimates.

2. Select the column that matches the initial estimate provided by the forecaster.

3. Perform a maximum likelihood estimation to obtain parameters for the beta distribution that best fits the empirical probability distribution.

Step 4 uses the beta parameters obtained in Step 3.

3.3.4 Step 4: Obtain posterior belief

Input: Prior(1) and Likelihood(1) beta distribution parameters

Output: Beta distribution parameters for the Posterior(1) distribution

Bayes' Theorem dictates that the posterior belief is proportional to the product of the prior belief and the likelihood. Since both the prior and the likelihood are expressed as beta distributions, the posterior can be computed as follows:

$$
\begin{aligned}
\text{Posterior} &= \text{Prior} \propto \text{Likelihood} \\
\text{Posterior}(n) &= \beta(\widehat{a}_n^{Post}, \widehat{b}_n^{Post}) \\
&= \beta(\widehat{a}_n^{Pri}, \widehat{b}_n^{Pri}) \times \beta(\widehat{a}_n^{L}, \widehat{b}_n^{L}) \\
&= k_n^{Post} x^{(\widehat{a}_n^{Pri}+\widehat{a}_n^{L}-2)}(1-x)^{(\widehat{b}_n^{Pri}+\widehat{b}_n^{L}-2)} \\
&= \beta(\widehat{a}_n^{Pri}+\widehat{a}_n^{L}-2, \widehat{b}_n^{Pri}+\widehat{b}_n^{L}-2) \qquad (3.2) \\
\text{Posterior}(1) &= \beta(\widehat{a}_1^{Post}, \widehat{b}_1^{Post}) \\
&= \beta(\widehat{a}_1^{Pri}, \widehat{b}_1^{Pri}) \times \beta(\widehat{a}_1^{L}, \widehat{b}_1^{L}) \\
&= \beta(\widehat{a}_1^{Pri}+\widehat{a}_1^{L}-2, \widehat{b}_1^{Pri}+\widehat{b}_1^{L}-2) \\
&= \beta(4.4245+4.5656-2, 8.9541+8.8003-2) \\
&= \beta(6.9901, 15.7544). \qquad (3.3)
\end{aligned}
$$

The end result (Equation 3.3) is a beta distribution with a new set of parameters, 6.9901 and 15.7544. The parameters for the posterior beta distribution is a sum of parameters from the prior and the likelihood distributions. This drastically simplifies the computation compared to using other distributions such as empirical distributions. Due to the iterative nature of the Bayesian forecasting technique, the posterior belief of the first iteration becomes the prior belief of the second iteration.

3.3.5 Step 5: Obtain $\mathcal{K}$-factor estimate from the model

Output: $\mathcal{K}$-factor estimate produced by the existing forecasting method for the target customer

Once the initial $\mathcal{K}$-factor estimate is provided by the expert in Step 2, subsequent estimates are provided by the existing forecasting method. Similar to Step 2, the historical record of the past estimates must be available at the time of this forecast.

As mentioned earlier in section 1.4, the existing forecasting method is an ensemble forecast model whose components include Linear Regression model, expert $\mathcal{K}$-factor, and tank size and $\mathcal{K}$-factor model. The ensemble model accepts weather data, delivery amount, and historical delivery record as inputs to produce its own $\mathcal{K}$-factor estimate for the target customer's latest delivery. Our Bayesian Heating Oil Forecaster accepts its latest $\mathcal{K}$-factor estimate as an observation for the second and subsequent deliveries.

3.3.6 Step 6: Update belief using the model likelihood

Input: Model's historical performance (Model's estimate vs. Latest estimate)

Output: Beta distribution parameters for the model likelihood

The overall process is exactly the same as in Step 3 when the belief was updated using the expert likelihood. Since the joint distribution for the expert likelihood is different from the model likelihood, we must generate a new joint distribution. Figure 3.16 is the joint frequency distribution, and Figure 3.17 is the joint probability distribution for the model likelihood. Instead of the estimates produced by the expert, the columns now refer to the estimates produced by the existing forecasting method. For each past forecast, we look at the past $\mathcal{K}$-factor estimate and the latest $\mathcal{K}$-factor estimate produced by the existing forecasting method. For instance, if the existing forecasting method estimated a $\mathcal{K}$-factor for an existing customer to be 6.0, and the latest $\mathcal{K}$-factor estimated for the same customer is 4.0, then we add a value one to row 9 column 13 of the joint frequency distribution. We repeat this process for all of the past forecasts until the joint frequency distribution is fully populated.

Observed k-factor (Model's k-factor estimate)

Latest k-factor (Estimated true k-factor)

	Col	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Row	k	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0
1	0.0	67														
2	0.5		236	15												
3	1.0		80	288	45											
4	1.5			42	293	30	1					1				
5	2.0	1			89	486	94	6	6			1				
6	2.5				5	101	396	81	9	9		1		1		
7	3.0			1	1	16	125	851	139	24	10	8	1			
8	3.5	1			1	7	39	231	969	244	43	4	3	5	3	2
9	4.0			1			14	31	263	1204	339	44	6	3	1	
10	4.5				1		2	11	57	414	1334	352	41	8	8	2
11	5.0							8	42	72	466	1217	382	65	35	8
12	5.5						1	1	11	26	102	490	1226	298	49	27
13	6.0					5	2	1	18	14	45	123	467	1021	377	94
14	6.5		1			1	1	3	3	5	17	47	132	451	936	254
15	7.0									1	10	17	34	116	394	604
16	7.5					1		1	8	1	1	5	29	35	110	272
17	8.0										1	14	22	21	29	74

Unit: Number of observations

Figure 3.16: Portion of the Model Likelihood Joint Frequency Distribution between November 14, 2007, and September 30, 2009 (24,121 observations)

Observed k-factor (Model's k-factor estimate)

Latest k-factor (Estimated true k-factor)

	Col	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Row	k	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0
1	0.0	0.28														
2	0.5		0.98	0.06												
3	1.0		0.33	1.19	0.19											
4	1.5			0.17	1.21	0.12	0.00					0.00				
5	2.0	0.00			0.37	2.01	0.39	0.02	0.02			0.00				
6	2.5				0.02	0.42	1.64	0.34	0.04	0.04		0.00		0.00		
7	3.0			0.00	0.00	0.07	0.52	3.53	0.58	0.10	0.04	0.03	0.00			
8	3.5	0.00			0.00	0.03	0.16	0.96	4.02	1.01	0.18	0.02	0.01	0.02	0.01	0.01
9	4.0			0.00			0.06	0.13	1.09	4.99	1.41	0.18	0.02	0.01	0.00	
10	4.5				0.00		0.01	0.05	0.24	1.72	5.53	1.46	0.17	0.03	0.03	0.01
11	5.0							0.03	0.17	0.30	1.93	5.05	1.58	0.27	0.15	0.03
12	5.5						0.00	0.00	0.05	0.11	0.42	2.03	5.08	1.24	0.20	0.11
13	6.0					0.02	0.01	0.00	0.07	0.06	0.19	0.51	1.94	4.23	1.56	0.39
14	6.5		0.00			0.00	0.00	0.01	0.01	0.02	0.07	0.19	0.55	1.87	3.88	1.05
15	7.0									0.00	0.04	0.07	0.14	0.48	1.63	2.50
16	7.5					0.00		0.00	0.03	0.00	0.00	0.02	0.12	0.15	0.46	1.13
17	8.0										0.00	0.06	0.09	0.09	0.12	0.31

Unit: Percentage (shown without the % symbol to conserve space)

Figure 3.17: Portion of the Model Likelihood Joint Probability Distribution between November 14, 2007, and September 30, 2009 (24,121 observations)

Once the table is populated, we select a single column that matches the current forecast produced by the existing forecasting method. Since the existing forecasting method estimated the $\mathcal{K}$-factor for the target customer to be 6.8, we select column 14 in the joint frequency distribution for the model likelihood. Following the same procedure as Step 3, we estimate the beta parameters that best fit the column. Figure 3.18 is the histogram representation of column 14 in the joint frequency distribution for the model likelihood. Figure 3.15 shows a beta distribution that best fits the empirical distribution, which is

$\text{Likelihood}(2) = \beta(\widehat{a}_2^L, \widehat{b}_2^L) = \beta(43.01, 82.11)$.

Step 7 uses the beta parameters obtained in Step 6 to compute the posterior belief.

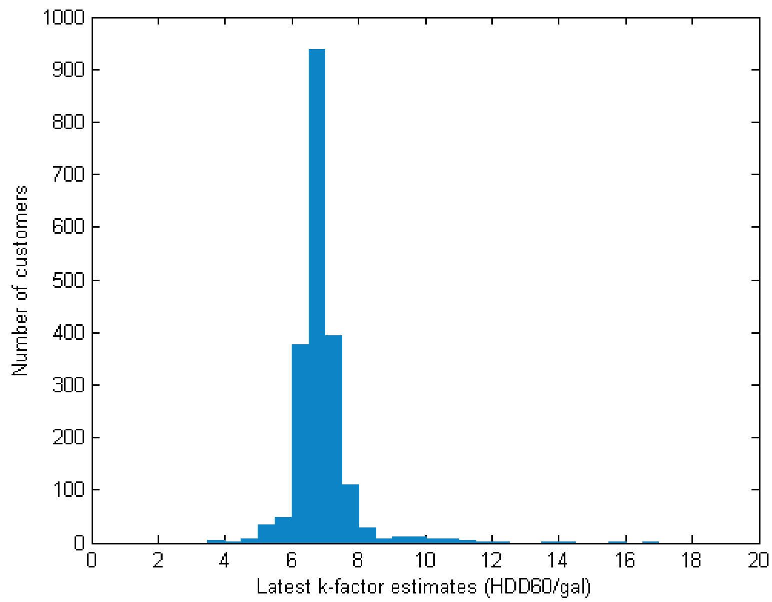

Figure 3.18: Histogram of the Marginal Frequency Distribution for the Model Likelihood

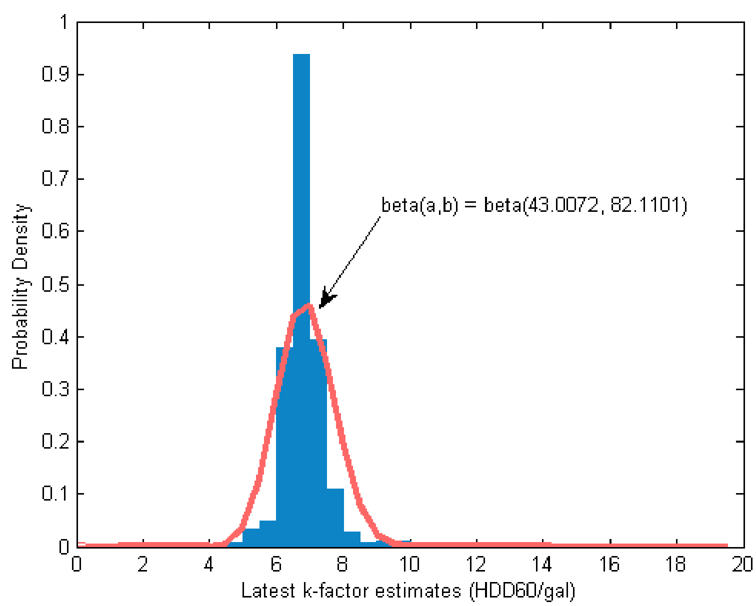

Figure 3.19: Empirical PDF (blue) and Beta PDF (red) for the Model Likelihood

3.3.7 Step 7: Obtain posterior belief

Input: Prior(2) and Likelihood(2) beta distribution parameters

Output: Beta distribution parameters for the Posterior(2) distribution

This step is identical to Step 4. Using Equation 3.2, we compute the posterior probability distribution using the prior distribution from Step 4 and the likelihood distribution from Step 6. Equation 3.4 is the new posterior belief.

$$\begin{aligned}
\text{Posterior}(2) &= \beta(\widehat{a}_2^{Post}, \widehat{b}_2^{Post}) \\
&= \beta(\widehat{a}_2^{Pri} + \widehat{a}_2^{L} - 2, \widehat{b}_2^{Pri} + \widehat{b}_2^{L} - 2) \\
&= \beta(6.9901 + 43.01 - 2, 15.7544 + 82.11 - 2) \\
&= \beta(48.00, 95.86) \qquad (3.4)
\end{aligned}$$

3.3.8 Step 8: Repeat steps 5 through 7

We repeat Steps 5 through 7 for each of the previous $\mathcal{K}$-factor estimates for the target customer. If the target customer has had 4 deliveries so far, then Steps 1 through 4 are performed once, and Steps 5 through 7 are repeated 3 times. The posterior belief is used to obtain the $\mathcal{K}$-factor estimate for the 5th delivery.

3.3.9 Step 9: Obtain $\mathcal{K}$-factor estimate

Input: Beta distribution parameters for the final posterior distribution

Output: Estimated $\mathcal{K}$-factor for the target customer

The posterior belief is a beta probability distribution, while a $\mathcal{K}$-factor is a single continuous non-negative number. Hence, the posterior belief itself cannot represent a single $\mathcal{K}$-factor estimate. Instead, we can compute an expected $\mathcal{K}$-factor from the posterior probability distribution. An expected $\mathcal{K}$-factor is the $\mathcal{K}$-factor estimate that is most likely to occur, given the posterior probability distribution. For a beta distribution, the expected value can be computed using Equation 2.3.

For example, if the posterior is $\beta(48.00, 95.86)$, then the expected $\mathcal{K}$-factor estimate is

$$\mathrm{E}(X) = \frac{48.00}{48.00 + 95.86} = 0.3337. \tag{3.5}$$

The result is scaled by the maximum $\mathcal{K}$-factor value of 20. Hence, the actual $\mathcal{K}$-factor estimate is $0.3337 \times 20 = 6.67$.

3.3.10 Step 10: Obtain estimated heating oil demand

Input: Estimated $\mathcal{K}$-factor for the target customer

Output: Estimated heating oil demand for the target customer

So far, we have estimated the $\mathcal{K}$-factor for the target customer. To estimate the heating oil demand for the target customer, we need to evaluate the regression model (Equation 1.1) using the estimated $\mathcal{K}$-factor. If the estimated $\mathcal{K}$-factor is 6.67, the estimated baseload is 0.15, and the heating degree day is 35, then the estimated demand for the target customer is $0.15 + 1/6.67 \times 35 \approx 5.40$ gallons.

The above calculation is for a daily estimate. If the estimated *cumulative* heating oil demand for the target customer becomes greater than 70% of the total tank capacity, then the target customer is flagged as a customer that requires heating oil delivery. Once the demand is computed, the steps above are repeated for each of the new customers.

This concludes the explanation of our Bayesian Heating Oil Forecaster Algorithm. The next section addresses the software implementation of the algorithm.

3.4 Software Implementation

There is a slight difference between the theoretical and the actual implementation of our Bayesian Heating Oil Forecaster. The theoretical implementation of our Bayesian Heating Oil Forecaster Algorithm computes the prior and the joint distribution for the two likelihoods during the estimation process. The prior probability distribution is generated for each customer. The likelihoods are generated for each repeated iteration. This is acceptable if the computation is instantaneous. However, repeating the calculation for every customer and for every iteration is extremely time consuming (Figure 3.20). To reduce the computation time, the actual implementation is modified in the following manner:

1. The prior probability distribution is computed once at the very beginning. This is possible because the prior probability does not change between target customers. The same prior probability distribution is reused for all of the target customers.

2. Instead of generating the likelihood tables for every iteration, the tables are computed once at the very beginning for ex-ante forecasts and for each target customer for ex-post forecasts

Figure 3.21 is a diagram of the actual implementation. Code 1 is a code implementation of Figure 3.21, where prior and two likelihoods are calculated before looping through each customer.

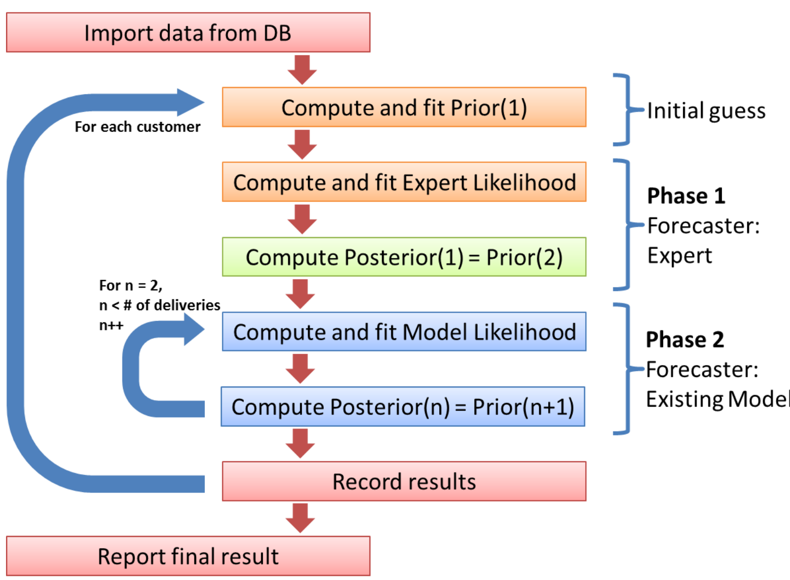

Figure 3.20: Theoretical implementation of our Bayesian Heating Oil Forecaster

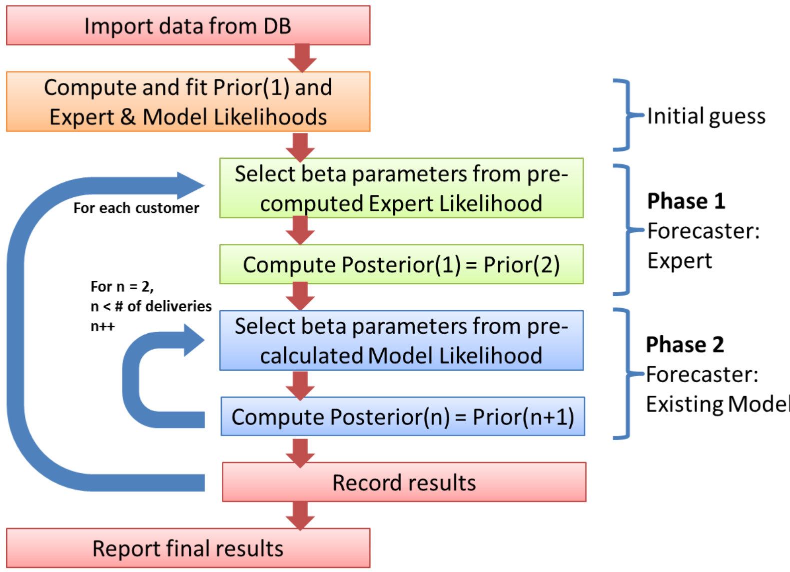

Figure 3.21: Actual implementation of our Bayesian Heating Oil Forecaster

Code 1 is the top-level pseudo-code of our Bayesian Heating Oil Forecaster. As seen in Figure 3.21, the section before the for loop pre-calculates the prior and the two likelihoods. The for loop repeats the calculation for each customer. Inside the for loop, the existing model performs its forecast first. The estimated $\mathcal{K}$-factor is then combined with previous $\mathcal{K}$-factor estimates and is passed to our Bayesian algorithm. The evaluation uses the baseload computed by the existing model and the $\mathcal{K}$-factor computed by the Bayesian model.

Code 2 shows the contents of the getBayesEstimate function. This function is responsible for generating the Bayesian $\mathcal{K}$-factor estimate for each customer. It implements our Bayesian Heating Oil Forecaster Algorithm with modifications to use pre-calculated prior and likelihoods. The section before the for loop corresponds to Steps 1 through 4 in our algorithm. The body of the for loop corresponds to

Steps 5 through 7, and the for loop itself corresponds to Step 8. The remainder of the code implements Steps 9 and 10. The function selectLikelihoodBeta is selecting a pre-calculated beta parameters that corresponds to the observed $\mathcal{K}$-factor. The parameter values are stored in the expertLikelihood and modelLikelihood variables.

This concludes the explanation of our Bayesian Heating Oil Forecaster and its implementation. The next chapter examines the evaluation method and evaluation results of our Bayesian Heating Oil Forecaster. The backtesting process and the evaluation method itself are the subject of the next chapter.

Code 1 Pseudo-code for the top-level Bayesian Heating Oil Forecaster Algorithm

```
prior1 = getPriorBeta(listOfLatestKFactors, maxKFactor);

expertLikelihood = getLikelihoodBeta(
    listOfExpertAndLatestKFactors, likelihoodBin, maxKFactor
);

modelLikelihood = getLikelihoodBeta(
    listOfModelAndLatestKFactors, likelihoodBin, maxKFactor
);

for i = 1:numOfCustomers

    % Run existing model
    [baseload kFactor] = getExistingModelEstimate(
        listOfCustomers(i),
        listOfDeliveries,
        hddPast
    );

    observedKFactor =
        [listOfPastExistingModelKFactorEstimates kFactor];

    % Run Bayes model
    kFactorBayes = getBayesEstimate(
        observedKFactor, prior1, expertLikelihood,
        modelLikelihood, likelihoodBin, maxKFactor
    );

    % Evaluate model
    listOfCustomers(i).estimatedDemand(today) =
        baseload + kFactorBayes * hddToday;

end
```

Code 2 Pseudo-code for getBayesEstimate

```
function estimatedKFactor = getBayesEstimate(
    observedKFactor, prior1, expertLikelihood,
    modelLikelihood, likelihoodBin, maxKFactor
)

    prior(1) = prior1;

    likelihood(2) = selectLikelihoodBeta(
        observedKFactor(1), expertLikelihood,
        likelihoodBin, maxKFactor
    );

    prior(2) = getPosteriorBeta(prior(1), likelihood(1));

    for i = 2:numberOfForecasts

        likelihood(i) = selectLikelihoodBeta(
            observedKFactor(i), modelLikelihood,
            likelihoodBin, maxKFactor
        );

        % Note: I don't have an array for Posterior(i) because
        %       Posterior(i) = Prior(i+1)
        prior(i+1) = getPosteriorBeta(prior(i), likelihood(i));

    end

    % Translate posterior to k-factor
    estimatedKFactor =
        getBetaExpectedValue(prior(numberOfForecasts+1)) * maxKFactor;

end
```

CHAPTER 4

Bayesian Heating Oil Forecaster Test Results

Chapter 3 discussed the design and implementation of our Bayesian Heating Oil Forecaster. This chapter discusses the details of the evaluation method that is used to measure the effectiveness of our Bayesian Heating Oil Forecaster. Additionally, this chapter reports the evaluation results. Our Bayesian Heating Oil Forecaster should reduce the forecasting error of the heating oil demand during the initial deliveries as defined in Section 1.2. The next section discusses the testing method used to evaluate the performance of our Bayesian Heating Oil Forecaster.

4.1 Evaluation Method

This section describes the evaluation method that is used to determine the effectiveness of our Bayesian Heating Oil Forecaster compared to the existing forecasting method. Section 4.1.1 on the backtesting process explains how the evaluation method using ex-post forecast is implemented, while the evaluation criteria lists attributes and performance measures that are used to compare our Bayesian Heating Oil Forecaster with the existing forecasting method.

4.1.1 Backtesting Process

Backtesting is a software implementation of the ex-post forecast that measures algorithm performance. The backtesting process uses historical delivery records and weather data to simulate how the algorithm would have performed in the past, and compares past forecasts with actual demand. The historical data used by the backtesting process is corrected using domain insights to reduce or remove the undesired effects from special cases, such as the overfill and underfill conditions.

Ex-ante: Today is October 1st, 2011

Year	2010		2011											
Month	Nov	Dec	Jan	...	Sep	Oct 1	Oct 2	Oct 3	Oct 4	Oct 5	Oct 6	Oct 7	Oct 8	Oct 9
Customer A						F1	F2	F3	F4	F5	F6	F7	F8	F9
Customer B						F1	F2	F3	F4	F5	F6	F7	F8	F9
Customer C						F1	F2	F3	F4	F5	F6	F7	F8	F9
	Past					Future								

= Training set is the same for all customers

Figure 4.1: Ex-ante Forecast Training Set

Ex-post: Re-estimate past deliveries

Year	Deliv	2008		2009						2010				
Month		Nov	Dec	Jan	...	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
Customer D	2							1	2		3		4	
Customer D	3							1	2		3		4	
Customer D	4							1	2		3		4	
Customer E	2								1	2		3		
Customer E	3								1	2		3		
		Training Set					Test Set							

= Training set expands to include days upto the next delivery being forecast

Figure 4.2: Ex-post Forecast Training Set

The behavior of the backtesting process is different from the ex-ante forecast covered in Chapter 3. Ex-ante forecast produces daily demand estimates in the *future*. The ex-post forecast covered in this chapter estimates the demand in the *past* and compares it against the actual demand for evaluation purposes. With an ex-ante forecast, we only need to calculate the prior and the two likelihoods once. Ex-post forecast, on the other hand, requires recalculation for each delivery for each customer. This is illustrated by Figures 4.1 and 4.2. Since the training period does not stay the same during the ex-post forecast, the implementation must account for this by introducing an additional loop in the algorithm, as shown in Figure 4.3.

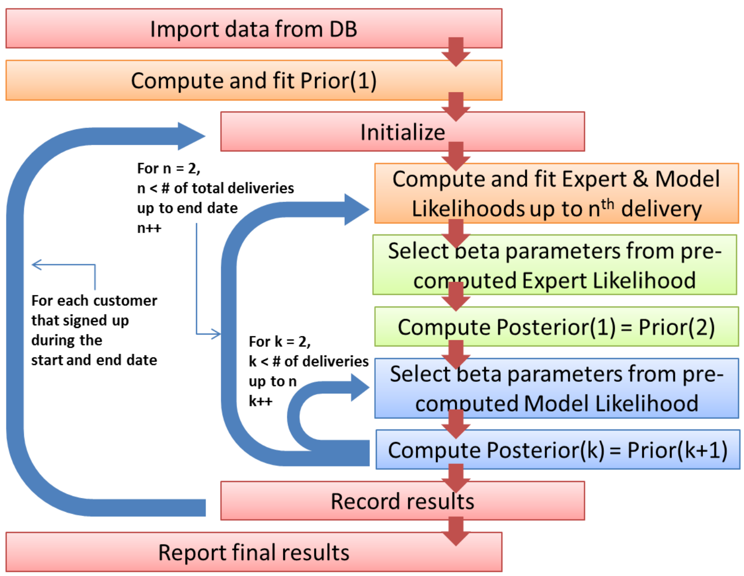

Figure 4.3: Ex-post implementation of our Bayesian Heating Oil Forecaster

The following example depicts the operation of the backtesting process. Suppose, we wish to test the algorithm by performing an ex-post forecast using deliveries that occurred between 11/1/2009 and 5/31/2010 as our test set. Historical data between 11/1/2008 and 10/31/2009 is used as our training set, as shown in Figure 4.2. To test the algorithm, the backtesting process first selects customers who signed up during the test period. In Figure 4.2, customers D and E are identified as new customers who signed up for the service during the test period. Suppose the second delivery for Customer D occurs on 11/15/2009. The backtesting process trains the algorithm using historical data between the training set start date (11/1/2008) and the day before the second delivery (11/14/2009). The backtesting process compares the demand forecast for the 15$^{\text{th}}$ with the actual demand that was observed on the 15$^{\text{th}}$. It continues the process by evaluating the third delivery on 12/18/2009 for customer D. The algorithm is trained again using historical data between the training set start date and the day before the third delivery (12/17/2009). The forecast is compared against the actual delivery that occurred on 12/18/2009. This process is repeated for every delivery and for each new customer. By selectively picking the historical data, the process simulates a forecast estimated on a particular day in the past.

Unfortunately, introducing the likelihood recalculation dramatically increases the processing time required to complete the backtesting process. Hence, the results reported in this chapter is based on a modified backtesting process, where the weather and historical delivery records are supplied for each delivery, but likelihoods are calculated only once at the beginning of the training set. When the likelihoods are not recalculated for each delivery, the Bayesian forecasting process is likely to perform worse since the likelihoods are not using all of the available historical data at the time of the forecast.

The difference between the estimate and the actual delivery is used to compute various error measures, such as the Root Mean Squared Error (RMSE) and the Mean Absolute Percentage Error (MAPE). Our Bayesian Heating Oil Forecaster and the existing forecasting method can be compared by assessing the difference in the error measures.

4.1.2 Evaluation Criteria

Three error metrics are used to compare the performance and the forecasting accuracy of our Bayesian Heating Oil Forecaster and the existing forecasting method. In all equations used in this section, d is the delivery number, i is the customer index, I_d is the total number of customers with d^{th} delivery, D_d is the total number of deliveries for the d^{th} delivery, $s_{i,d}$ is the estimated demand for the i^{th} customer's d^{th} delivery, , $\widehat{s}_{i,d}$ is the estimated demand for the i^{th} customer's d^{th} delivery, and c_i is the tank capacity for the i^{th} customer.

Root Mean Squared Error (RMSE) measures the average magnitude of the error. RMSE for the d^{th} delivery is computed using Equation 4.1. RMSE emphasizes large errors since it squares the error before it takes the average. The unit of the RMSE is the same as the data itself.

$$\text{RMSE}_d = \sqrt{\frac{\sum_{i=1}^{I_d}(\widehat{s}_{i,d} - s_{i,d})}{D_d}}. \tag{4.1}$$

Mean Absolute Percentage Error (MAPE) measures the accuracy of the forecast. MAPE Actual measures the average amount of error relative to the actual amount. MAPE Actual for the d^{th} delivery is computed using Equation 4.2.

$$\text{MAPE}_d^{actual} = \frac{1}{D_d}\sum_{i=1}^{I_d}\left|\frac{\widehat{s}_{i,d} - s_{i,d}}{s_{i,d}}\right|. \tag{4.2}$$

We also report MAPE Capacity, which measures the average amount of error relative to the tank capacity. MAPE Capacity for the d^{th} delivery is computed using Equation 4.3.

$$\text{MAPE}_d^{capacity} = \frac{1}{D_d}\sum_{i=1}^{I_d}\left|\frac{\widehat{s}_{i,d} - s_{i,d}}{c_i}\right|. \tag{4.3}$$

With both measures, lower error values indicate that the algorithm produced accurate estimates. When comparing the two algorithms, the algorithm with lower error values is desirable.

4.2 Models

To better compare the results, the following four models are used during the performance analysis in addition to our Bayesian Heating Oil Forecaster.

Bayes with Simple Average

The simple average model combines the estimated demand of the existing forecasting method and our Bayesian Heating Oil Forecaster by averaging the two estimates with equal weights.

Bayes with Expanded Expert Likelihood

One of the drawbacks of our Bayesian Heating Oil Forecaster is that the joint distribution for the expert likelihood is only sparsely populated. This is due to the limited availability of expert's historical initial $\mathcal{K}$-factor estimates. This can negatively affect the forecasts for the earlier deliveries, which rely on the expert likelihood. Since expert's estimates are available for most deliveries, this model uses expert's $\mathcal{K}$-factor estimates for not just the initial deliveries but for all deliveries. As a result, the joint distribution for the expert likelihood is populated with 28,572 observations instead of 175.

Bayes with Expanded Expert Likelihood, Simple Average

This model averages the estimates generated by the existing forecasting method and Bayes with Expanded Expert Likelihood.

Bayes with Simple Average, Season Flag

Over time, we have observed that our models tend to perform worse during the sixth and seventh deliveries for no apparent reason. Upon closer investigation, we discovered that these deliveries tend to occur during the summer because most customers sign up for the service at the beginning of the heating season. Since there is a significant difference in the weather pattern between seasons, models whose observations only include winter tends to perform worse during the summer. This model flags deliveries that occur during the summer months (April through September) and uses the existing forecasting method during the summer. Otherwise, it uses the estimates from Bayes with Simple Average.

The evaluation section compares the performance of the existing forecasting method, our Bayesian Heating Oil Forecaster, and the four models that are mentioned above.

4.3 Data Sets Used During the Test

The test uses two separate sets of historical heating oil delivery data that are provided by two different heating oil sales and distribution companies (Company A and Company B). Historical data from each company is split into training and test data sets. The training data is used to compute the prior and expert and model likelihoods. The test data set is used to evaluate the performance of the forecasts by comparing the estimated and actual heating oil demand. The development of our Bayesian Heating Oil Forecaster involved repeated testing using the data from company A. The data set from company B is reserved as a validation set and is used only for evaluation purposes.

The training data consists of historical delivery and estimation data between November 14, 2007, and September 30, 2009. Data between October 1, 2009, and September 30, 2010, is used as the test data. These dates are selected based on the availability of past delivery records. November 14, 2007, is the oldest date in which the company's past forecast results are recorded. From the discussions in Chapter 1, a data set with at least 18 months of data is required for forecasting purposes. Hence, the training data must include or exceed May 14, 2009. An end date of September 30, 2009, is selected to ensure that the training data is sufficient. The test data is between October 1, 2009, and September 30, 2010, so that one full winter heating season is observed.

For company A, 397 deliveries are observed for 120 customers who signed up for the service during the test period. There are about 3,000 existing customers that appear during the training period. For company B, 1,139 deliveries are observed for 291 customers who signed up for the service during the test period. There are about 2,300 existing customers that appear during the training period. For both companies, Figure 4.4 depicts the number of deliveries by delivery number.

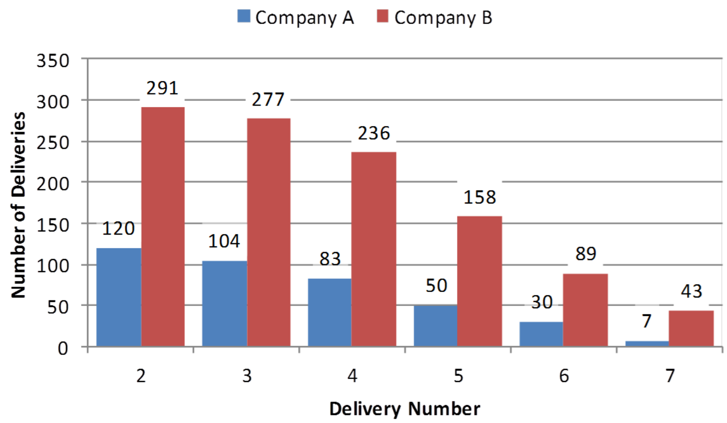

Figure 4.4: Number of Deliveries by Delivery Number

4.4 Trimming

To prevent unusual customer behaviors and other outliers from influencing the error measures, we report both trimmed and untrimmed results. 10% trimming of the *delivery* removes 5% of the deliveries with largest positive and negative errors. 10% trimming of the *customer* removes 5% of the customers whose deliveries have the largest positive and negative errors. Since 10% of the worst performing deliveries for one model is different from that of another, trimming by delivery causes each model to be evaluated on a different set of deliveries and customers. This can be avoided by trimming 10% of the customers with worst performing deliveries. By removing the same set of customers and all of its deliveries from the test set, all of the models are evaluated on the same set of customers and deliveries.

4.5 Chi-Square Goodness-of-Fit Test and Beta Probability Distributions

The Bayesian Heating Oil Forecaster uses beta distributions to represent beliefs and likelihoods. As it was discussed in Sections 2.2.2, 3.3.1, 3.3.3, and 3.3.6, the initial prior belief and the two likelihoods are obtained by performing a maximum likelihood estimation of the beta parameters. The beta distribution was chosen not because it best represents the empirical data, but because it drastically simplifies the update process. In general, we expect the performance of our Bayesian Heating Oil Forecaster to improve if the beta distribution fits the empirical data. Hence, we compare the beta distribution with other probability distributions, such as Normal, Weibull, Rayleigh, and Log Normal distributions, to evaluate the goodness of fit using the Chi-Square goodness-of-fit test.

The Chi-Square goodness-of-fit test is used to check the likelihood of a set of data coming from a specific distribution [20]. It is a statistical test whose null hypothesis states that a set of data comes from a specified distribution. The alternative hypothesis states that a set of data does not come from a specified distribution. If we fail to reject the null hypothesis, we conclude that there is **insufficient** evidence to state that the data does **not** follow a specified distribution. If we reject the null hypothesis, we conclude that it is unlikely that the data came from the specified distribution.

Readers interested in a more detailed discussion of the Chi-Squared goodness-of-fit test should consult a book by D'Agostino and Stephens [13]. This book covers the mathematical theories behind various goodness-of-fit techniques, including the Chi-Squared test. The NIST handbook [20] offers a more brief introduction to Chi-Squared and other goodness-of-fit techniques as well.

Figures 4.5 and 4.6 show different kinds of distributions that are fitted to the empirical distributions used by our Bayesian Heating Oil Forecaster. The parameters for the distributions are determined using the Maximum Likelihood Estimation. The goodness of fit of each distribution is tested using the Chi-Square goodness-of-fit test, whose results are shown in Figure 4.7. The p-values in the table indicate the probability of committing a type I error. Hence, a lower value indicates high confidence that the data does **not** come from a specified distribution. Since most entries in Figure 4.7 are zero, beta distribution, as well as all other distributions, have a poor fit to the empirical distribution. The test results suggest that other forms of more complex distributions, such as the Gaussian Mixture distributions, might exhibit a better fit to the empirical distributions used by our Bayesian Heating Oil Forecaster.

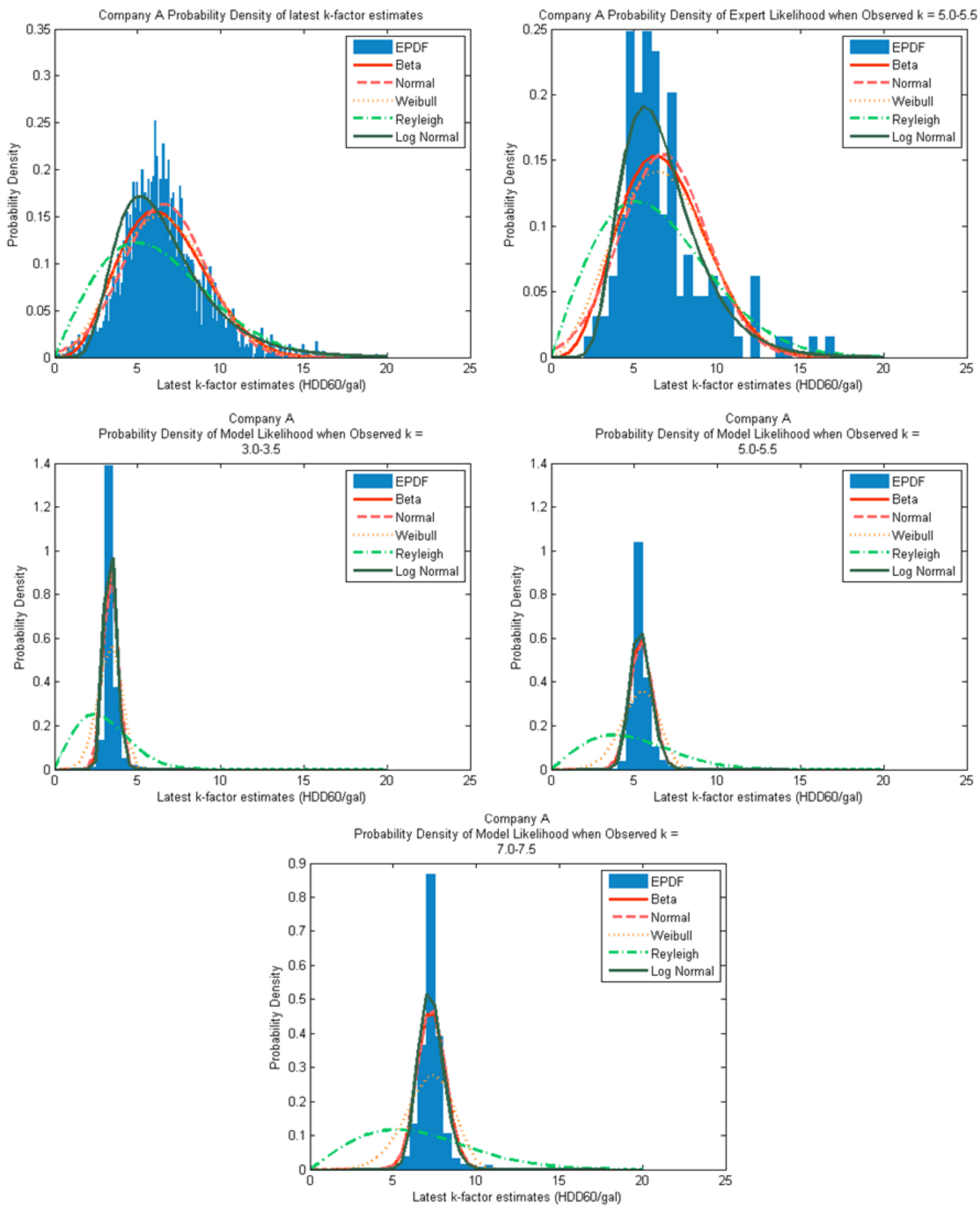

Figure 4.5: Comparison of the Fitness of Various Distributions for Company A

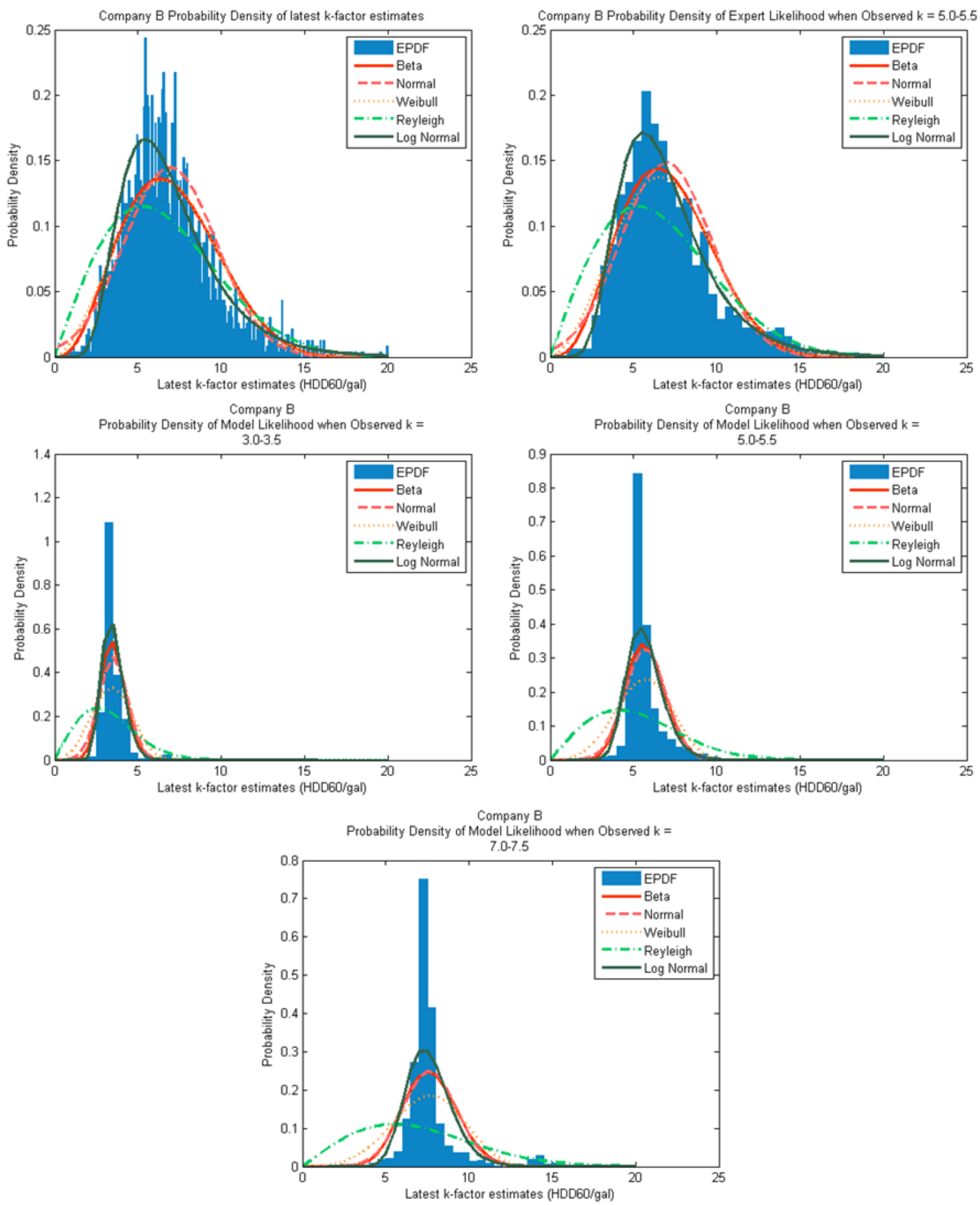

Figure 4.6: Comparison of the Fitness of Various Distributions for Company B

Company A	Beta	Normal	Weibull	Rayleigh	Log Normal
Initial Prior	4.24E-35	0	2.63E-45	0	0
Expert Likelihood (k=5.0-5.5)	0.001	3.00E-05	4.89E-05	7.0257E-08	**0.24**
Model Likelihood (k=3.0-3.5)	2.45E-38	3.11E-48	0	0	5.48E-32
Model Likelihood (k=5.0-5.5)	0	0	0	0	0
Model Likelihood (k=7.0-7.5)	2.28E-16	6.89E-17	0	0	4.81E-05

Company B	Beta	Normal	Weibull	Rayleigh	Log Normal
Initial Prior	2.14E-35	0	3.82E-39	0	7.41E-08
Expert Likelihood (k=5.0-5.5)	4.05E-10	1.04E-18	2.21E-12	9.24E-25	**0.12**
Model Likelihood (k=3.0-3.5)	0	0	0	0	0
Model Likelihood (k=5.0-5.5)	0	0	0	0	0
Model Likelihood (k=7.0-7.5)	0	0	0	0	0

Figure 4.7: P-values of the Goodness-of-Fit Chi-Squared Test

4.6 Results

This section discusses the results of the performance analysis by comparing the error metrics of the existing forecasting method, our Bayesian Heating Oil Forecaster, and four additional models proposed in Section 4.2.

Figure 4.8 compares RMSE, Figure 4.9 compares MAPE Actual, and Figure 4.10 compares MAPE Capacity of the six models. The x-axis lists two numbers: the top row is the delivery number, and the bottom row is the number of deliveries. For example, a top row with a value of 3 and a bottom row with a value of 150 indicates that 150 third deliveries were observed during the test set. 'All' indicates error measures taken across all of the deliveries regardless of their delviery numbers. The bar graphs on the left are for company A, while the graphs on the right are for company B. The top graphs show untrimmed results. The middle graphs show the results with 10% of the worst performing *deliveries* trimmed. The bottom graphs show the results with 10% of the worst performing *customers* trimmed. Trimming is based on the worst performing customers for the Bayes model with simple average. Each bar represents a particular model. From left to right: existing model, Bayes model, Bayes model with simple average, Bayes model with expanded expert likelihood, Bayes model with simple average expanded expert likelihood, and Bayes model with simple average season flag. For all of these figures, lower bars indicate better performance.

Figures 4.11, 4.12, and 4.13 list error metric percent changes of different models compared to the existing model. Negative percentage change indicates improved performance. Rows that contain Bayes with simple average models are highlighted, and negative percentage changes are maked with darker cells.

When we compare the trimmed results (graphs on the middle and bottom rows) with the untrimmed results (graphs on the top row), we notice that trimming reduces the overall RMSE by about 5 to 10 gallons. This indicates that a small number of deliveries with large positive and negative errors heavily influence the overall RMSE. This observation also holds true for other error measures. Trimming, however, did not significantly alter the behavior of the models: Bayes with simple average models remained either the best or the second best models with or without trimming.

Looking at Figures 4.11, 4.12, and 4.13, we see that Bayes models generally

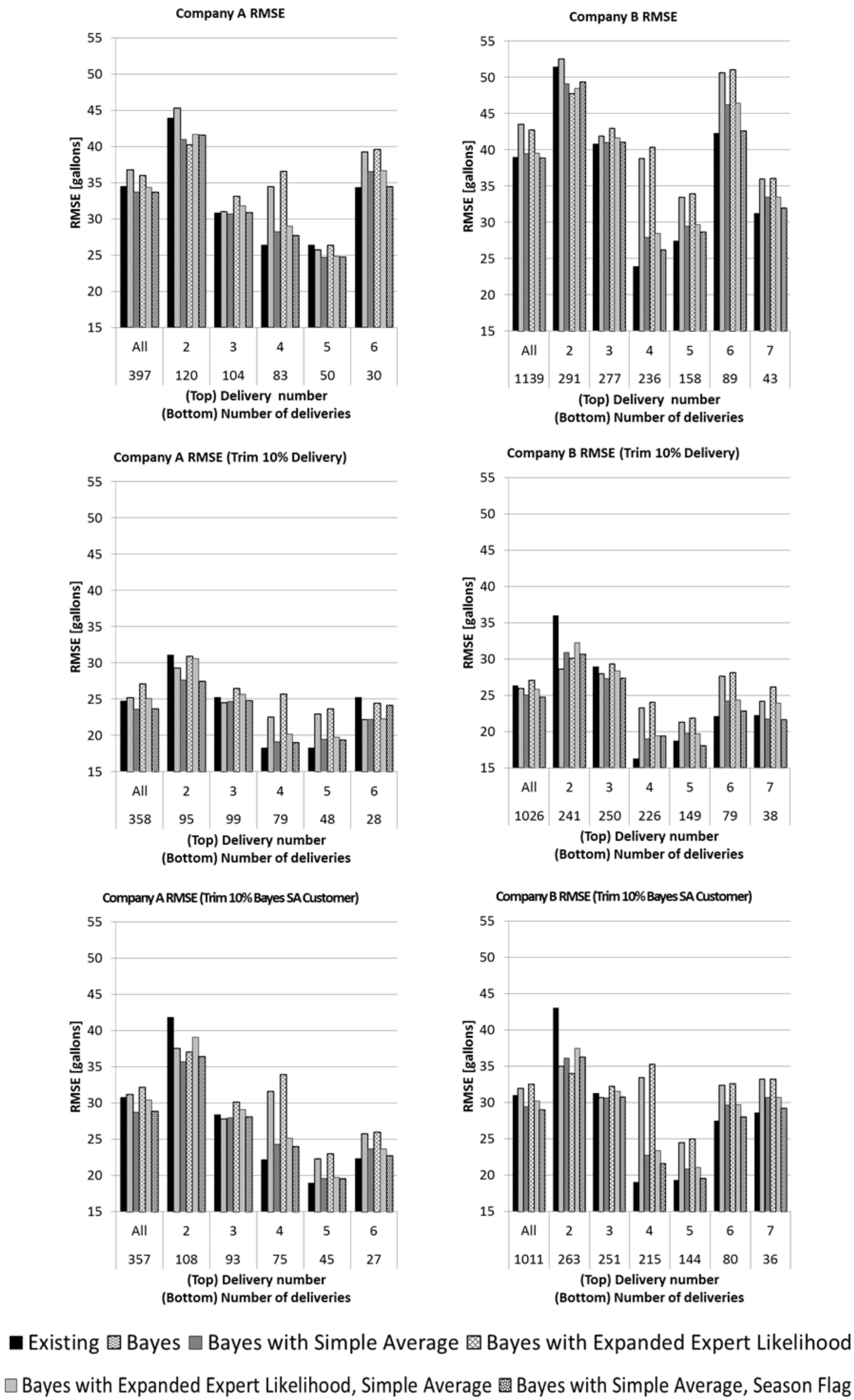

Figure 4.8: RMSE

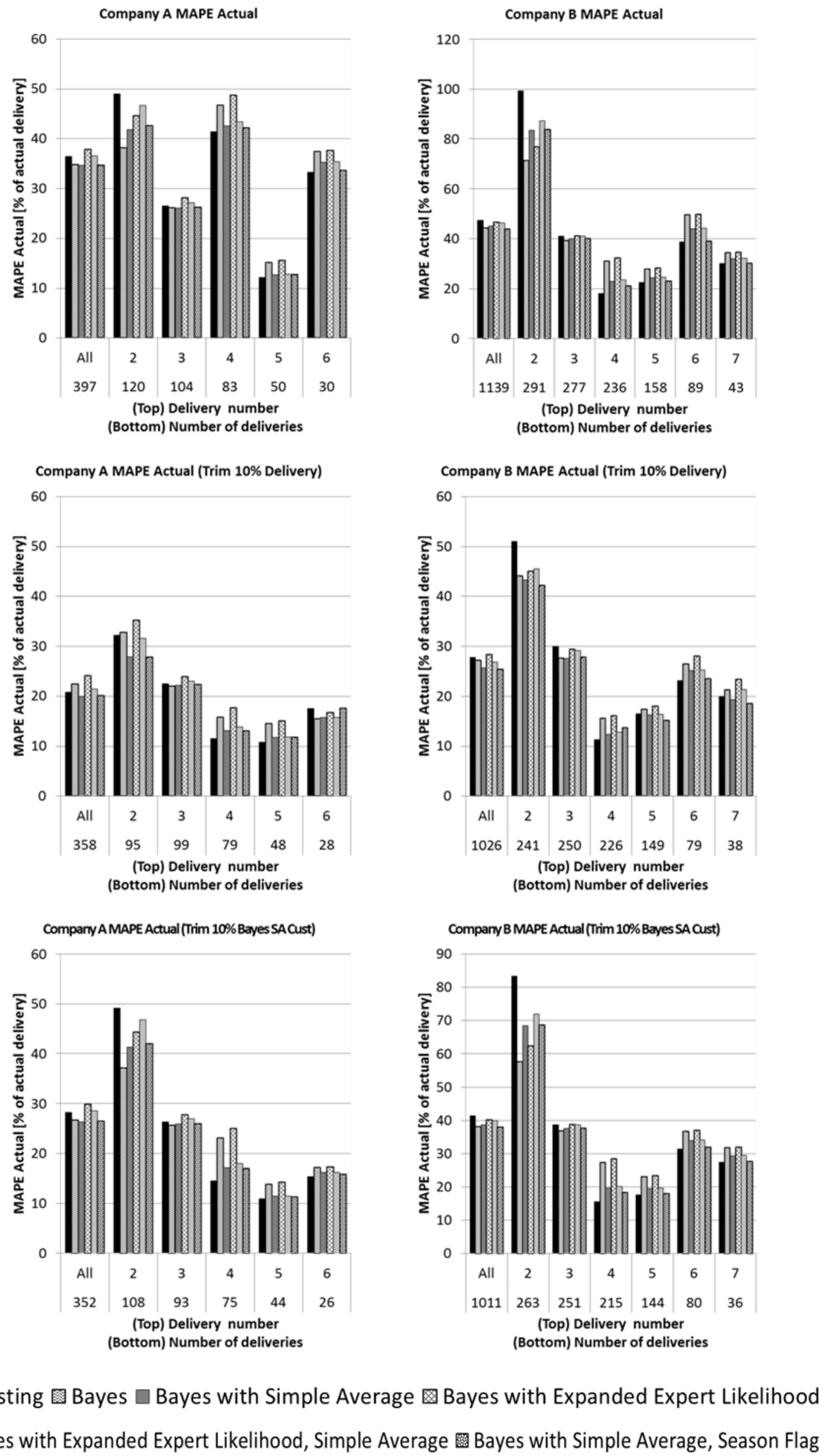

Figure 4.9: MAPE Actual

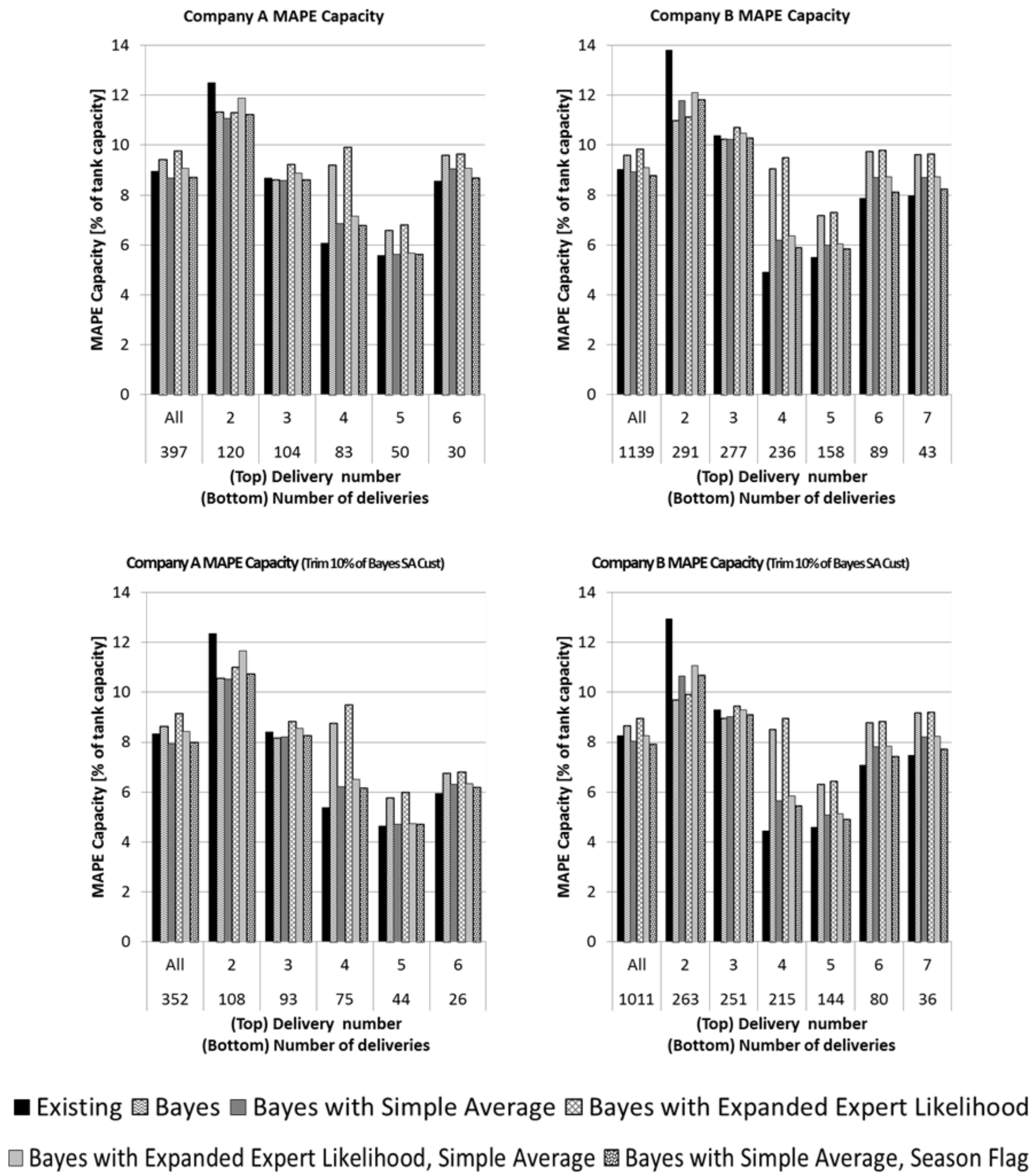

Figure 4.10: MAPE Capacity

RMSE % Change	Company A						Company B						
Delivery number	All	2	3	4	5	6	All	2	3	4	5	6	7
Number of deliveries	397	120	104	83	50	30	1139	291	277	236	158	89	43
Bayes	6.5%	3.0%	0.4%	30.5%	-2.6%	14.2%	11.7%	2.1%	2.7%	62.5%	22.1%	19.8%	15.3%
Bayes with SA	-2.4%	-6.7%	-0.3%	7.1%	-6.2%	6.3%	1.3%	-4.6%	0.6%	17.0%	7.6%	9.3%	7.1%
Bayes with Exp. Likelihood	4.2%	-8.6%	7.3%	38.5%	-0.3%	15.1%	9.7%	-7.2%	5.4%	69.1%	23.9%	20.8%	15.5%
Bayes with Exp. Likelihood, SA	-0.4%	-5.1%	3.2%	9.8%	-5.8%	6.7%	1.5%	-5.8%	2.1%	19.3%	8.3%	9.8%	7.2%
Bayes with SA, Season	-2.4%	-5.5%	0.1%	4.8%	-6.3%	0.3%	-0.3%	-4.0%	0.8%	9.8%	4.6%	0.7%	2.4%

MAPE Actual % Change	Company A						Company B						
Delivery number	All	2	3	4	5	6	All	2	3	4	5	6	7
Number of deliveries	397	120	104	83	50	30	1139	291	277	236	158	89	43
Bayes	-4.5%	-22.0%	-1.6%	12.6%	24.3%	12.3%	-6.2%	-28.1%	-3.8%	72.4%	24.4%	28.1%	15.2%
Bayes with SA	-5.0%	-14.6%	-1.2%	2.7%	4.9%	6.0%	-5.3%	-15.9%	-2.4%	26.6%	8.7%	13.5%	6.8%
Bayes with Exp. Likelihood	3.7%	-9.1%	6.3%	17.6%	28.1%	13.0%	-1.6%	-22.6%	0.8%	79.3%	26.3%	29.0%	15.5%
Bayes with Exp. Likelihood, SA	0.4%	-4.6%	2.7%	4.8%	5.8%	6.3%	-2.4%	-12.1%	0.1%	29.7%	9.6%	13.9%	6.9%
Bayes with SA, Season	-4.8%	-13.2%	-0.9%	1.8%	4.7%	1.0%	-7.1%	-15.6%	-2.0%	17.0%	2.8%	0.9%	1.2%

MAPE Capacity % Change	Company A						Company B						
Delivery number	All	2	3	4	5	6	All	2	3	4	5	6	7
Number of deliveries	397	120	104	83	50	30	1139	291	277	236	158	89	43
Bayes	5.3%	-9.5%	-0.9%	51.7%	18.2%	12.2%	6.2%	-20.5%	-1.4%	84.5%	30.3%	23.8%	20.5%
Bayes with SA	-2.8%	-11.5%	-1.2%	13.1%	1.2%	5.8%	-1.3%	-14.6%	-1.5%	26.2%	9.0%	10.6%	9.2%
Bayes with Exp. Likelihood	9.1%	-9.6%	6.2%	63.3%	22.1%	12.8%	9.0%	-19.5%	3.0%	93.3%	32.3%	24.4%	20.8%
Bayes with Exp. Likelihood, SA	1.6%	-4.9%	2.3%	17.7%	1.9%	6.1%	0.9%	-12.2%	1.0%	29.8%	9.9%	10.9%	9.3%
Bayes with SA, Season	-2.7%	-10.2%	-0.8%	11.8%	1.1%	1.6%	-2.8%	-14.4%	-1.0%	20.1%	5.9%	3.2%	3.2%

Figure 4.11: RMSE and MAPE Percent Change

performed poorly on the 4^{th} and later deliveries. The results of the 4^{th} delivery was usually the worst, and performance improved on subsequent deliveries. Since our Bayesian Heating Oil Forecaster is an iterative algorithm, the poor results of the 4th delivery seems to be negatively affecting subsequent deliveries. This observation remained true for both trimmed and untrimmed results.

Company B responded better to the seasonal model than company A. With company A, simple average model with and without the season flag exhibited very similar performance. With company B, simple average model with the season flag usually performed better than the simple average model without the season flag.

Bayes with simple average models exhibited performance improvements with

RMSE % Change Trim 10% Bayes SA Customer	Company A						Company B						
Delivery number	All	2	3	4	5	6	All	2	3	4	5	6	7
Number of deliveries	352	108	93	75	44	26	1011	263	251	215	144	80	36
Bayes	1.4%	-10.2%	-1.8%	42.2%	18.1%	15.4%	3.1%	-18.7%	-1.6%	75.6%	26.7%	18.0%	16.1%
Bayes with SA	-6.7%	-14.6%	-1.4%	9.5%	3.3%	5.9%	-4.9%	-16.0%	-1.8%	19.4%	8.0%	7.9%	7.4%
Bayes with Exp. Likelihood	4.5%	-11.4%	6.3%	52.9%	21.6%	16.3%	5.0%	-20.9%	3.3%	85.0%	29.5%	18.8%	16.3%
Bayes with Exp. Likelihood, SA	-1.0%	-6.6%	2.6%	13.1%	4.3%	6.2%	-2.4%	-13.0%	0.9%	22.7%	9.1%	8.3%	7.4%
Bayes with SA, Season	-6.1%	-12.9%	-0.9%	8.1%	3.3%	1.9%	-6.4%	-15.8%	-1.5%	13.4%	1.5%	1.9%	2.3%

MAPE Actual % Change Trim 10% Bayes SA Customer	Company A						Company B						
Delivery number	All	2	3	4	5	6	All	2	3	4	5	6	7
Number of deliveries	352	108	93	75	44	26	1011	263	251	215	144	80	36
Bayes	-5.3%	-24.3%	-2.8%	59.8%	27.4%	12.7%	-7.6%	-30.7%	-4.8%	76.6%	31.1%	17.2%	16.3%
Bayes with SA	-6.5%	-15.9%	-1.9%	18.4%	4.5%	5.8%	-6.4%	-17.8%	-2.9%	27.1%	10.8%	7.8%	7.3%
Bayes with Exp. Likelihood	6.2%	-9.6%	5.5%	72.6%	31.5%	13.3%	-2.5%	-25.0%	0.4%	84.5%	33.6%	18.1%	16.7%
Bayes with Exp. Likelihood, SA	1.1%	-4.8%	2.3%	23.7%	5.3%	6.1%	-3.2%	-13.6%	-0.1%	30.6%	12.0%	8.3%	7.4%
Bayes with SA, Season	-5.8%	-14.4%	-1.5%	17.2%	4.3%	3.6%	-7.9%	-17.4%	-2.5%	19.0%	2.8%	1.9%	1.3%

MAPE Capacity % Change Trim 10% Bayes SA Customer	Company A						Company B						
Delivery number	All	2	3	4	5	6	All	2	3	4	5	6	7
Number of deliveries	352	108	93	75	44	26	1011	263	251	215	144	80	36
Bayes	3.6%	-14.6%	-2.9%	62.2%	24.5%	13.8%	4.7%	-25.3%	-3.7%	91.4%	37.5%	23.8%	22.8%
Bayes with SA	-4.8%	-14.7%	-2.4%	15.2%	1.4%	6.4%	-2.9%	-17.8%	-2.8%	27.4%	10.8%	10.4%	10.2%
Bayes with Exp. Likelihood	9.6%	-10.8%	5.0%	76.1%	28.9%	14.4%	8.3%	-23.5%	1.6%	101.7%	40.1%	24.5%	23.1%
Bayes with Exp. Likelihood, SA	1.3%	-5.5%	1.6%	20.6%	2.0%	6.7%	-0.2%	-14.6%	0.2%	31.6%	12.0%	10.7%	10.3%
Bayes with SA, Season	-4.2%	-13.2%	-1.9%	14.4%	1.4%	4.0%	-4.1%	-17.5%	-2.2%	22.9%	6.7%	4.8%	3.4%

Figure 4.12: RMSE and MAPE Percent Change (Trim 10% Customer)

RMSE % Change Trim 10% Delivery	Company A						Company B						
Delivery number	All	2	3	4	5	6	All	2	3	4	5	6	7
Number of deliveries	358	95	99	79	48	28	1026	241	250	226	149	79	38
Bayes	1.9%	-5.8%	-2.6%	23.6%	25.9%	-12.1%	-1.6%	-20.5%	-3.4%	43.1%	13.7%	24.8%	8.7%
Bayes with SA	-4.7%	-11.2%	-2.2%	4.9%	6.8%	-12.1%	-4.9%	-14.0%	-5.4%	17.0%	5.6%	9.8%	-2.0%
Bayes with Exp. Likelihood	9.6%	-0.5%	5.2%	40.9%	29.9%	-3.1%	2.5%	-16.4%	1.2%	47.6%	16.6%	27.2%	17.5%
Bayes with Exp. Likelihood, SA	1.2%	-1.7%	1.7%	10.6%	8.0%	-12.0%	-1.8%	-10.3%	-2.0%	19.5%	5.3%	10.2%	7.7%
Bayes with SA, Season	-4.3%	-11.7%	-1.5%	4.3%	6.4%	-4.4%	-6.4%	-14.8%	-5.4%	19.0%	-3.6%	3.2%	-2.8%

MAPE Actual % Change Trim 10% Delivery	Company A						Company B						
Delivery number	All	2	3	4	5	6	All	2	3	4	5	6	7
Number of deliveries	358	95	99	79	48	28	1026	241	250	226	149	79	38
Bayes	7.9%	1.6%	-2.2%	37.4%	33.5%	-12.1%	-2.4%	-13.6%	-7.8%	36.8%	5.1%	13.9%	6.7%
Bayes with SA	-3.7%	-13.3%	-1.4%	13.8%	8.3%	-10.3%	-7.7%	-15.0%	-7.9%	9.3%	-1.5%	8.5%	-3.2%
Bayes with Exp. Likelihood	16.0%	8.9%	6.0%	54.0%	38.0%	-4.9%	1.7%	-11.8%	-2.0%	41.8%	8.8%	20.6%	17.1%
Bayes with Exp. Likelihood, SA	3.3%	-1.9%	2.7%	20.3%	9.4%	-10.1%	-3.6%	-10.6%	-2.6%	12.6%	-1.1%	9.0%	7.1%
Bayes with SA, Season	-3.4%	-13.8%	-0.9%	13.2%	8.0%	0.0%	-8.9%	-17.3%	-7.2%	20.6%	-8.7%	1.3%	-7.1%

Figure 4.13: RMSE and MAPE Percent Change (Trim 10% Delivery)

the second, third, and overall deliveries. This result remained true regardless of trimming, error measures, and company (except for untrimmed RMSE for company

B). This is a strong evidence that the Bayes with simple average models are effective during the initial deliveries. The largest improvements can be seen with the second delievery, with 10% to 16% reduction in MAPE for both companies. The overall error is reduced by about two to eight percent. The largest overall improvement of 8.9% was observed with MAPE Actual for company B when the 10% of the worst deliveries were trimmed (Figure 4.13).

Next chapter concludes the thesis by summarizing its findings and results. The chapter also discusses potential extensions and improvements to our Bayesian Heating Oil Forecaster.

CHAPTER 5

Conclusions and Future Research

5.1 Conclusions

The goal of this thesis was to develop an algorithm that, compared to the existing forecasting method, reduces the error between the new customers' forecast and actual heating oil demand during initial deliveries. We have presented a novel forecasting algorithm in Chapter 3 which uses forecasters' past performances for existing customers to adjust the current forecast for target customers. We have adapted a Bayesian approach to forecasting [4; 5] combined with domain knowledge and original ideas to develop our Bayesian Heating Oil Forecaster which forecasts demand for target customers without relying on their historical deliveries.

Performance evaluation presented in Chapter 4 demonstrated that our Bayesian Heating Oil Forecaster showed increased performance over the existing forecasting method when the two techniques are combined. We used Root Mean Squared Error (RMSE), Mean Absolute Percent Error (MAPE) Actual, and MAPE Capacity to compare the performance of the two algorithms. Compared to the existing forecasting method alone, our Simple Average model, which combines the forecasts from the existing forecasting method and our Bayesian Heating Oil Forecaster, recorded an overall improvement of 6.7% in RMSE, 6.5% in MAPE Actual, and 4.8% in MAPE Capacity when 10% of the worst performing customers for company A are removed. When using all of the customers for company A, the improvements were 2.4%, 5.0%, and 2.8%, respectively. Company B reported similar results using the simple average model with season flag. When 10% of the worst performing customers were removed, the RMSE, MAPE Actual, and MAPE Capacity improved by 6.4%, 7.9%, and 4.1% respectively. When untrimmed data was used, the improvements were 0.3%, 7.1%, and 2.8% respectively. This improvement was attained without requiring additional information about the customers. Furthermore, the algorithm succeeded in reducing the overall error across three different error measures for two different companies with or without trimming the test results. This is a strong evidence that our Bayesian Heating Oil Forecaster is effective in reducing the error during the initial deliveries.

It should also be noted that, due to the limited availability of the training data, the training set is less than two years long. Additionally, the backtesting process did not update the likelihoods during the test period. The Bayes models were forecasting deliveries that occurred in September 2010 using likelihoods that were trained between November 2007 and October 2009. In the actual operation, the likelihoods will be updated each day, allowing the algorithm to forecast demand using current data. Hence, it is expected that the model's performance will be better under actual operating condition as it accumulates more data and it uses up-to-date likelihoods during daily forecast.

A direct impact of improved forecasts is a reduction in operational expenses. The majority of the cost savings is assumed to be the result of reducing the number of unnecessary deliveries. Using the existing forecasting method, company A made approximately 18,000 deliveries a year. Since most customers are living in rural areas, assume that the average delivery time is 30 minutes and its travel distance is 10 miles. Furthermore, assume that the delivery person works for $15 per hour, fuel economy of delivery trucks is around 10 miles per gallon, and a gallon of diesel fuel costs 4 dollars per gallon. 18,000 deliveries requires 9,000 hours and 18,000 gallons of fuel for a total cost of $207,000. If reduction of error directly results in reduction of deliveries, a 5% reduction in deliveries would save approximately $10,000 annually.

5.2 Recommendations

The Bayesian Heating Oil Forecaster was found to be most effective when it was combined with the existing forecasting method. Hence, Bayes model with simple average should be used as the primary forecasting method instead of the existing forecasting method. Demand forecasts for company B should use Bayes model with simple average and season flag since it performed better compared to the model without the season flag. We were only able to test the performance of our Bayesian Heating Oil Forecaster for the first six to seven deliveries. Hence the model's performance is not known beyond the first year and a half. It is recommended that the performance of the model be reevaluated after one year to test its effectiveness beyond the first few deliveries.

5.3 Future Research

Although we proposed a feasible method to reduce the forecast error for new customers' heating oil demand during initial deliveries, there may still be some improvements that can be made to our method. Listed below are some possible improvements and enhancements to our Bayesian Heating Oil Forecaster.

- Our Bayesian Heating Oil Forecaster uses beta distributions because of its multiplicative properties. As demonstrated in Section 4.5, beta distributions do not fit our empirical distributions very well. Using other forms of distributions, such as Gaussian or beta mixture models, can improve the fit and performance of our algorithm.
- When we fit beta distributions to the joint distributions for the two likelihoods, we fit individual beta distributions to the columns of the joint distributions. Since the joint distributions form a surface across rows and

columns, fitting a surface to the entire joint probability distribution can improve the fit, and in turn improve the performance of our algorithm.

- Since the performance analysis indicates a lower performance on the fourth deliveries, a rule-based forecasting method that adjusts the forecasts based on the delivery number is likely to improve the performance of our algorithm.
- Our Bayesian Heating Oil Forecaster adjusts the heatload coefficient ($\mathcal{K}$-factor) but not the baseload coefficient. If our Bayesian Heating Oil Forecaster is extended to adjust both the baseload and the heatload coefficients, the overall forecast is expected to improve.
- The model with the best performance was the model that combined the forecasts from the existing forecasting method and our Bayesian Heating Oil Forecaster. Use of other combination techniques, such as rule-based unequal weighting, can improve the performance of the model.
- Although our Bayesian Heating Oil Forecaster reports point estimates of the estimated demand, the algorithm uses probability distribution throughout the computation process. The algorithm can be extended to provide additional information about its estimates by reporting probability forecasts instead of point forecasts.

BIBLIOGRAPHY

[1] M. B. Araújo and M. New, "Ensemble forecasting of species distributions," *Trends in Ecology & Evolution*, vol. 22, no. 1, pp. 42–47, 2007.

[2] J. S. Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners.* Boston: Kluwer Academic, 2001.

[3] R. J. Beckman and G. L. Tiet jen, "Maximum likelihood estimation for the beta distribution," *Journal of Statistical Computation and Simulation*, vol. 7, no. 3, p. 253, 1978.

[4] D. A. Berry, *Statistics: A Bayesian Perspective.* Belmont, CA: Duxbury Press, 1996.

[5] W. M. Bolstad, *Introduction to Bayesian Statistics.* Hoboken, NJ: Wiley-Interscience, 2007.

[6] B. L. Bowerman, R. T. O'Connell, A. B. Koehler, and B. L. Bowerman, *Forecasting, Time Series, and Regression: An Applied Approach.* Belmont, CA: Thomson Brooks/Cole, 2005.

[7] R. H. Brown, "Improved forecasting by leveraging external weather and consumption data," October 2008, presented at Gas Forecaster's Forum.

[8] G. Casella, "An introduction to empirical Bayes data analysis," *The American Statistician*, vol. 39, no. 2, pp. 83–87, May 1985.

[9] R. T. Clemen, "Combining forecasts: A review and annotated bibliography," *International Journal of Forecasting*, vol. 5, no. 4, pp. 559–583, 1989.

[10] A. S. Cofino, R. Cano, C. Sordo, and J. M. Gutiérrez, "Bayesian networks for probabilistic weather prediction," in *The 15th Eureopean Conference on Artificial Intelligence. Proceedings.*, 2002, pp. 695–699.

[11] G. F. Cooper and T. Dietterich, "A Bayesian method for the induction of probabilistic networks from data," in *Machine Learning. Proceedings.*, 1992, pp. 309–347.

[12] K. Crombecq, L. De Tommasi, D. Gorissen, and T. Dhaene, "A novel sequential design strategy for global surrogate modeling," in *Winter Simulation Conference (WSC). Proceedings.*, 2009, pp. 731–742.

[13] R. B. D'Agostino and M. A. Stephens, *Goodness-of-fit techniques.* New York: Marcel Dekker, 1986.

[14] E. de Alba, "Disaggregation and forecasting: A Bayesian analysis," *Journal of Business & Economic Statistics*, vol. 6, no. 2, pp. 197–206, 1988.

[15] N. Dhillon, "Natural gas load forecasting using ensembles of multiple models to improve accuracy," 2008, Personal correspondence.

[16] G. Duncan, W. Gorr, and J. Szczypula, "Bayesian forecasting for seemingly unrelated time series: Application to local government revenue forecasting," *Management Science*, vol. 39, no. 3, pp. 275–293, Mar. 1993.

[17] D. W. Fan, P. K. Chan, and S. J. Stolfo, "A comparative evaluation of combiner and stacked generalization," in *In Proceedings of AAAI-96 workshop on Integrating Multiple Learned Models*, 1996, pp. 40–46.

[18] D. Fenn, "Reinventing your family business," Oct. 2010, www.bnet.com/blog/entrepreneurs/reinventing--your--family--business/1345 (Accessed 1 Apr. 2011).

[19] J. J. Filliben, "Beta distribution from NIST/SEMATECH e-handbook of statistical methods," Jun. 2010, www.itl.nist.gov/div898/handbook/eda/section3/eda366h.htm (Accessed 1 Apr. 2011).

[20] ——, "Chi-Square goodness-of-fit test from NIST/SEMATECH e-handbook of statistical methods," Jun. 2010, http://itl.nist.gov/div898/handbook/eda/section3/eda35f.htm (Accessed 1 Jun. 2011).

[21] A. Garcia-Ferrer, R. A. Highfield, F. Palm, and A. Zellner, "Macroeconomic forecasting using pooled international data," *Journal of Business & Economic Statistics*, vol. 5, no. 1, pp. 53–67, Jan. 1987.

[22] B. Ghosh, B. Basu, and M. O'Mahony, "Bayesian time-series model for short-term traffic flow forecasting," *Journal of Transportation Engineering*, vol. 133, no. 3, pp. 180–189, Mar. 2007.

[23] M. A. Gluck and C. Myers, *Gateway to Memory: An Introduction to Neural Network Modeling of the Hippocampus and Learning*. Cambridge, MA: MIT Press, 2001.

[24] M. Green and P. J. Harrison, "Fashion forecasting for a mail order company using a Bayesian approach," *Operational Research Quarterly*, vol. 24, no. 2, pp. 193–205, Jun. 1973.

[25] K. Gurney, *An Introduction to Neural Networks*. London: UCL Press, 1997.

[26] J. Harrison and M. West, "Practical Bayesian forecasting," *Journal of the Royal Statistical Society Series D (The Statistician)*, vol. 36, no. 2/3, Special Issue: Practical Bayesian Statistics, pp. 115–125, 1987.

[27] P. J. Harrison and C. F. Stevens, "A Bayesian approach to short-term forecasting," *Operational Research Quarterly*, vol. 22, no. 4, pp. 341–362, Dec. 1971.

[28] ——, “Bayesian forecasting,” *Journal of the Royal Statistical Society Series B (Methodological)*, vol. 38, no. 3, pp. 205–247, 1976.

[29] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*. New York: Wiley & Sons, 1994.

[30] F. R. Johnston and P. J. Harrison, “An application of forecasting in the alcoholic drinks industry,” *The Journal of the Operational Research Society*, vol. 31, no. 8, pp. 699–709, Aug. 1980.

[31] H. Migon and A. Monteiro, “Rain-fall modeling: An application of Bayesian forecasting,” *Stochastic Hydrology and Hydraulics*, vol. 11, no. 2, pp. 115–127, Apr. 1997.

[32] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*. Hoboken, NJ: Wiley-Interscience, 2006.

[33] A. W. Moore, “Bayes nets for representing and reasoning about uncertainty,” p. 46, Oct. 2001, www.autonlab.org/tutorials/bayesnet09.pdf (Accessed 1 Apr. 2011).

[34] ——, “Learning with maximum likelihood,” p. 25, Sep. 2001, www.autonlab.org/tutorials/mle13.pdf (Accessed 1 Apr. 2011).

[35] I. J. Myung, “Tutorial on maximum likelihood estimation,” *Journal of Mathematical Psychology*, vol. 47, no. 1, pp. 90–100, 2003.

[36] R. E. Neapolitan, *Probabilistic Methods for Bioinformatics with an Introduction to Bayesian Networks*. Boston: Morgan Kaufmann/Elsevier, 2009.

[37] C. Otrok and C. H. Whiteman, “Bayesian leading indicators: Measuring and predicting economic conditions in Iowa,” *International Economic Review*, vol. 39, no. 4, pp. 997–1014, Nov. 1998.

[38] S. Pezzulli, P. Frederic, S. Majithia, S. Sabbagh, E. Black, R. Sutton, and D. Stephenson, “The seasonal forecast of electricity demand: a hierarchical Bayesian model with climatological weather generator,” *Applied Stochastic Models in Business and Industry*, vol. 22, no. 2, pp. 113–125, 2006.

[39] A. Pole, M. West, and J. Harrison, *Applied Bayesian Forecasting and Time Series Analysis*. Boca Raton, FL: Chapman & Hall/CRC, 1999.

[40] O. Pourret, P. Naim, and B. Marcot, *Bayesian Networks: A Practical Guide to Applications*. Hoboken, NJ: John Wiley, 2008.

[41] R. Prado and M. West, *Time Series Modeling, Computation, and Inference*. Florida: Chapman & Hall/CRC, 2010.

[42] F. Robert, “An evaluation of Bayesian forecasting,” *Journal of Forecasting*, vol. 2, no. 2, p. 137, Apr. - Jun. 1983.

[43] F. Ruggeri, R. Kenett, F. W. Faltin, and Knovel, *Encyclopedia of Statistics in Quality and Reliability*. Chichester, England: John Wiley, 2007.

[44] J. Q. Smith, "Forecasting accident claims for an assurance company," *Journal of the Royal Statistical Society Series D (The Statistician)*, vol. 32, no. 1/2, pp. 109–115, Mar. - Jun. 1983.

[45] C. Spieth, "Mathematical models chapter 3 algorithms," 2004, www.ra.cs.uni-tuebingen.de/software/JCell/tutorial/ch03s03.html (Accessed 1 Apr. 2011).

[46] C. F. Stevens, "On the variability of demand for families of items," *Operational Research Quarterly*, vol. 25, no. 3, pp. 411–419, Sep. 1974.

[47] S. M. Stigler, *The History of Statistics: The Measurement of Uncertainty Before 1900.* Cambridge, MA: Belknap Press of Harvard University Press, 1986.

[48] S. R. Vitullo, R. H. Brown, G. F. Corliss, and B. M. Marx, "Mathematical models for natural gas forecasting," *Canadian Applied Matehmatics Quarterly*, vol. 19, no. 3, 2011, pending publication Fall 2011.

[49] M. West and J. Harrison, *Bayesian Forecasting and Dynamic Models.* New York: Springer, 1997.

[50] M. West, P. J. Harrison, and H. S. Migon, "Dynamic generalized linear models and Bayesian forecasting," *Journal of the American Statistical Association*, vol. 80, no. 389, pp. 73–83, Mar. 1985.

[51] C. Zhang, S. Sun, and G. Yu, "A Bayesian network approach to time series forecasting of short-term traffic flows," in *The 7th International IEEE Conference on Intelligent Transportation Systems. Proceedings.*, 2004, pp. 216–221.