SUPPLY CHAIN OPTIMIZATION THROUGH BLOCKCHAIN UNDER BAYESIAN FRAMEWORK USING MCMC METHODS

A STUDY PRESENTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTERS OF SCIENCE
IN
MANUFACTURING ENGINEERING AND MANAGEMENT

BY

IQRA JOHIM
NUST-MS-2015-00000117715

SUPERVISOR

CAPT. DR. TARIQ MAIRAJ RASOOL KHAN PN

INDUSTRIAL & MANUFACTURING ENGINEERING DEPARTMENT

PAKISTAN NAVY ENGINEERING COLLEGE, KARACHI

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY, ISLAMABAD

MAY,2019

DECLARATION

I hereby certify that the research work titled "Supply Chain Optimization through Blockchain under Bayesian Framework using MCMC Methods" is my own work and none of the material in this thesis/ research work has been submitted in support for another degree or qualification from another institute, organization or University. Except that research paper based on this thesis work has been submitted to international conference/ journal for acceptance of research paper for publication.

IQRA JOHIM

RESEARCH TEAM MEMBERS

Thesis Supervisor

Dr. Tariq Mairaj

Guidance and Evaluation Committee (GEC) Members

Dr. Aleem Mushtaq

Dr. Tahir Hussain

Dedication

This write-up is dedicated to my treasured parents, my endeared siblings with whose unsurpassed support, encouragement and love I have made this wonderful accomplishment.

Acknowledgment

All glory to ALLAH for the His countless blessings in completing this thesis in time. I would like to give my sincere gratitude to my supervisor Dr. Tariq Khan for the throughout guidance and support. I attribute the level of my Master's degree to his inspiration and determination without him, this work would not be possible. His consistency, promptness, regularity help me in completing my thesis in an efficient way.

I appreciate Dr. Salman Nisar for his helpful suggestions, comments, encouragement and guidance on the various aspects on research thesis. I would also like to offer thanks to Dr. Tahir Hussain, and Dr. Aleem Mushtaq for serving on my thesis committee. I would also like to thank Engr. Qasim Ali & Engr. Rabia Ghani for their motivation and support. Moreover, I wish to thank my family members, specially my father for his support and incentive for enrolling into Masters and giving me chance to work on their business data without which, this thesis was not possible.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence
AHP	Analytical Hierarchy Process
AAA	Authentication Authorization and Accounting.
BC	Block chain
BF	Bayes' Factor
BIC	Bayesian Information Criterion
BDMCMC	Birth & Death MCMC
B2B	Business to Business
BMS	Business Models
DSS	Decision Support System
DIC	Deviance Information Criterion
ES	Enterprise System
EBC	Enterprise Block Chain
FMCMC	Fusion MCMC
GUI	Graphical User Interface
KLD	Kullback Libler Divergence
MLE	Maximum Likelihood Estimation
MCMC	Markov Chain Monte Carlo
PDSA	Plan Do Study Action cycle
RJMCMC	Reversible Jump MCMC
SME	Small Medium Enterprise
SCM	Supply Chain Management
SoR	System of Records

Abstract

Improved supply chain management has gained immense significance in recent years because of the associated countless benefits. Supply Chain management is a challenging task in today's world due to the encountered uncertainties. In Pakistan forecasting in supply chain is more challenging due to smaller industry base, longer lead times, limited client tail, political instability, lack of warehousing facilities and the smaller market quantum. The aim of the proposed research is to overcome the problems associated with supply chain in Pakistan. A model will be developed to optimize the supply chain using Bayesian inference Markov chain Monte Carlo (MCMC) method. During the research, a well- established supply chain business was taken as a case study whose historical as well as current supply chain data was used in the research. MCMC methods are developed and validated for better forecasting. The Particle Filter (PF) technique is used to estimate posterior PDF (probability density function) from prior PDF. PF is a sequential Monte Carlo technique which represents probability densities in samples or particles along with their weights. So computed Posterior Estimates are then considered as predicted forecast. This prediction scheme evolves with passage of time as more data is available. A framework is developed which facilitates the application of PF technique iteratively.

This project comprises of two parts. First part is to construct an enterprise system (ES). Second part is to produce correct forecasting models for an Enterprise and update it along time with current data set. The methodology follows the Plan, Do, Study and Action (PDSA) cycle. During the research, a well- established supply chain business will offer

the historical as well as current data where MCMC methods will be developed and validated for better forecasting.

The research commences with the development of framework and data acquisition in ‘Plan’ phase. The second step is the development of ES which will move towards ‘Do’ part. Afterwards, the research will proceed towards MCMC algorithm development. Furthermore, the model (output) will be verified and improved in the ‘Study’ phase. Finally, documentation and updation will be done under the ‘Action’ phase.

The research outcome will be a model that can be used to increase the profit in a supply chain business through improved stock levels in the inventory management.

Chapter 1 Introduction

1.1 Background and Motivation

The Great Recession of 2007-2009 is known to be the greatest downturn of the last decade. The impact was huge [1]. The financial reports published in 2011 suggested that the recession was avoidable [2].

Due to the great recession, market heavily collapse around the world which leaves the large audience in debt. This recession motivates to have such a system which evolves with time and has the capacity to learn from the given parameters. These parameters help in directing the short-comings of any system with-in the supply chain. So that, such an uncertainty can be avoided.

The major in-country challenges in optimization of supply chain includes, international market steel prices fluctuation, commodity prices fluctuation, US Dollar rate fluctuation, shipment costs, warehouse rents, law and order situation, fluctuating Government regulations, fluctuating custom clearance regulation/policies/ rates, customer of commodity shift related policies, taxation regimes.

The aim of the thesis is to optimize the supply chain by ensuring following actions, including, meeting up the orders on time, minimize the overhead cost and correct inventory forecast. This will be ensured through Bayesian inference Markov chain Monte Carlo (MCMC) methods which has the ability to learn and evolve with time.

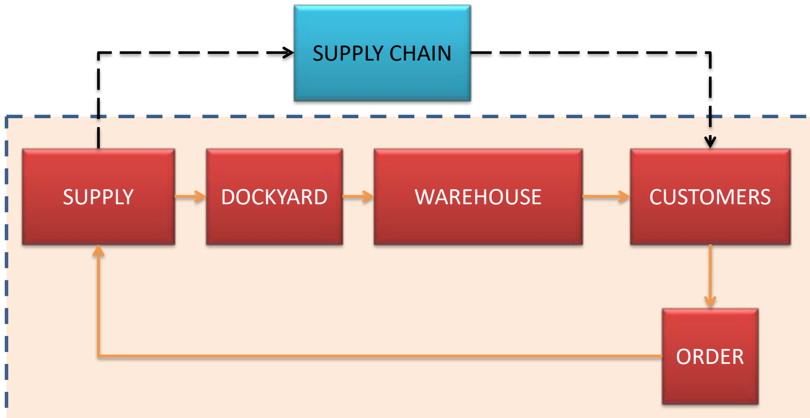


Figure 1-1 MHCO Steel Supply Chain

The uncertainty in capturing orders is modeled using stochastic techniques. Moreover, MCMC based self-corrective scheme is placed which learns as the time evolves. The scheme continuously optimizes the supply chain with respect to time. The scheme will be applied on available historical (last eighteen months) as well as current database of M/s MHCO [3]. M/s MHCO steel historical as well as current databases would be used as a case-study in this thesis. M/s MHCO steel was established in 1975 and is certified from Karachi Chamber Of Commerce and Industry (KCCI). The net revenue of the company is between USD 2,000,000 – 5,000,000 with around six to ten employees. This organization usually deals in prime steel and secondary steel of different flat products including cold rolled coils (CRC), hot rolled coils (HRC), galvanized iron (GI), pre-painted GI (PPGI) etc. These flat products are imported from China generally and sold in domestic market as per consumer demand. The improved predictions will also be validated.

1.2 Bayesian Inference Based MCMC Methods for Supply Chain Optimization

Bayes' system is relatively simple. Initial belief + new data= Improved belief, which can also be termed as prior + likelihood = posterior [4]. Markov Chain Monte Carlo (MCMC) methods are used for sampling from a probability distribution built on making a Markov chain. A discrete state space defined is for **Markov chain** which is a type of Markov process. Markov process fulfil the Markov property i.e. the future predictions can be made of the process founded exclusively on its current state. The chain state after numerous steps is used as a sample of the chosen distribution. The sample quality gets better as a function of the steps.

1.3 Organization of Thesis

This thesis organizes as follows:

Chapter 2: Literature Review on Supply chain and MCMC Implementation

Chapter 3: Describes the Research Methodology. In this section the complete theory of Particle Filter algorithms is explained with necessary equations. It also introduces the goodness of fit of degradation models which are used for prediction. It analyzes the behavior of degradation model under one market factor which sets the direction for different factors selection

Chapter 4: Problem Statement and Case Study of uncertain demand data taken from MHCO Steel

Chapter 5: This section explains the case study of MHCO Steel demand / supply data & this section implements the proposed methodology on acquired data and displays the prediction fitness of distribution and present the results followed by discussion on results.

Chapter 6: This section summarizes the thesis and presents ideas for future work

Chapter 2 Literature Review

2.1 Supply Chain

Supply chain is a concatenation of operations from customer orders to distribution of products. Supply chain management for every organization is moderately different depending upon their size and operations. It may include activities such as sourcing, procurement, transportation, ware-housing, Inventory control forecasting, scheduling, production planning, customer services and order processing [5]. The proposed research includes critical factors affecting supply chain [6]. In supply chain, warehouse and inventory plays an important role. Basic warehouse functions comprise of receiving, storage, order picking, and shipping [7] . Sometimes the order of a product that has least demand in future is produced in a greater quantity, resulting in increase of inventory and its cost. This is the bull-whip effect [8]. In shipping there are delays studied by Tamang [7] [9]. There is an ongoing significant research on supply chain optimization. Recently, Artificial Intelligence (AI) and Decision Support System (DSS) are being used for decision making in optimization of supply chain [10] [12].

2.2 MCMC Implementation

The foremost step is **data preparation**. Second is the **prior inspection and integration** which includes consistency and credence test. Consistency check includes significance tests that helps in making statistical inference [13] [15]. Credibility of prior represents that the collected data belongs to same population. The total data values correspond to frequency distribution characterized by histogram [16]. Diagnosis of estimator bias is known as Bootstrap [17]. Monte Carlo theory of several iterations is also executed on the chosen sample [18].

Next, fusion of priors is a significant step. Fusion can either be parallel or serial [19]. The weighted average approach assigns weight to various priors. Rank based methods including Borda count which is defined as a voting system based on preferences [20]. Evidence based framework with uncertainty called Dampster-Shaffer theory [21], fuzzy Bayesian theory [22], probabilistic schemes and fusion through neural networks [23] are used for fusion of priors. Analytic Hierarchy Process (AHP) is another popular fusion method due to its ability of making decisions both on subjective and objective data with reduced bias [24].

The third step is the **selection of priors**. There are mainly two types of prior namely informative prior and non-informative prior. Conjugate prior is most popular informative prior [25]. The distributions in latter includes Bayes hypothesis [26], Jeffery's rule [27], reference prior, maximum entropy prior, probability matching prior, Monte Carlo method, relative likelihood approach, random weighting simulation method, inverse reference prior, cumulative distribution function, bootstrap method, and use of marginal distributions [25]. However, conjugate prior is a common prime. Besides conjugate, log concave prior is known for its flexibility [28] [29].

Model selection determines an accurate model (parametric or nonparametric), selecting from 'n' candidates for the studied system. One of the categories is parametric model which is the most commonly used statistical model. Bayesian exponential model is extensively being used. Bayesian extreme value model deals with the extreme deviations from the median of probability distributions. Bayesian Faulty tree analysis which starts with actual or potential failures and deducing what might have caused them. Bayesian Reliability growth model deals with the reliability changes over time. Bayesian accelerated failure model provides an alternative to the proportional hazards' models. Other models include Bayesian gamma model, Bayesian Weibull model and Bayesian log-normal [30].

The second category is semi-parametric models which includes some parametric and non-parametric components. These models offer more all-purpose modeling plan with fewer conventions [31]. Frailty models are used in cases where the data-sets may come from associated topics of the similar group, a key expansion in modeling this type of data is to create frailty models in which the data-set are conditionally self-governing [32].

Fourth step is **posterior sampling** which determines a sampling method to implement MCMC simulation. Particle filters [33] are sequential Monte Carlo methods centered on point mass (or "particle") depictions of probability densities by a set of random samples along with-it weights. The concept is to represent the Probability Density Function (PDF) by a set of random particles with its associated weights. With large number of samples, the Monte Carlo depiction becomes an equal illustration of the posterior probability function. Thus, obtaining optimal Bayesian estimate. In particular, Particle filter with Weibull modeling technique will be used for prediction. Weibull is known best for predicting failures [34] or in this case uncertainty.

For the Particle Filter, we present the Sequential Importance Sampling (SIS) procedure which includes a resampling step [35]. The SIS algorithm uses importance density, which is another proposal density representing another distribution that can't be exactly computed, that is, the required posterior density. Therefore, samples from importance density is drawn instead of the actual density.

Previously, several Monte Carlo (MC) based sampling methods have been established which includes Metropolis-Hastings sampling, MC importance sampling, Gibbs sampling, and other hybrid algorithms [36]. The other methods include sample importance re-sampling [37] rejection distribution [38] and adaptive rejection sampling.

The fundamental theory of MCMC ensures that the distribution of the output will converge to the posterior distribution as the number of samples increases to infinity. However, it doesn't ensure that the chain will converge after N iterations. The convergence problems can be addressed by applying any of the **MCMC convergence diagnostic tools** [39-52].

There are two common uncertainties under **MC error diagnostic** namely; statistical uncertainty which can be mitigated using maximum likelihood estimation (MLE) [39]and MC uncertainty which can be mitigated using standard error (SE) method [40]. Other error estimation techniques include Mean Squared Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD) & Mean absolute percent error (MAPE).

Model comparison is an important part of interpretation and its purpose is to select a better model after comparing with the other models. Comparison of model is divided into three wide categories. Distinct assessment including posterior predictive distribution, using methods that are Bayes' factors (BF) and its approximate estimation-Bayesian information criterion (BIC) and deviance information criterion (DIC) [41]. Comparative estimation including different distance measures: entropy distance, L-measure, Kullback-Leibler divergence (KLD) and weighted L-measure. Simultaneous estimation including birth-and-death MCMC (BDMCMC), reversible Jump MCMC (RJMCMC) and fusion MCMC (FMCMC).

The role of **model improvement** is to check the model against the prior knowledge and the other prior distributions. Posterior predictive checks are used to observe the matching of model prediction with the data, and if the parameter estimates (e.g. the differences between two groups) are consistent with the available prior knowledge. Prior distributions can also be changed, or more robust models can be used. **Data updating and inference** is a continuous step which evolves with the time.

Chapter 3 Research Methodology

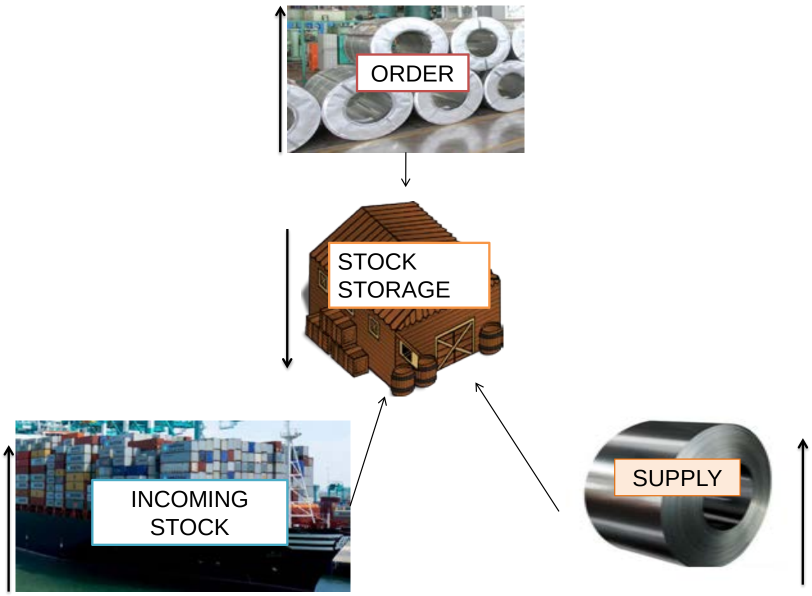


Figure 3-1 Big Picture

The main goal is to optimize the supply chain. It is possible by maintaining inventory up to a standard level to fulfill demands in order to produce a smooth supply to clients as displayed in Figure 3-1.

Statistical Bayesian inference using MCMC model will be used to optimize the inventory to maximize profit. With BC technology, decentralized approach will be followed to organize database. As mentioned above, the proposed project comprises of two parts:

First part is to construct an Enterprise System (ES).

Second part is to produce correct forecasting models for an Enterprise and update it along time with current data set. The whole flowchart is displayed in Figure 3-2.

From historical database, structures will be created for suppliers, customers and organization. This data includes cost and type of product, lead time, revenue generated, profit earned, inventory level at warehouse, product demand etc. This historical database will help in developing ES. All the mentioned information is embedded in the data itself. We can model on previous as well as on current data. Currently, on the previous data modeling will take place and will be validated through the current data.

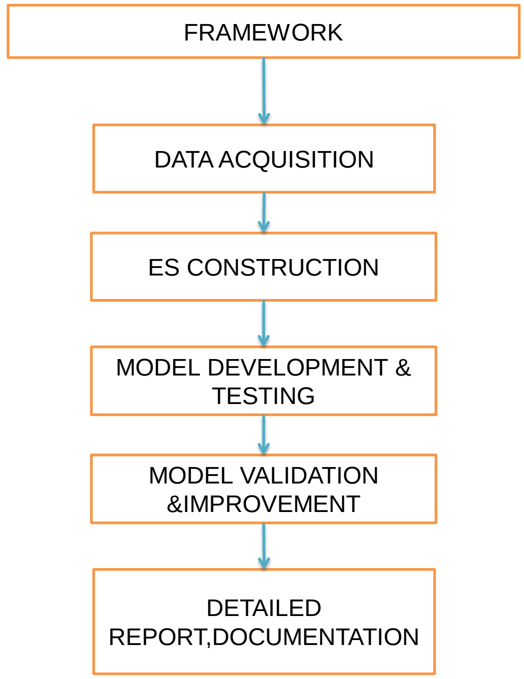


Figure 3-2 Proposed Research Flowchart

3.1 Bayesian Framework

Under Bayesian framework, shown in fig 7b optimization will be done in four steps, namely, plan, do, study and action (Figure 3-3).

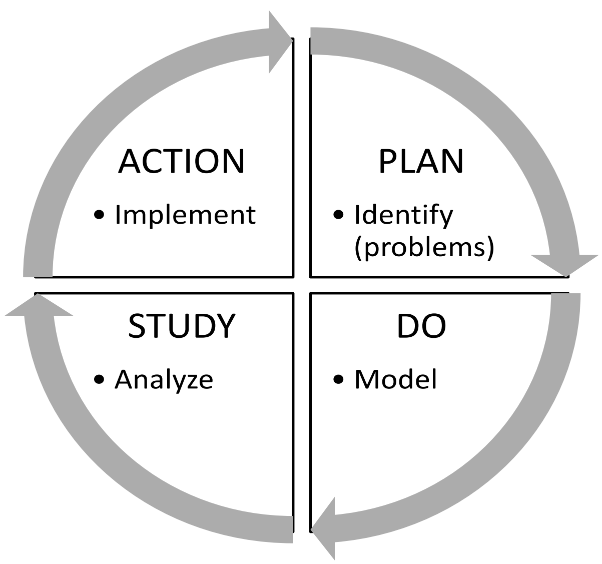


Figure 3-3 PDSA Cycle

'Plan' part formulates ES while preparing data. Afterwards, prior inspection and integration and prior selection occurs. **'Do'** part includes MCMC modeling. The **'Study'** part incorporates MCMC convergence diagnostic & Monte Carlo error diagnostic. The last part of **'Action'** includes inference making and data updating that evolves with time and the cycle continues as shown in Figure 3-4.

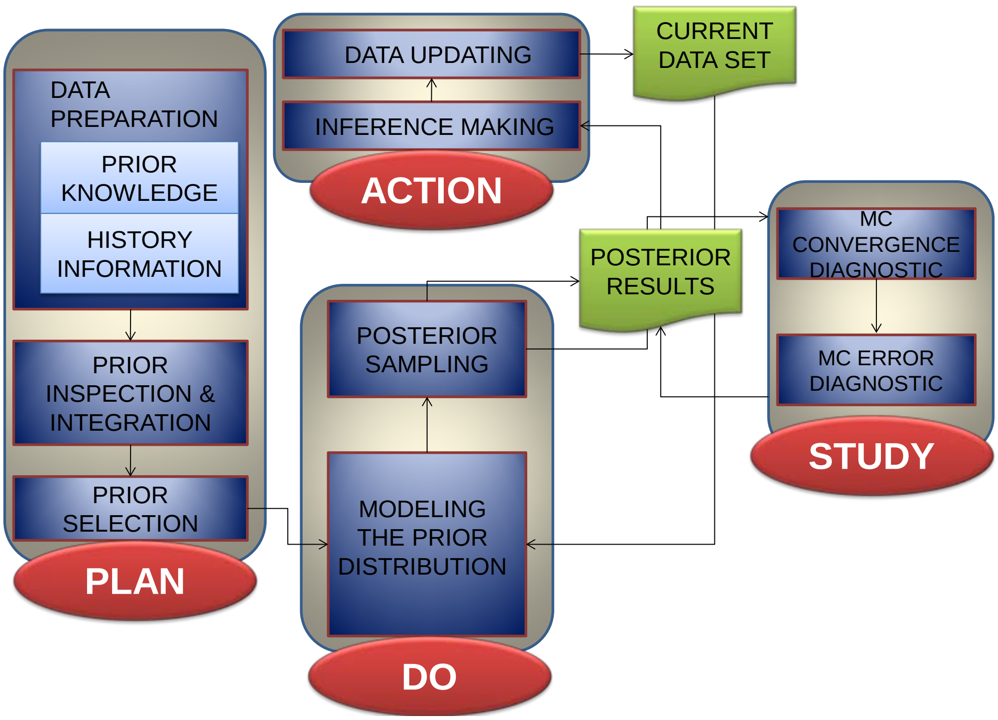


Figure 3-4 Generation of Enterprise BC under Bayesian framework [42]

3.1.1 Plan:

The plan is to develop an ES.

3.1.1.1 Enterprise System (ES):

An enterprise system is an application software package that incorporates different business areas such as gathering and analyzing data, reporting and executing the business procedures in complex organizations. With the gathered data, ES will be created. At this stage, everything is stored in a centralized database as shown in Figure 3-5.

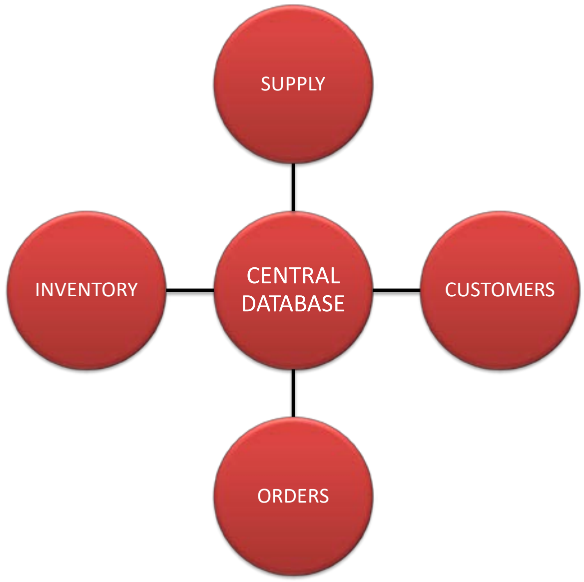


Figure 3-5 The Structure of a centralized Enterprise System

3.1.1.1.1 Data Preparation:

In data preparation, through prior formulation, ES will be constructed. For data preparation; organizational database & warehouse inventory database is required.

The second step is to perform consistency and credence test on priors leading to formulation of prior distribution.

3.1.1.1.1.1 Prior Formulation:

For the purpose of generating ES, priors are to be gathered. Prior formulation includes market demand, orders acquired and revenue generated. The formulation is elaborated in Figure 3-6.

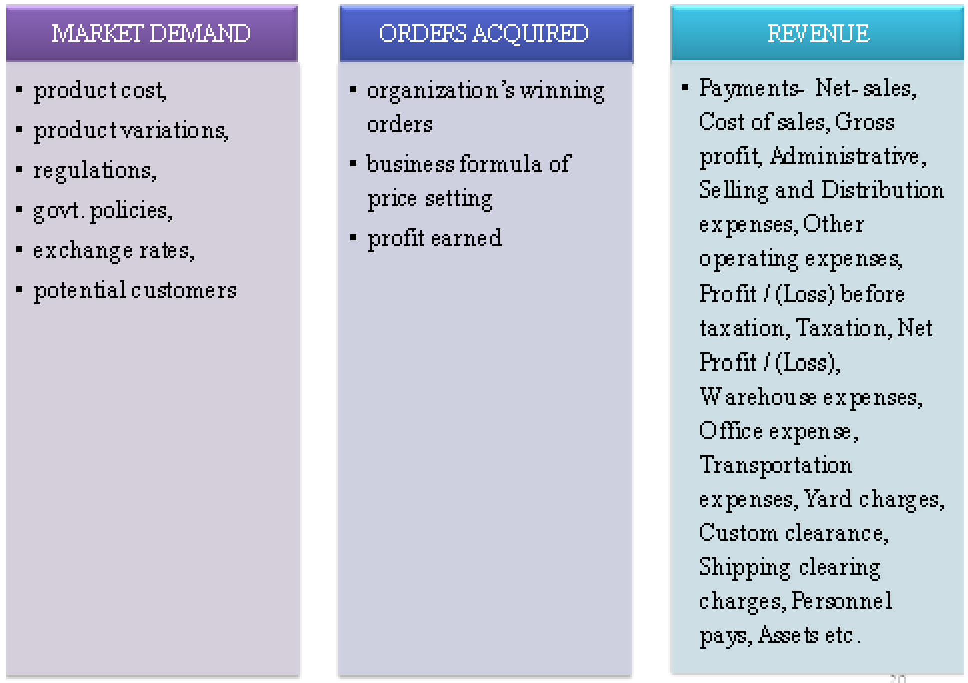


Figure 3-6 Details of Prior Formulation

3.1.1.1.1.1.1 Data Acquisition:

In this step, data would be acquired from different sources as tabulated in Table 3-1

Table 3-1 Data Acquisition Sources

TYPE OF DATA	SOURCE OF DATA
Orders, Revenue, Overhead costs, Profit, Potential customers, Payments	Company Database
Inventory, delivery records	Warehouse Database
Competitors revenue	Competitors firm
Government regulations and policies, exchange rates, International steel market prices, Opportunities	Internet surfing

3.1.1.1.1.1.2 History Information:

Available historical data does not contain all the ES information, as mentioned in ES section. Therefore, the historical database cannot be termed as ES. ES will only be formed in the current timeframe.

3.1.1.2 Prior Inspection and Integration

The next step is to inspect the priors. Inspection and fusion will finalize the prior. The tests for inspection and integration are discussed in literature review. The information gathered will be transferred to knowledge as shown in Figure 3-7 extracted from information.

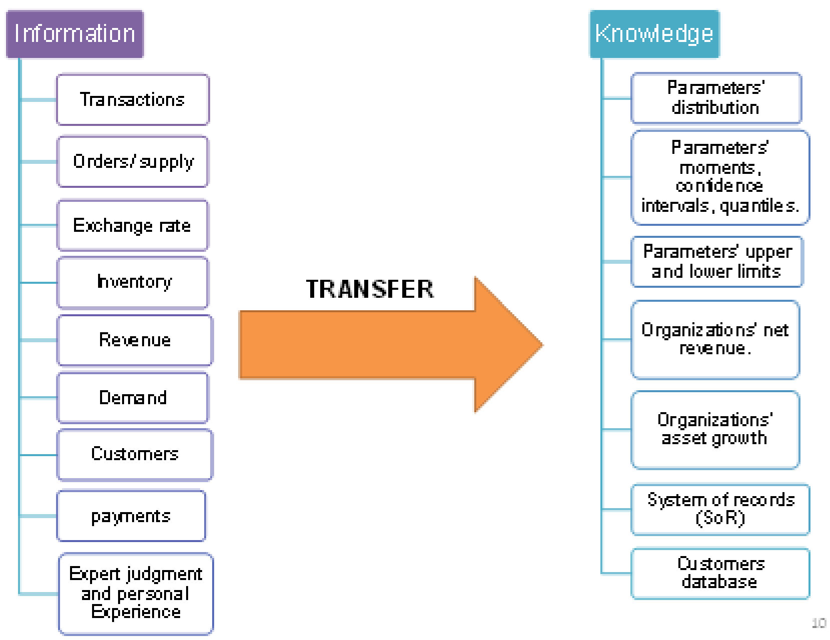


Figure 3-7 Information transfer to prior knowledge [42]

Gathering Prior

The prior gathered is basically the supply/ demand data which itself contains the history.

the following are the key aspects,

- MATLAB, an efficient and reliable software was used for coding
- The results are in graphical form, easy to grasp relevant information

- The Bayesian Inference MCMC technique was applied on the database
    - Market attribute was also incorporated by keeping 30% margin on upper and lower bounds leading to accurate results

3.1.2 Do

3.1.2.1 MCMC Model Development

To apply the MCMC methods, the models can be divided into four categories namely parametric, semi-parametric, frailty and non-traditional Bayesian models discussed in detail in literature review.

In the thesis, Parametric Modeling is used and fitted as it best replicates the real-world behavior. As parametric models are built from the mathematical equations set, changing one variable update the other variables itself. Therefore, it leads to the easiness of use.

Particle filter:

Linear measurement models or the Gaussian noise does not restrict the Particle Filter (PF) algorithms. The posterior Probability Density Function (PDF) can be estimated by PF by extrapolating the prior PDF. However, PF uses Sequential Importance Sampling (SIS) to determine the next state.

The uncertainty x_k is consider as state and z_k is the observation corresponding to the state where k is the current state index and K is the maximum index of available updates. Therefore, Kth order MCMC is used. All the observations for the past states $z_{1:k}$ must be known.

Via prior density $p(x_k \mid x_{k-1})$ the update function $p(z_k \mid x_k)$, one can determine the posterior density $p(x_k \mid z_{1:k})$ using the Bayes' Theorem as in 3-1,

$$p(x_k \mid z_{1:k}) = \frac{p(z_k \mid x_k).p(x_k \mid z_{1:k-1})}{p(z_k \mid z_{1:k-1})} \qquad \textbf{3-1}$$

Next step is to predict the future states using the predictive density $p(x_{K+F} \mid z_{1:k})$. The predictive density can be determined as:

$$p(x_{K+F} \mid z_{1:K}) = \int p(x_k \mid z_k).\prod_{F=K+1}^{K+F} p(x_F \mid x_{F-1})\, dx_F \qquad \textbf{3-2}$$

Where x_{K+F} represents the predicted state at ($K+F$) cycles. ('F' is the number of predicted states in future) The operation of PF executes in two steps i.e. Update and Propagate. The state transition model predicts the next state based on the information of the previous state. This predicted state is then updated using the available measurement.

The weights are updated only when the measurements are available. The weights corresponding to particles can be updated using the weight update equation 3-3 as:

$$w_k \propto w_{k-1} \frac{p(z_k \mid x_k)\, p(x_k \mid x_{k-1})}{q(x_k \mid x_{k-1}, z_k)} \qquad \textbf{3-3}$$

where $q(x_k \mid x_{k-1,} z_k)$ is the importance density [43]. In PF framework, the prior density can be chosen as the importance density. Therefore, the samples can be drawn from the prior density.

The Weibull distribution, named after mathematician Waloddi Weibull, is a continuous probability distribution. Weibull is generally used to evaluate reliability**,** model failure times, predict uncertainty and analyze life data. The three-parameter general Weibull distribution is [44]:

$$\boldsymbol{f(x) = \frac{\gamma}{\alpha} \frac{(x-\mu)}{\alpha}^{(\gamma-1)} exp^{(-((x-\mu)/\alpha)^{\gamma})} \qquad x \geq \mu\ ;\ \gamma\ , \alpha\ > 0,} \qquad 3\text{-}4$$

- γ is the shape parameter (*Weibull slope* or *the threshold parameter*).
- α is the scale parameter, (*characteristic life parameter)*.
- μ is the location parameter, (waiting time parameter or the shift parameter)

When μ = 0 and α = 1, the formula for the pdf reduces to:

$$f(x) = \gamma\ x^{(\gamma-1)} exp^{(-(x^{\gamma})} \qquad x \geq 0\ ;\ \gamma > 0, \qquad 3\text{-}5$$

which is the **standard Weibull distribution**.

The Weibull as a prior density $p(x_k|x_{k-1})$ [45] which is centered at the previous state as shown in equation 3-6:

$$= \gamma\ (x_k\text{-}x_{k\text{-}1})^{\ (\gamma\text{-}1)}\ e^{\ (\text{-}(x_k\text{-}x_{k\text{-}1})'\ \gamma)} \qquad 3\text{-}6$$

Where γ, is a shape parameter that controls the slope of the predicted state. Weights are assigned and updated using equation 3-3.

Observations using the actual observations & the measurement model are the inputs to the weight assignment block. The weight of a specific sample is small if the variance between the actual and the computed measurement is high or vice versa.

A suitable measurement model is required which transform the states into measurements. A linear measurement model is used in this work which the transforms the states into measurements with some Gaussian noise. The measurements are obtained from the actual states as in [45] using equation 3-7and 3-8.

$$z_k = f(x_k, n_k) \quad \textbf{3-7}$$

Using the database of the available states and their measurements, a Dth order polynomial is used to represent the measurement model as:

$$z_K = \sum_{d=1}^{D} c_d \, x^d \quad \textbf{3-8}$$

This measurement model is used to determine the measurements which are useful in updating the weights. The weights are updated till the measurements are available.

Modeling will be carried out for forecasting. The forecasting for the last month will be recorded in database. ES will be updated with current data set. Testing of the model will be carried out for the upcoming months and recorded in ES. Thus, maintaining the organization database. Hence, the cycle continues shown in Figure 3-8.

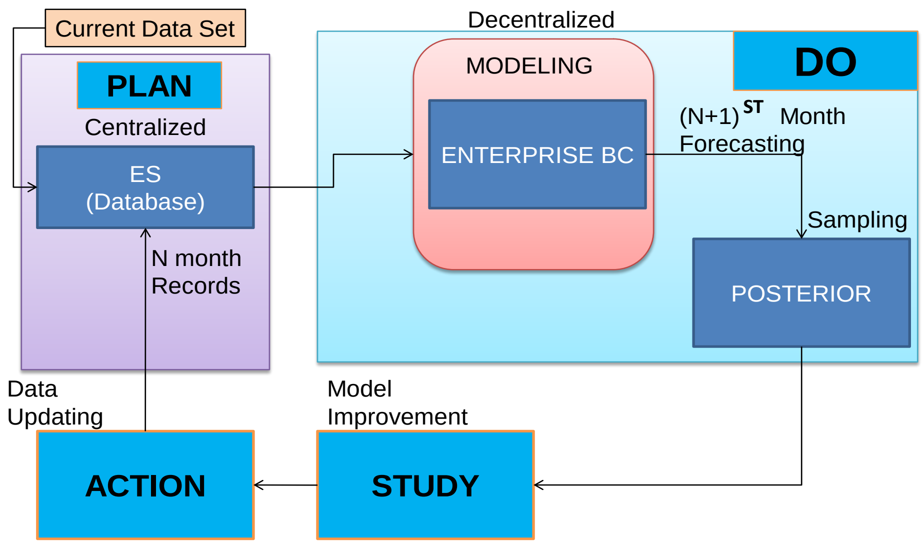


Figure 3-8 Modeling Prior

3.1.2.2 Posterior Sampling

Next step is posterior sampling in which a sampling method is determined to implement MCMC simulation through different techniques.

In Monte Carlo Simulation, thousand samples were drawn from the data to obtain maximum objective function. Sampling from a distribution p(x) and again using another uniform distribution q(x) for easy-to-sample proposal distribution q(x) that satisfies Equation 3-9

$$p(x) <= M\, q(x) \qquad \textbf{3-9}$$

using an accept/reject procedure.

3.1.3 Study

The study part shown in Figure 3-15 is the improvement of model by performing MCMC convergence diagnostic, MC error diagnostic, model comparison and finally making decision on improved model for an organization.

3.1.3.1 Error Reduction Scheme

For MCMC learning scheme a learning-based error model is proposed. The error is the difference between the predicted state and the actual state. In this the error is approximated in regard to the number of future estimate steps.

3.1.3.1.1 F' Step Prediction Error:

First, second and third step error is predicted. It is defined by F step Prediction error.

'F' step Prediction error, is the difference between predicted $(1+F)^{th}$ state and the actual $(1+F)^{th}$ state when first state is updated. When second state is updated, error is the difference between predicted $(2+F)^{th}$ state and the actual $(2+F)^{th}$ state. Similarly, the error as the third state is updated is the difference between predicted $(3+F)^{th}$ state and the actual $(3+F)^{th}$ state. On the similar manner, a series of 'F' step prediction error is generated till the update is available shown in equation 3-10.

$$E_{k,F} = \left(x_{k+F} - \hat{x}_{k+F} \right)\Big|_{k=1}^{K} \qquad \textbf{3-10}$$

where x is the actual state and $\hat{x}$ is the predicted state. ('k' is representing the number of updated states and 'F' is the number of future predicted state)

The general expression for calculating the error series is given under equation 3-11:

$$E_{k,i} = \left(x_{k+i} - \hat{x}_{k+i} \right) \Big|_{i=1}^{F} \Big|_{k=1}^{K} \tag{3-11}$$

Where,

F, is the number of future predicted states and

K, is the index at which the measurements are available for updating the weights.

3.1.3.1.2 Error Correction Factor:

As mentioned earlier, that the error is the difference between the actual and the predicted state, all the predicted states error series are determined correspondingly. The first, second and third error series are shown in Figure 3-9.

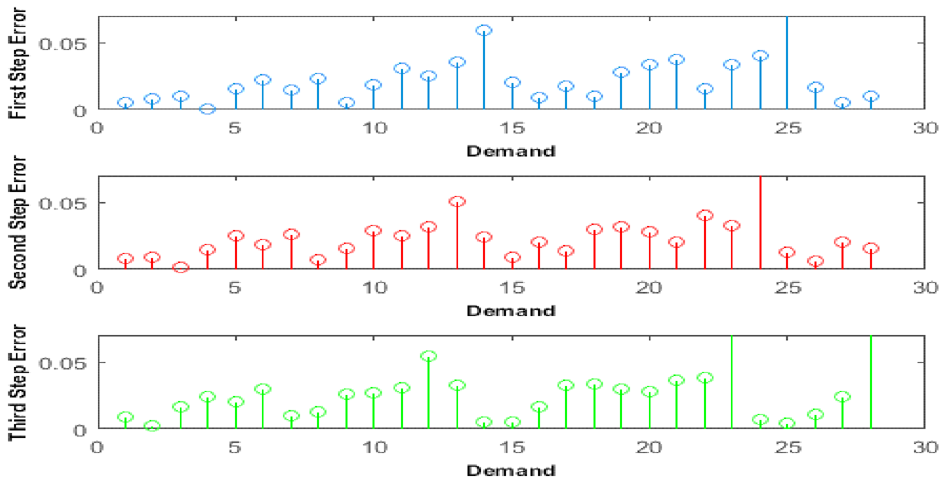


Figure 3-9 First, Second & Third Step Prediction Error

It has been observed that the predicted errors depend on the previous predicted state errors. Further, these errors reduce with increasing number of updated states. The difference between consecutive errors determine the behavior of the error slope which remarkably helps in the prediction of the next error.

The consecutive errors of each series are shown below in Figure 3-10, Figure 3-11 and Figure 3-12.

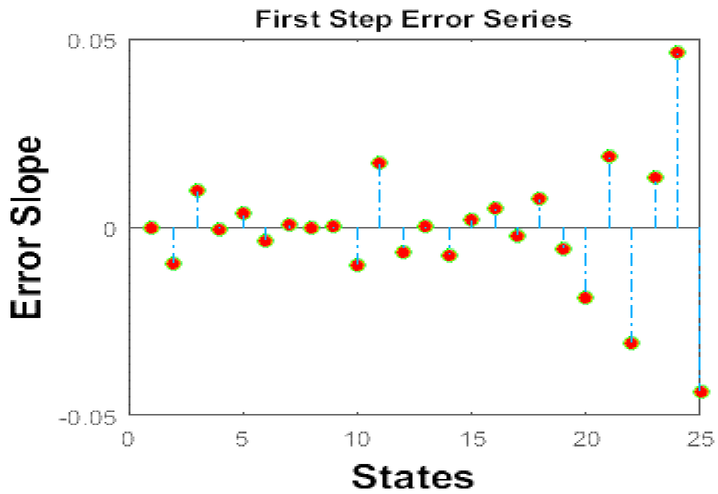


Figure 3-10 Error Slope of First Step Error Series

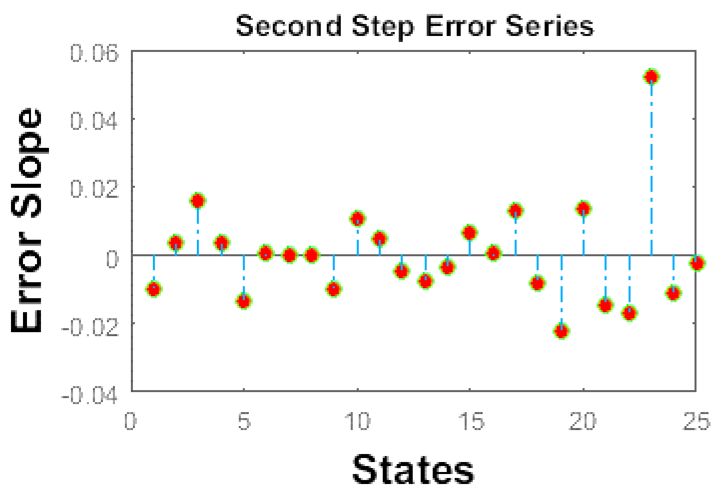


Figure 3-11 Error Slope of Second Step Error Series

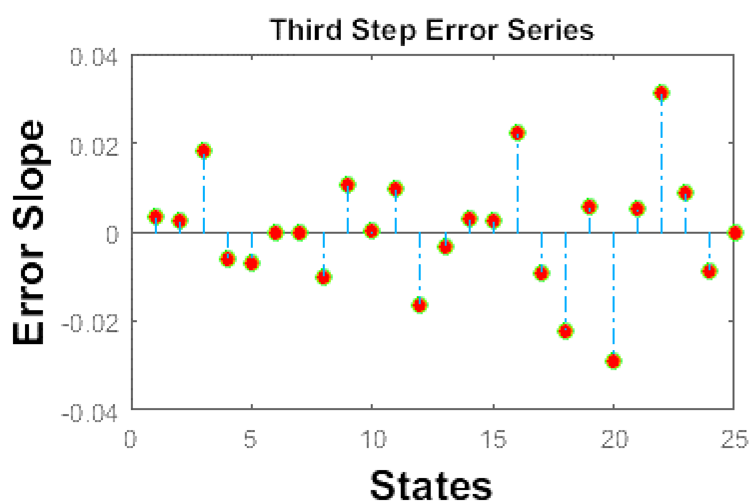


Figure 3-12 Error Slope of Third Step Error Series

By analyzing error slope figures, it is assumed that the error at every state is the mean of errors of all previous states. However, through keen observation one can say, that, the error mainly depends on the last two or three error readings. Therefore, for first step prediction, two error slopes are used. The next state error slope is calculated by taking the mean of the previous two error slopes as given in equation 3-12:

$$err_{s(j+1)} = \frac{err_{s(j)} + err_{s(j-1)}}{2} \qquad \textbf{3-12}$$

Where,

$err_{s(j+1)}$ is the next state error slope,

j is the index of the current state

The error correction factor can be calculated by adding the calculated error slope into the error at the current demand *K*:

$$e_i = E_K + err_{s(j+1)} \qquad \textbf{3-13}$$

The prediction results can be improved using these error correction factors as:

$$\bar{x}_{K+i} = \hat{x}_{K+i} + e_i \qquad \textbf{3-14}$$

Where, $\bar{x}$ is the corrected state and *i* is the index of the predicted state.

As, the study part signifies the convergence. The above used method implies that the error converges with the provided data. It also shows that it will converge as time evolves.

3.1.4 Action

In last step, inference will be made and data will be updated in ES database with respect to time as shown in Figure 3-13.

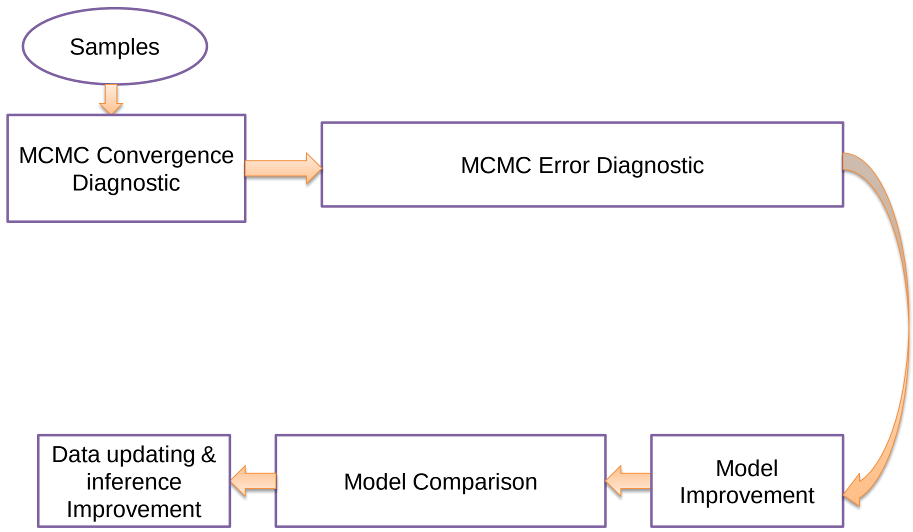


Figure 3-13 Modeling the Posterior

Chapter 4 Case Study

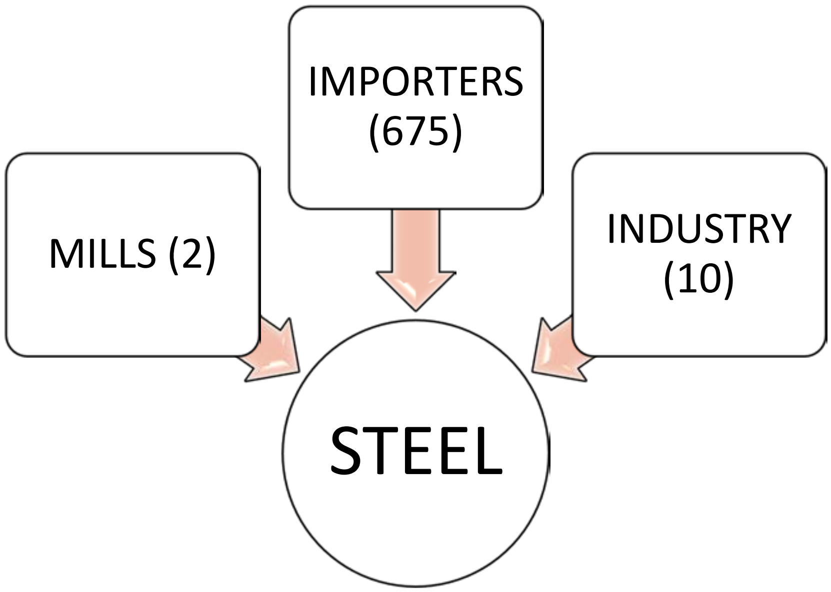
4.1 Problem Statement

Improved supply chain management has gained immense significance in the recent years because of the associated countless benefits. Supply Chain management is a challenging task in today's world due to the encountered uncertainties. In Pakistan, forecasting in supply chain is more challenging due to **smaller industry base, longer lead times, limited client tail, political instability, lack of warehousing facilities and the smaller market quantum**. **Truncated/censored/incomplete historical data**, a situation widespread in Pakistani supply chains, is also a major impediment in accurate demand forecasting.

The existing system of transactions record keeping, in public as well as private sectors' supply chains, is prone to tempering as well as obliteration at different hierarchal levels. The traceability of financial as well as material transactions in the existing supply chains also requires improvement.

4.2 Historical Data

The steel in Pakistan is mainly produced by mills, industry and importers. There are mainly two mills and few industries. In Karachi there are almost 675 importers that are dealing in steel as indicated in Figure 4-1.



.

Figure 4-1 Steel Consumption

In Pakistan, the need of steel is either fulfilled by mills, industry or imports. The raw material is itself imported by China. The price rates are influenced by Chinese market. The production of flat steel products did not meet the capacity. About 77% of flat steel products are imported. The imports ratio is greater than that of production [46] as shown in Figure 4-2.

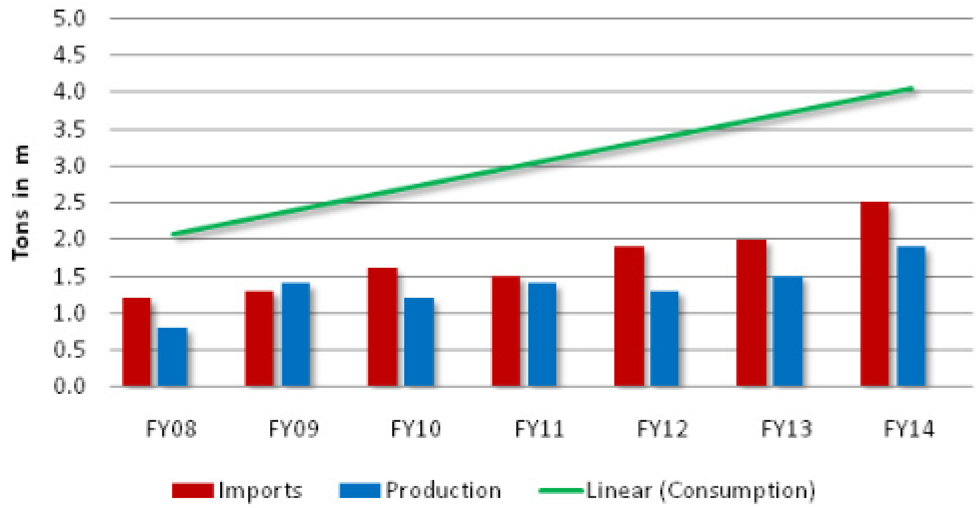


Figure 4-2 Domestic Steel Production and Imports [46]

In 2015, as report states, the yearly demand in Pakistan is approximately 8.4 million tons, while the present production is about 4.9 million tons per annum [47] as shown in Figure 4-3.

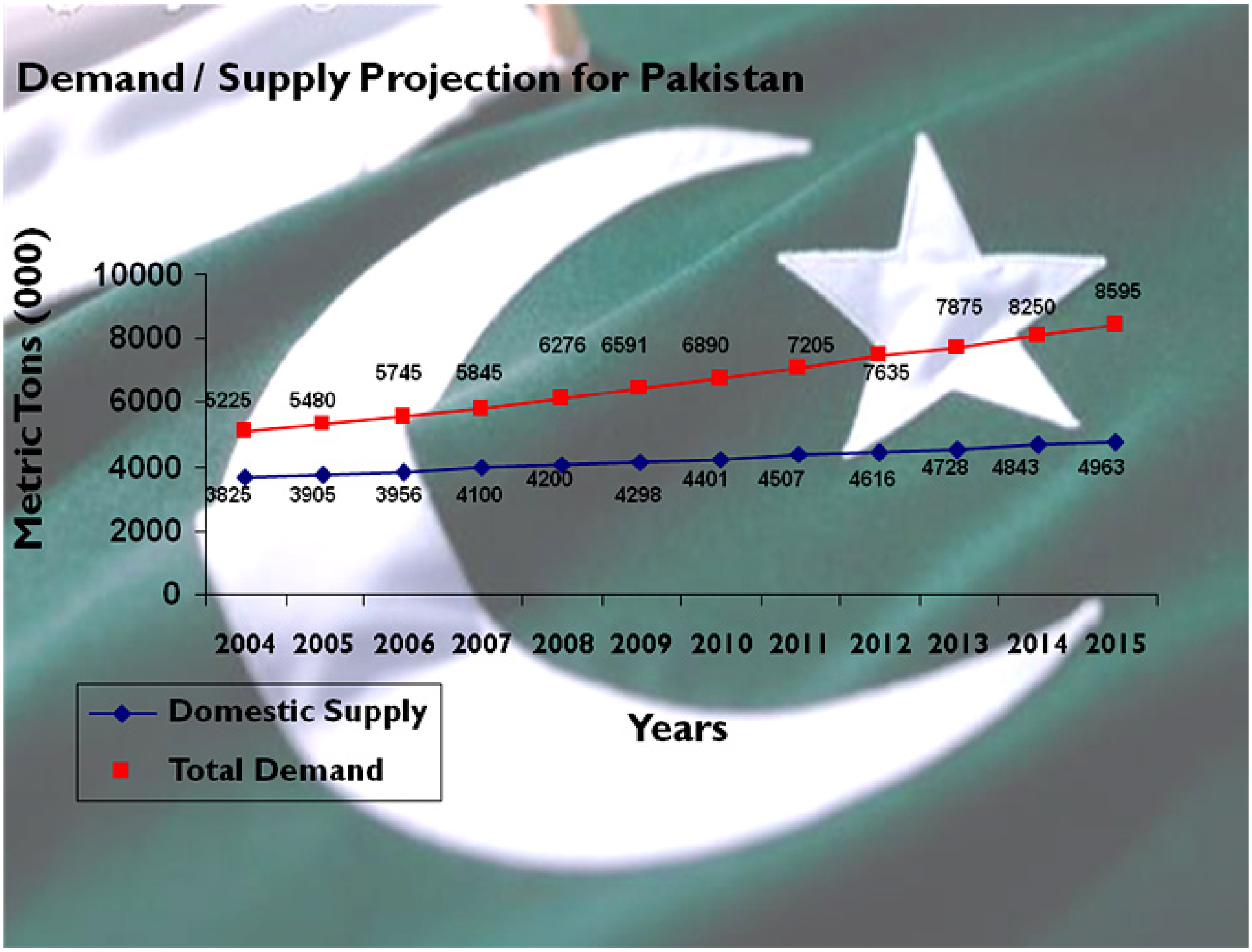


Figure 4-3 Steel Supply/ Demand [47]

Moreover, the uncertainty can be seen in the production of steel in Figure 4-4. The supply/demand data of one of the importers, namely, MHCO Steel has been captured to learn and predict the uncertainty as they are old in the market. Their data set or prior will be an asset to predict the future demand.

Figure 4-4 Global Crude Steel Production [46]

4.3 Types of Steel & Applications

There are many types of steel. Some are listed below:

- **Hot Roll (HR)**: HR used in welding, Pipes, Tanks, Floating Structures, Wheel Rims, construction trade to make railroad I-beams & tracks. For no precise shapes and tolerances requirements HR is used in such situations.
- **Cold Roll (CR)**: Cold rolled sheets are used for the production of enamelled wares, bicycles, steel containers, machinery parts, steel fabrication, drums, jerry-cans, barrels, vehicle and bus bodies, steel furniture, sand appliances, oil and gas appliances, also used for the production of galvanized sheets, and tin plates and black plates.
- **Galvanized Iron (GI)**: widely utilized in rural and military buildings such as sheds and water tanks and for pneumatic roofing guns.

- **Pre-Painted Galvanized Iron (PPGI)**: utilized in architectural adornment and outward beautification of piano, several domestic appliances and electrical appliances
- **Pattern PPGI**: Wood grain pattern PPGI for furniture, Ship plate, boiler plate, container plate.

As Government encourages industrialization. They place anti-dumping on selected products for imports. They have applied anti-dumping on Cold Roll (CR) and Hot Roll (HR) which are of most use. It is processed and use by industries to produce bars, billets and pipes.

HR is the raw material which was only produced by Steel Mills which is now only imported. Currently, no raw steel production is done in country. The two mills, namely, Aisha Mills [48] and International Steel limited (ISL) [49] are producing specific sizes and specific amounts of other flat products. They do not even fulfill half of the capacity.

There are many types of steel. In this research flat products are targeted. Moreover, data of only PPGI is taken. There is no anti-dumping on this product and less exploitation. Through this product we can accurately capture the market trends.

The demand data along with respected months in which orders are placed.

Chapter 5 Results

The discussed methodology is then applied. The first step was to select the prior from the data set. The data was taken from an import/export company. The supply/demand data which is considered same at this time is opted for busting uncertainty. There are thirty-two data points altogether signifying 2.5 years data. Different assessments were made by varying training data and validation data. As more data entered the training data, more accurate results were obtained. Histogram was made by the data and three different distributions were made fit. From which Weibull duplicates the histogram well. Furthermore, goodness of fit also implies Weibull to be the best fit for this data-set.

5.1 Fitting distribution using MLE (Maximum Likelihood Estimate)

Different distributions named Weibull, Lognormal and Normal were fitted on the acquired data

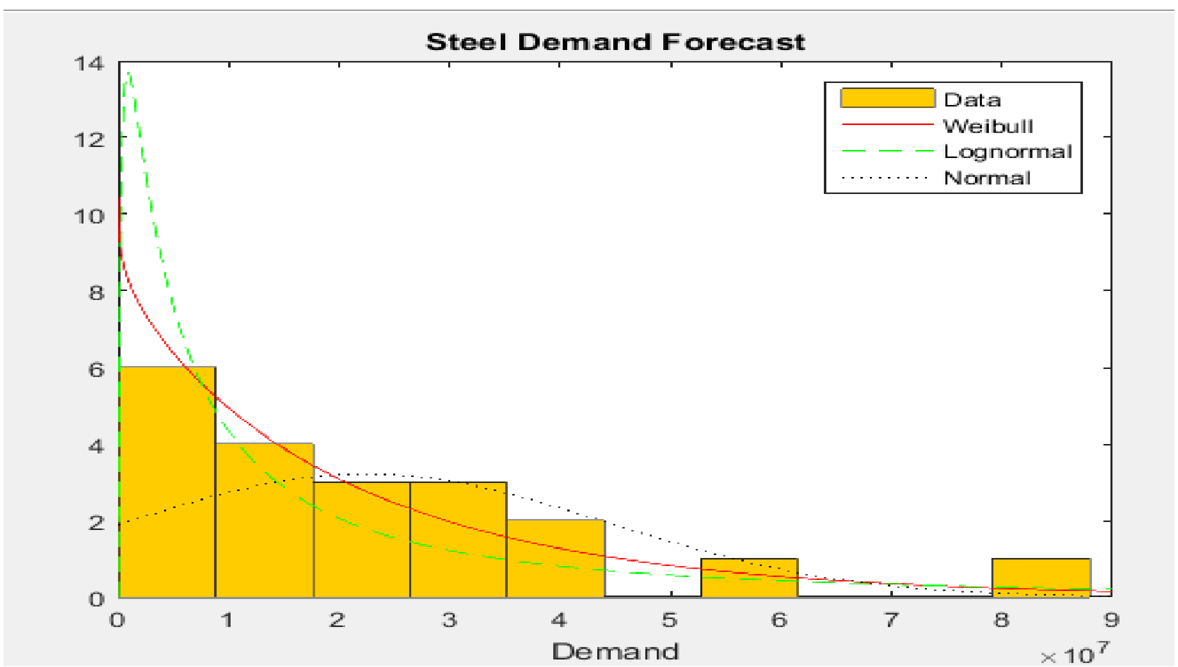


Figure 5-1 Distribution Fit

5.2 Goodness of Fit

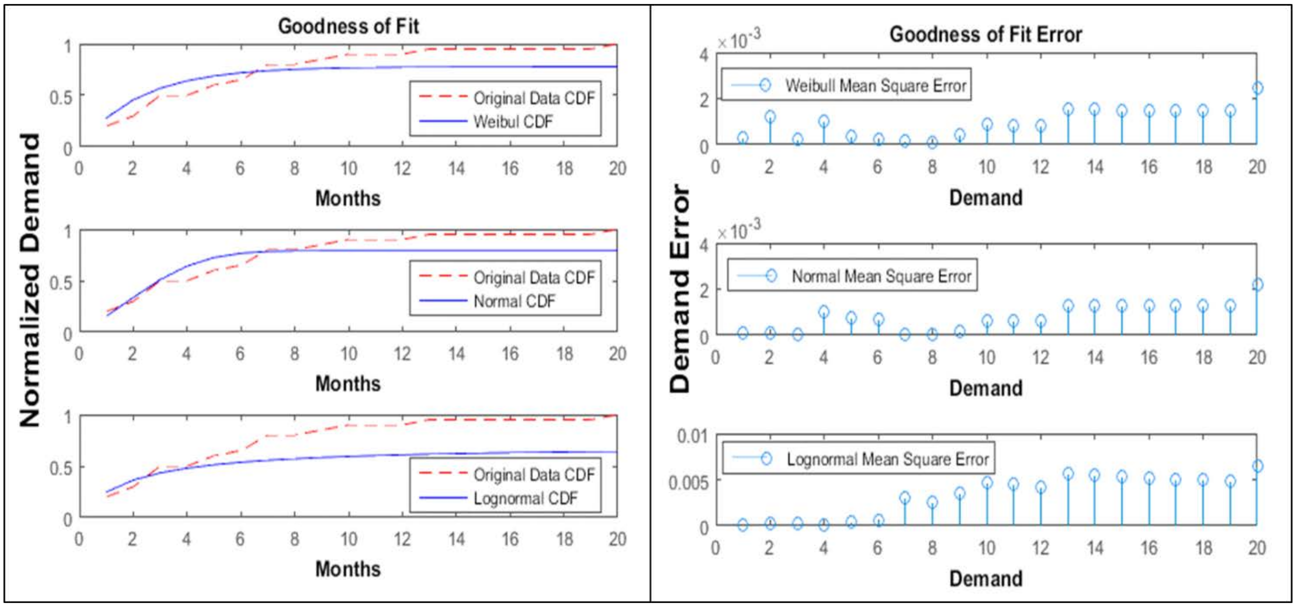


Figure 5-2 Goodness of Fit

Best fitted distribution was ***Weibull,*** evident from Figure 5-1 and Figure 5-2***.*** ***Maximum likelihood estimate*** of A & B parameters with 95% confidence Intervals were,

A (Scale Parameter) = 2.20965e+07 [1.36995e+07, 3.56406e+07]

B (Shape Parameter) = 0.959145 [0.673832, 1.36526]

The next step was to predict next state which was done through Rejection Sampling and Maximum Likelihood Estimation (MLE) from which Mean and Maximum A posterior probability (MAP) was estimated indicating the next state.

5.3 Mean and Maximum Aposteriori probability (MAP)

The **Mean** & **maximum Aposteriori probability** (**MAP**) was taken of these samples and posterior was generated. This was helpful in taking accurate decision for predicting next demand.

5.3.1 Assessment # 01

Provide twenty months training data (20 data points) and predict next five months demand, as shown in Figure 5-3,

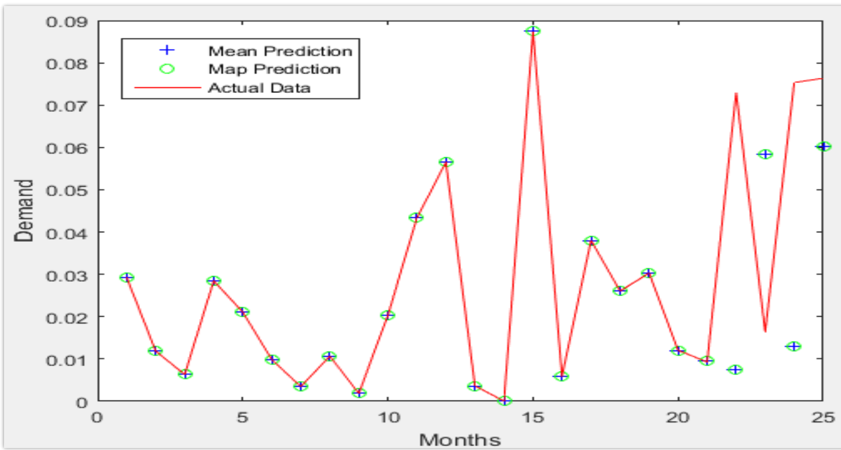


Figure 5-3 Predicted vs Actual demand (20 training points)

5.3.2 Assessment # 02

Provide two years, 24 training data points and predict next five years demand, as shown in Figure 5-4,

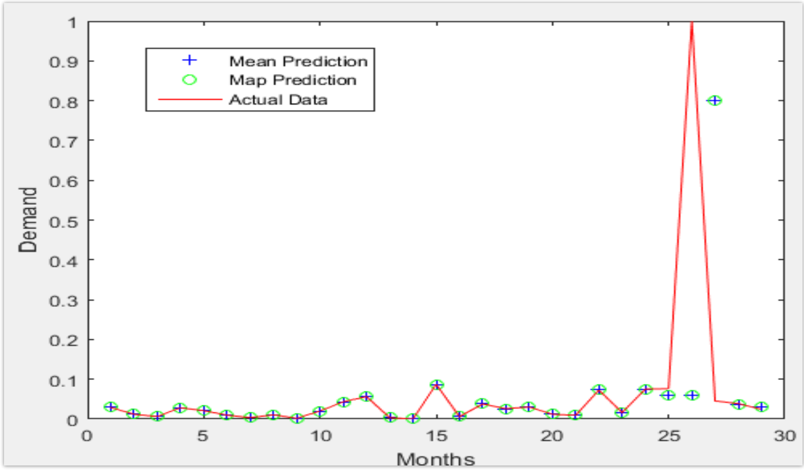


Figure 5-4 Predicted vs Actual demand (24 training points)

5.3.3 Assessment # 03

Provide two and half years, 27 training data points and predict next five years demand, as shown in

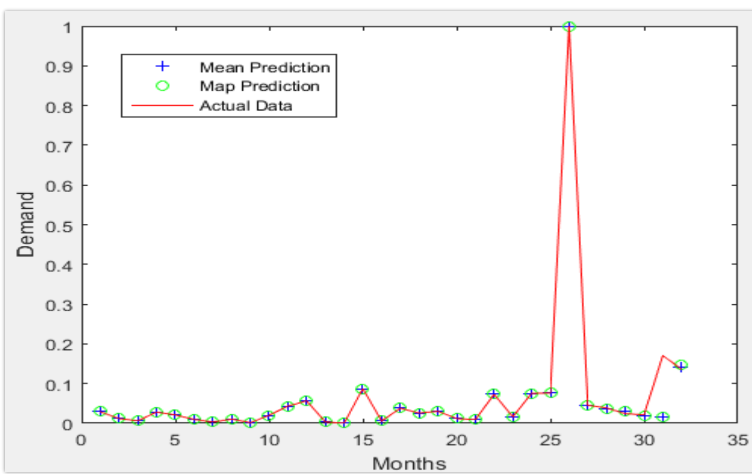


Figure 5-5 Predicted vs Actual Demand (27 training points)

5.4 Errors:

By the passage of time, as there is increase in data more accurate results are acquired.

Two types of errors are used in this research namely,

- Mean Error (ME)
- Root Mean Square Error (RMSE)

It is justified through these errors countered in each assessment as shown in figures respectively.

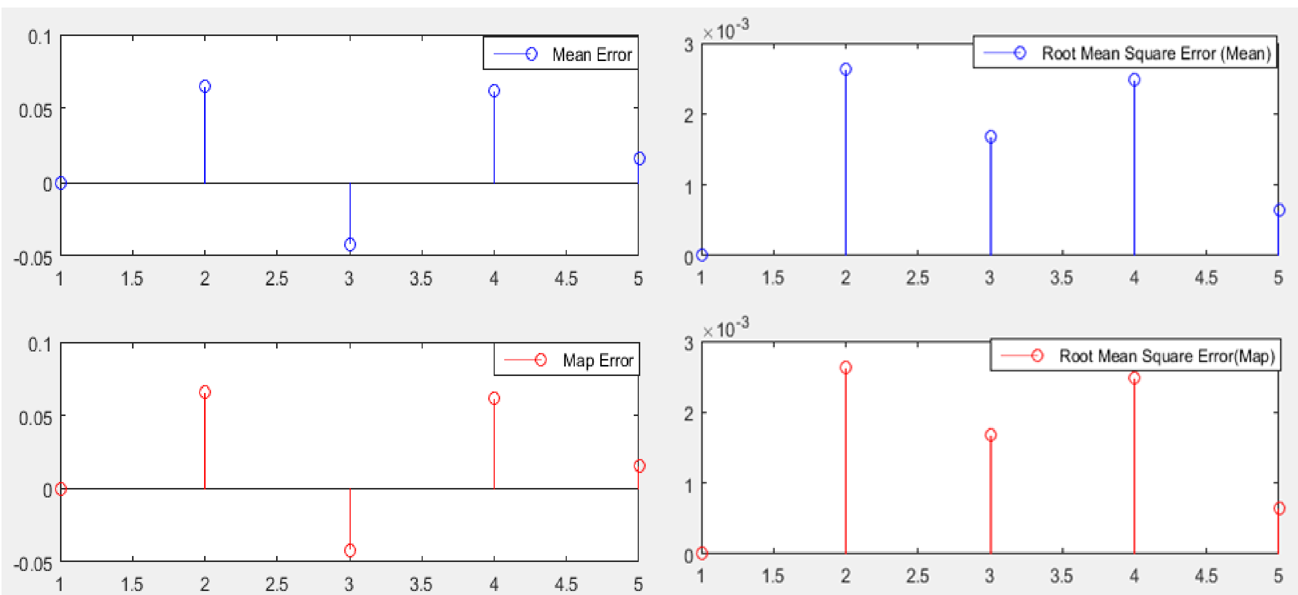


Figure 5-6 Error of Predicted vs Actual demand (20 training points)

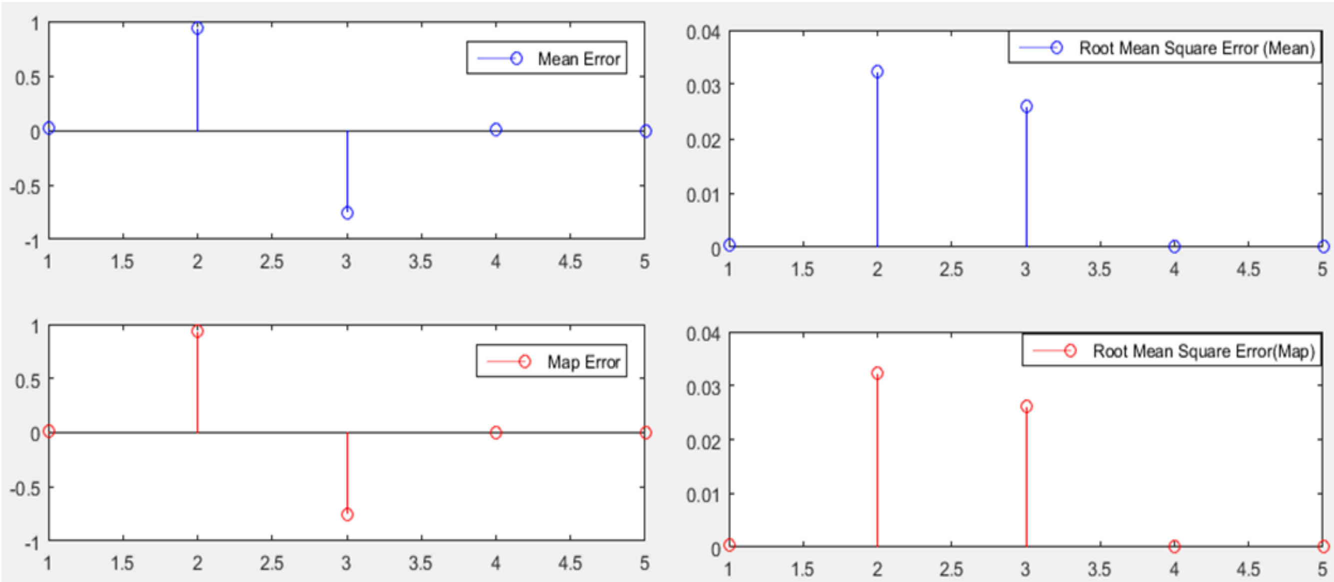


Figure 5-7 Error of Predicted vs Actual demand (24 training points)

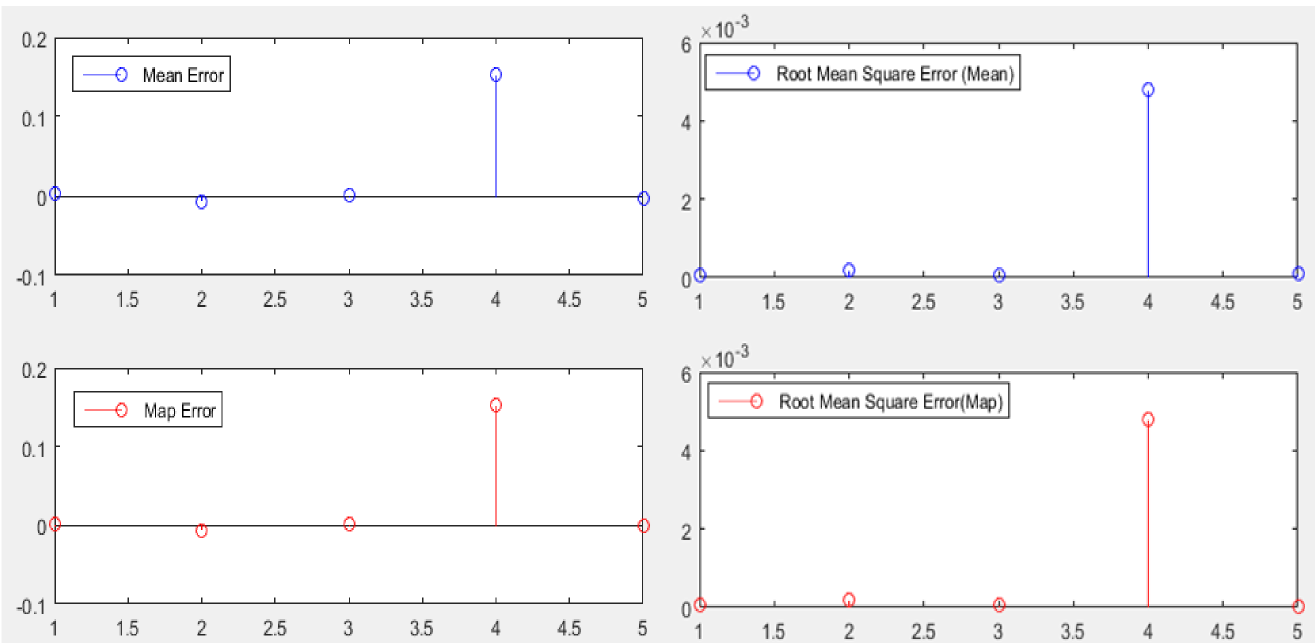


Figure 5-8 Error of Predicted vs Actual Demand (27 training points)

5.5 Particle Filter Results

Moving one step further, particle filter is an advanced technique which itself incorporates MLE. Using Particle filter on the demand data-set one can easily obtain predicted as well as corrected states. It makes easy for decision makers to take right decisions to meet the next demand.

As enough data was available for plotting so there was no need to interpolate the data points. But if less data is available Cubic Spline interpolation can be used to fit curve as shown in Figure 5-9

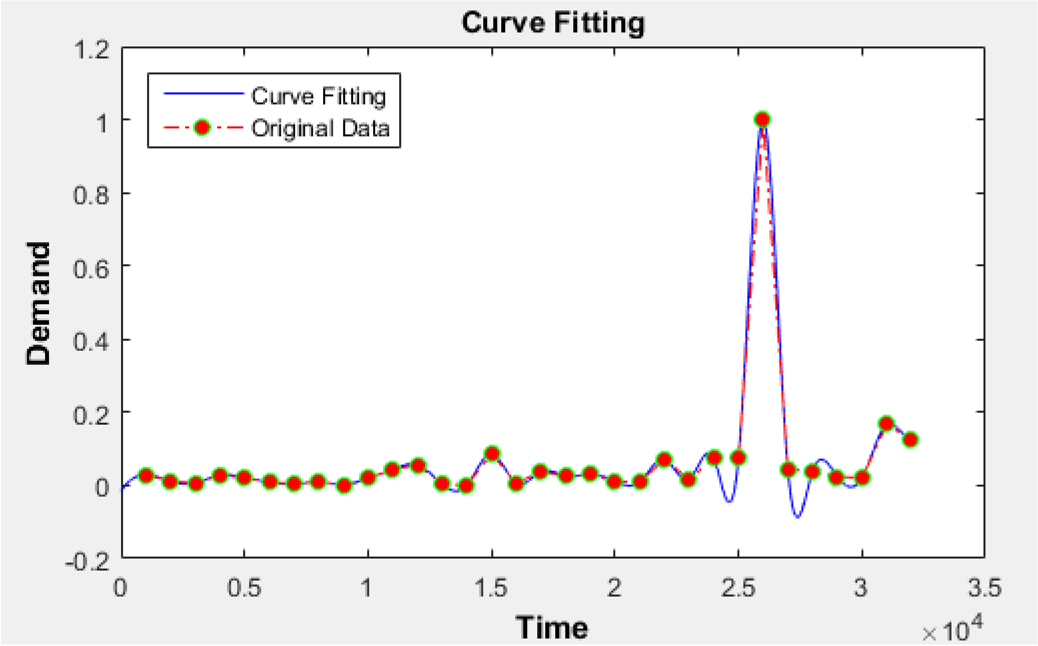


Figure 5-9 Curve Fitting

Out of thirty-two points, twenty-eight were used for training. The rest points were used for validation. The predicted states are given as

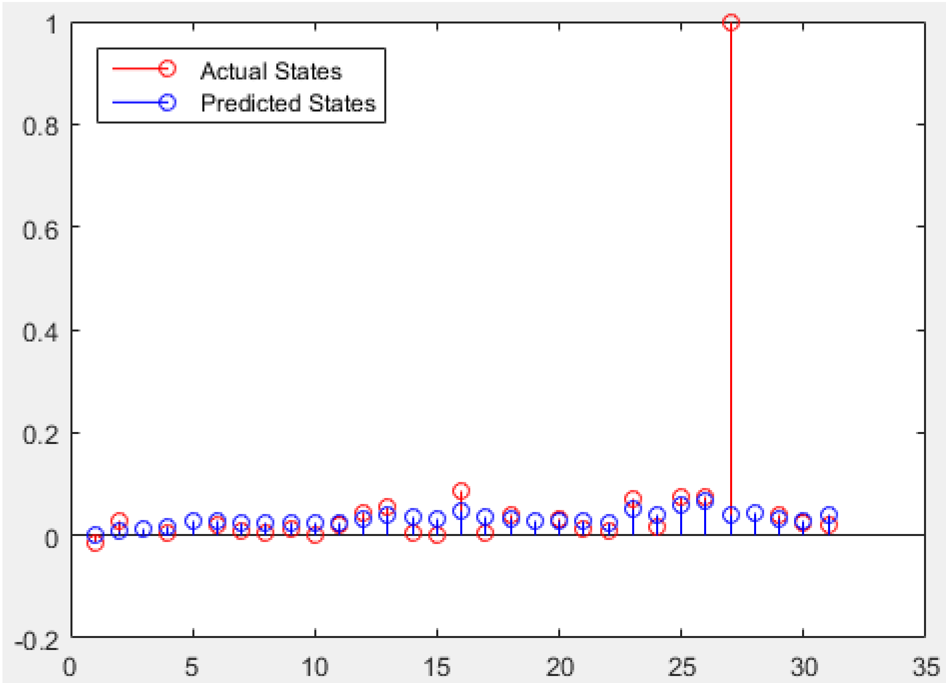


Figure 5-10 Predicted States

Different scenarios were created to find the best fitting. The parameters that were used includes:

- Market alteration
- Nth Order Polynomial
- Random Sampling

<u>Market Alteration</u>:

The data that is selected, obviously contains market traits. It was observed that considering this factor of state tremendously effects the results. Different rates were experimented. After which it is kept constant at 30% upper and lower state as the best results were acquired.

<u>Nth Order Polynomial</u>:

The second factor was of the Polyfit fitting. As PF states, functional model is the combination of data and noise. Therefore, noise was added to the data. Afterwards, the important step was to approximate function. Curve fitting is one of the very useful tools for creating such a function.

$$\mathbf{Dt_2 = f(Dt_1)} \qquad \mathbf{5\text{-}1}$$

Where,

Dt_2= Approximate Data

Dt_1=Original data

Therefore, Curve fitting was used to approximate the functional model corresponding to time scale.

As the data is non-linear as shown in Figure 5-9 this factor too plays an active role in order to achieve best fitting. This is altered on 3rd, 4th & 5th order.

Random Sampling:

While doing the rejection sampling, many samples or infinite samples have to be drawn. This is the key feature of Monte Carlo Sampling. The equation signifies that altering the parameter greatly affects the results. This was kept constant on 0.8 as in equation.

$$\mathbf{Y=0.8 * R\ (1) * (1\text{-}xk)} \qquad \mathbf{5\text{-}2}$$

Where, R is the random number generator,

xk is the state at time step k

RESULTS:

CASE 1**:3rd Order Polynomial**

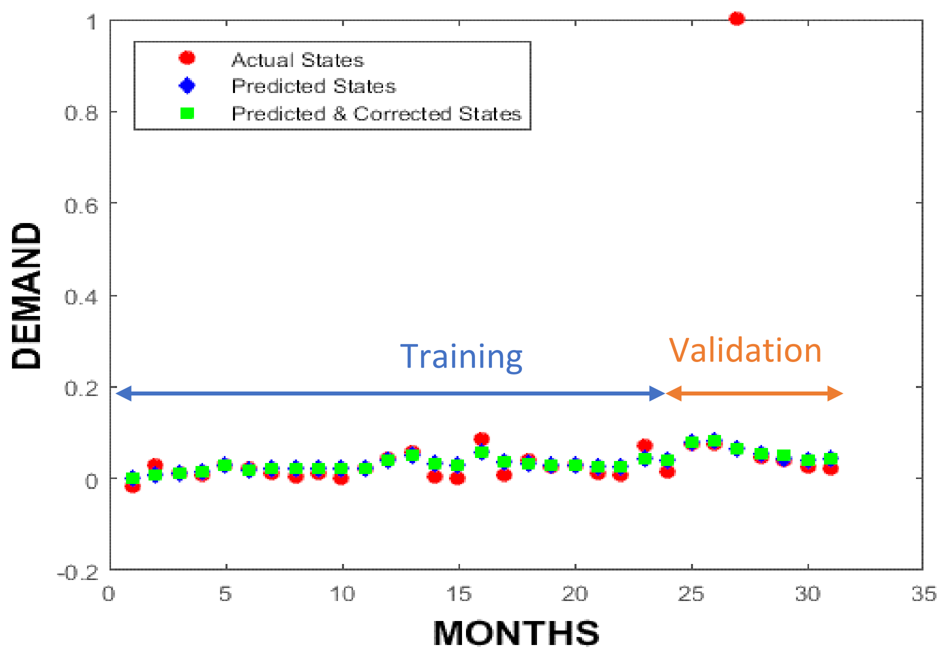


Figure 5-11 Forecasting Demand at 3rd Order Polynomial

CASE 2: **4th Order Polynomial**

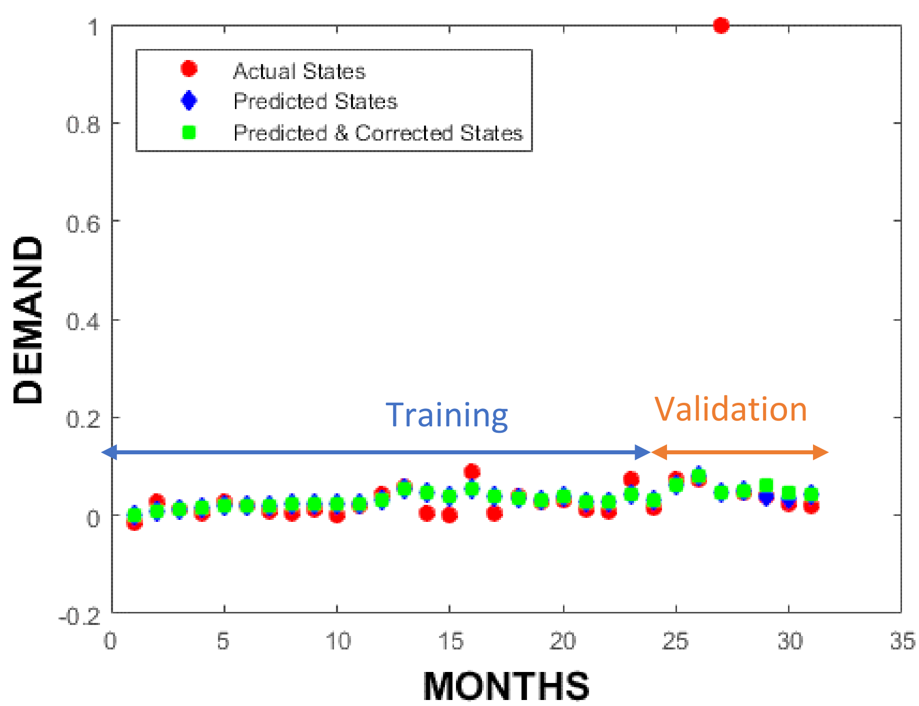


Figure 5-12 Forecasting Demand at 4th Order Polynomial

***CASE 3:* 5th Order Polynomial**

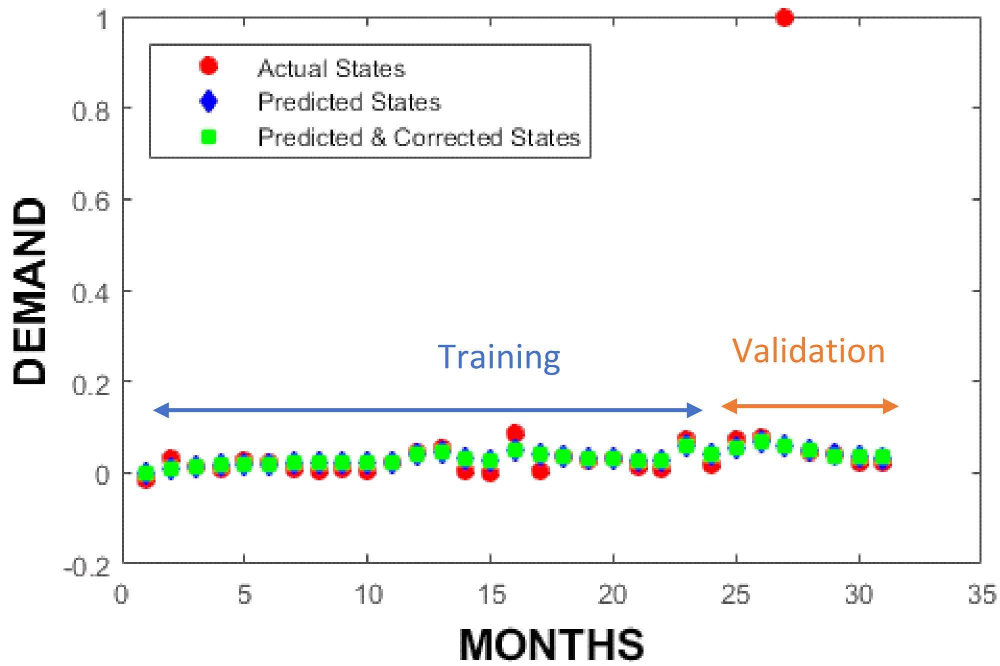


Figure 5-13 Forecasting Demand at 5th Order Polynomial

By keeping two factors constant and one variable we achieved the results shown in Figure 5-11 Forecasting Demand at 3rd Order Polynomial Figure 5-11, Figure 5-12 and Figure 5-13.

The 3rd order polynomial reflects the best results so far. As 5th order and 3rd are almost same. Therefore, 3rd order is opted for this research.

Chapter 6 Conclusions & Discussions

This research produces very meaningful results in terms of forecasting & prediction. This is a very strong project producing very powerful results. The techniques used can be use on any data set. Especially, it can be use on any current supply chain business. It can be applied on any individual sector governed under umbrella of supply chain. This project will not only predict the near future but will also help businesses lead the future.

There are many parameters involved in any business which includes international market steel prices fluctuation, commodity prices fluctuation, US Dollar rate fluctuation, shipment costs, warehouse rents, law and order situation, fluctuating Government regulations, fluctuating custom clearance regulation/policies/ rates, customer of commodity shift related policies, taxation regimes.

Countering these parameters one can accurately forecast the future. In this research, such form of one parameter is taken into account which is the fluctuating market parameter. The scale is set taking account 30% ups and downs or fluctuating rate into account and then made predictions. Similarly, there are other important parameters involved which can be taken as the future work for this research.

Furthermore, such a database is required which is temper-proof and uncensored. To achieve this objective this project can move towards encountering Blockchain feature into it. The Blockchain as considered new internet for the coming generation. The decentralization feature of BC can be well used in this project. The un-gagged, un-censored and un-tempered database can be achieved producing more sounding results. As new data adds on with the previous data the database increases. To manage database intelligently without any data tempering Blockchain feature will add accuracy in the results. It brings transparency with privacy which is much important in this era. Taking blockchain along with MCMC techniques it becomes a very influential tool for any kind of supply chain business.

References

[1] P. E. H. a. B. T. M. Neeta P. Fogga, "The impact of the Great Recession upon the unemployment of Americans with disabilities," *Journal of Vocational Rehabilitation ,* pp. 193-202, 2010.

[2] "Investopedia," [Online]. Available: https://www.investopedia.com/terms/g/great-recession.asp.

[3] "MHCO Steel," Import & Export, 1975. [Online]. Available: http://www.mhcosteel.com/.

[4] lukeprog, "A History of Bayes' Theorem. lesswrong. [Online].," 2011. [Online]. Available: http://lesswrong.com/lw/774/a_history_of_bayes_theorem/..

[5] G. S. D. G. A. S. B. J. Jinesh Jain, "Supply Chain Management: Literature Review and Some Issues," *Journal of Studies on Manufacturing,* vol. Vol. 1., 2010.

[6] R. G. a. S. S. [. a. D. A. G. s. :. I. Henry Quesada, "Critical Factors Affecting Supply Chain Management: A Case Study in the US Pallet Industry.," *Pathways to Supply Chain Excellence.,* 2012.

[7] M. G. ,. L. F. M. Jinxiang Gu, "Research on warehouse operation: A comprehensive review," *European Journal of Operational Research: Elsevier,* pp. 1-21, 2006.

[8] Z. D. J. K. R. D. S.-L. Frank Chen, "Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead Times, and Information.," *Management Science,* pp. 436-443, 2000.

[9] P. L. a. P. P. K. Tamang, "A DBMS based inventory model and its timeframe study in automobile spare parts import management.," *IEEE,* 2015.

[10] S. P. Validi, "Intelligent Decision Support Systems in Supply Chain Management," *European Conference on Operational Research,* 2016.

[11] C. Glock, "Decision support models for managing returnable transport items in supply chains: A systematic literature review," *Elsevier,* vol. Vol. 183., 2017.

[12] C.-S. L. J. J. H. P. B. M. Chung-Shan Yanga, "The effect of green supply chain management on green performance and firm competitiveness in the context of container shipping in Taiwan," *Elsevier,* vol. Vol. 55, 2013.

[13] "Tests of Significance. Statistical Topics," 1997. [Online]. Available: http://www.stat.yale.edu/Courses/1997-98/101/sigtest.htm..

[14] "Kolmogorov-Smirnov Test," [Online]. Available: http://www.physics.csbsju.edu/stats/KS-test.html..

[15] D. K. Neal, "rank-sum test. Statistical Inference on the TI-83+/86/89," 2003 . [Online]. Available: http://people.wku.edu/david.neal/statistics/nonparametric/ranksum.html..

[16] "Understanding the Uses for Mood's Median Test. iSixSigma.," 2000. [Online]. Available: https://www.isixsigma.com/tools-templates/hypothesis-testing/understanding-uses-moods-median-test/..

[17] "Courtney Taylor. Frequencies and Relative Frequencies. thoughtCo.," 2014. [Online]. Available: https://www.thoughtco.com/frequencies-and-relative-frequencies-3126226..

[18] A. E. P. &. R. H. Arni Magnusson, "Measuring uncertainty in fisheries stock assessment: the delta method, bootstrap, and MCMC," *Marine Research Institute,,* 2012.

[19] R. T. Bradley Efron, "Bootstrap Method. An Introduction to the Bootstrap.," 1993..

[20] J. S. Lin, "An Integrated Procedure for Bayesian Reliability Inference using MCMC.," 2014.

[21] P. Emerson, "Collective Decision-making: The Modified Borda Count," *Springer Verlag,,* vol. Part 1, 2007.

[22] T. Denoeux, "A k-nearest neighbor classification rule based on Dempster-Shafer theory," *IEEE,* vol. Vol. 25., 2002.

[23] R. Viertl, "Fuzzy Bayesian Inference," *springer.*

[24] H. C. D. D. T. K. M. F. Abhijit Sinhaa, "Estimation and decision fusion: A survey," *Elsevier.*

[25] "Analytical Hierarchy Process Overview," [Online]. Available: http://www.thequalityportal.com/q_ahp.htm.

[26] M. A. a. A. P. S. Jing Lin, "Reliability Analysis for Degradation of Locomotive Wheels using Parametric Bayesian Approach," *Quality and Reliability Engineering International,* 2013.

[27] [Online]. Available: http://stattrek.com/probability/bayes-theorem.aspx., bayes theorem. stat trek.

[28] A. Shemyakin, "Hellinger Distance and Non-informative Priors".

[29] G. Walther, "Inference and Modeling with Log-concave Distributions," p. 2009.

[30] M. A. D. J.G.IBRAHIM, "Bayesian Survival Analysis.," 2001.

[31] N. a. R. C. Balakrishnan, " Advances in Survival Analysis," *Handbook of Statistics. s.l. : Elsevier,* 2004.

[32] D. K. D. a. D. S. H. Aslanidou, "Bayesian analysis of multivariate survival data using Monte Carlo methods," *Canadian Journal of Statistics,* vol. vol. 26 no. 1, p. 33–48, 1998.

[33] S. M. N. G. a. T. C. M. Sanjeev Arulampalam, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking.," *IEEE TRANSACTIONS ON SIGNAL PROCESSING,* vol. Vol. 50, 2002.

[34] D. Alexander, "APPLICATION OF MONTE CARLO SIMULATIONS TO SYSTEM RELIABILITY ANALYSIS," 2003.

[35] N. E. N. Gordon, "A Tutorial on Particle Filters for On-line Nonlinear/Non-Gaussian," *The Institution of Engineering and Technology,,* 2001.

[36] W. R. a. W. P. Gilks, "Adaptive Rejection Sampling for Gibbs Sampling," *Journal of the Royal Statistical Society. Series C (Applied Statistics),* 1992.

[37] B. Hall, " Sampling Importance Resampling," 1993. [Online]. Available: https://artax.karlin.mff.cuni.cz/r-help/library/LaplacesDemon/html/SIR.html..

[38] Tsperry, " Rejection Sampling. R- bloggers," 2005. [Online]. Available: https://www.r-bloggers.com/rejection-sampling/.

[39] I. J. Myung, "Tutorial on maximum likelihood estimation," *Journal of Mathematical Psycology,* p. 90–100, 2003.

[40] J. S. Lin, "An Integrated Procedure for Bayesian Reliability Inference Using MCMC.," *Journal of Quality and Reliability Engineering,* 2013.

[41] "Statistical & Financial Consulting by Stanford PhD. stanfordphd.," [Online]. Available: http://www.stanfordphd.com/BIC.html.

[42] J. Lin, "An Integrated Procedure for Bayesian Reliability Inference Using MCMC," *Journal of Quality and Reliability Engineering,* vol. Volume 2014, 2014.

[43] M. S. e. a. Arulampalam, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *Signal Processing,IEEE Transactions on,* vol. 50, no. 2, pp. 174-188, 2002.

[44] "Statistics how to," [Online]. Available: https://www.statisticshowto.datasciencecentral.com/weibull-distribution/.

[45] T. L. U. a. S. U. Khan, "Particle filter based prognosis study for predicting remaining useful life of steam generator tubing," in *Prognostics and Health Management (PHM), 2011 IEEE Conference on*, 2011.

[46] J.-V. C. R. C. Limited, "Steel Industry," 2016.

[47] "Tuwairqi Steel Mills Limited," [Online]. Available: http://tuwairqi.com.pk/new/steel-industries/demand-supply.php.

[48] "Aisha Steel Mills Limited," [Online]. Available: http://www.aishasteel.com/.

[49] "International Steels Limited," [Online]. Available: https://www.isl.com.pk/.

[50] "Investopedia," [Online]. Available: https://www.investopedia.com/terms/s/shadow-banking-system.asp.

[51] G. Rubin, "Inference from iterative simulation using multiple sequences," vol. Vol. 7.

[52] R. U. S.Lewis, "How many iterations in gibbs sampler," vol. Vol. 4., 1992.

[53] S. N. C. Garren, "Convergence Diagnostics for Markov Chain Samplers," 1993.

[54] G. Brooks, "General Methods for mointoring convergence of iterative simulations," vol. Vol. 7, 1998.

[55] J.Geweke, "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments," 1992.

[56] V.E.Johnson., "Studying convergence of markov chain monte," 1994.

[57] Mykland, "Regeneration in markov chain samplers," vol. Vol. 90, 1995.

[58] M. Ritter, "Facilitating the gibbs sampler.," vol. Vol. 87., 1992,.

[59] . G.O.Roberts. UK, "Convergence diagnostics of the gibbs sampler," vol. Vol. 4, 1992.

[60] .. B. USA, "Monitoring the Convergence of Markov Samplers Based," 1994.

[61] P. U. B.Yu, "Looking at Markov Samplers through Cusum path plots," 1994.

[62] C. U. . A.Zellner, "Gibbs sampler convergence criteria," vol. Vol. 90, 1995.

[63] P. W. P.Heidelberger, "Simulation run length control in the presesnce of an intial transient.," vol. Vol. 31..

[64] C.Liu., "A variational control variable for assessing the convergence of the gibbs sampler".

[65] K. a. D. J. K. Sobczyk, Stochastic modelling of microstructures, Springer Science and Business Media, 2012.