**ANL 488 PROJECT PROPOSAL**

**Modeling Energy Mix and Average Monthly Household Electricity Consumption in Singapore using Monte Carlo Simulation**



**Submitted by**

**Ernest Pek Jun Wei**

**B1981095**

**SCHOOL OF BUSINESS Singapore University of Social Sciences**

**Presented to Singapore University of Social Sciences in partial fulfillment of the requirements for the**

**Degree of Bachelor of Science in Business Analytics**

**2023**

Contents

[**Introduction** 1](#_Toc127223726)

[Business Problem and Objective: 3](#_Toc127223727)

[Literature Review 3](#_Toc127223728)

[Article 1: Integration of Regression Analysis and Monte Carlo Simulation for Probabilistic Energy Policy Guidelines in Pakistan 4](#_Toc127223729)

[Article 2: Electricity Price Forecasting Using Monte Carlo Simulation: The Case of Lithuania 5](#_Toc127223730)

[Article 3: ElecSim: Monte-Carlo Open-Source Agent-Based Model to Inform Policy for Long-Term Electricity Planning 6](#_Toc127223731)

[Data Understanding and Preparation 8](#_Toc127223732)

[a. Outlier Replacement: 9](#_Toc127223733)

[b. Missing Value Replacement: 9](#_Toc127223734)

[Proposed Modelling and Evaluation 10](#_Toc127223735)

[Modelling the Energy Mix: 10](#_Toc127223736)

[a. Probability Distribution Estimation: 10](#_Toc127223737)

[b. Random Sample Generation: 11](#_Toc127223738)

[c. Probability Distribution of the Future Energy Mix Estimation: 11](#_Toc127223739)

[Modelling Average Monthly Household Consumption: 12](#_Toc127223740)

[a. Stationary Time Series Creation: 12](#_Toc127223741)

[b. ARIMA or SARIMA Model Identification: 13](#_Toc127223742)

[c. ARIMA or SARIMA Model Fitting: 13](#_Toc127223743)

[d. Random Sample Generation: 14](#_Toc127223744)

[Integrating the Two Models: 14](#_Toc127223745)

[a. Results Integration: 14](#_Toc127223746)

[b. Probability Calculation: 15](#_Toc127223747)

[Proposed Schedule 16](#_Toc127223748)

[References 17](#_Toc127223749)

# Introduction

The energy sector plays a critical role in the development and growth of modern city-states. As urbanization and population growth continue, it is essential to make informed decisions about electricity policy and planning so that a secure, reliable, and affordable energy supply can be provided. Accurate forecasting of electricity demand is crucial for the proper design, construction, and operation of energy infrastructure as well as for effective energy policies, programs, and implementation.

Researchers have turned to Monte Carlo simulation to account for uncertainty and variability that may impact future energy demand. Monte Carlo simulation is a statistical approach that allows for the incorporation of multiple scenarios and variables into projections, making it a valuable tool in energy demand forecasting.

Monte Carlo simulation is a technique for making predictions about the future, based on historical data and trends. The idea is to generate multiple scenarios and variables to account for the uncertainty and variability that may impact future demand. Monte Carlo simulation allows for the estimation of the probability distribution of future energy demand based on historical data and trends. This information is then used to make predictions about future consumption patterns, considering different scenarios and variables.

Monte Carlo simulation is a model-based method for predicting future energy demand. It allows for the incorporation of multiple scenarios and variables, which can be used to assess the impact of different factors on future energy demand, such as changes in population growth, economic growth, and energy efficiency measures. Additionally, Monte Carlo simulation can be used to assess the impact of different energy policies and programs on future energy demand.

In this study, Monte Carlo simulation will be used to model the probability distribution of the future energy mix and average monthly household electricity consumption in Singapore. The study will analyse historical electricity consumption data and make predictions about future consumption patterns based on trends and patterns in the data. The results of this study will provide valuable insights into the challenges and opportunities of ensuring a stable and sustainable energy supply in a rapidly growing city-state.

# Business Problem and Objective:

The proposed academic study aims to address the complex challenge of forecasting electricity demand in Singapore, a rapidly growing and developing city-state. The demand for electricity continues to increase as the city-state urbanizes, making it essential to make informed decisions about energy policy and planning to ensure a secure, reliable, and affordable energy supply.

Accurate forecasting of electricity demand is critical for the proper design, construction, and operation of energy infrastructure, as well as the development of effective energy policies and programs. However, traditional forecasting methods are limited in their ability to account for the uncertainty and variability that may impact future energy demand. To address this challenge, the study proposes the use of Monte Carlo simulation as a method for making informed predictions about the future energy mix and average monthly household electricity consumption in Singapore.

Monte Carlo simulation is a statistical method that allows for multiple scenarios and variables to be incorporated into projections. This makes it possible to account for the uncertainty and variability that may impact future energy demand. By using Monte Carlo simulation to analyse historical electricity consumption data, the proposed study aims to provide valuable insights into the future energy mix and average monthly household electricity consumption in Singapore.

The results of the simulation will provide valuable insights into the challenges and opportunities of ensuring a stable and sustainable energy supply in Singapore, advancing the academic understanding of electricity demand forecasting and its application in energy policy and planning.

# Literature Review

## Article 1: Integration of Regression Analysis and Monte Carlo Simulation for Probabilistic Energy Policy Guidelines in Pakistan

Pakistan is currently facing a severe energy crisis, due in part to the country's lack of power. The objective of this study is to develop a predictive model for energy demand and supply in the country. The study focuses on the later stages of the Integrated Energy Planning (IEP) process, including demand analysis, energy supply, energy balance, and policy formulation. It aims to integrate a stochastic approach to account for uncertainties in data measurement methods and tools, as well as previous policies in Pakistan.

The methodology of this study consists of three steps: data sources, statistical analysis, and stochastic analysis. Historical data for energy demand and supply were collected from various sources, and regression analysis was performed to develop energy supply and demand models. The best fit model was selected based on R-square value, residual plot, and p-value; the exponential equation was found to be the best fit for historical population data from Pakistan. Monte Carlo simulation (MCS) was performed using a random number generator to produce uncertainty values around the best fit model's predictions; these predictions were then used in stochastic analysis with the MCS results. Overall, 86% of the original energy supply value could be achieved or exceeded by 2025 according to these results.

The study's results indicated that Pakistan is expected to be heavily dependent on natural gas in 2050 and that there will also be a significant increase in the supply of oil, coal, and hydroelectricity. The contribution of renewable energy will continue to be low. The Consumer Price Index (CPI) is expected to increase by 36.46% by 2050; transmission and distribution losses in the country are projected to decrease by 18.36%.

The study offers several policy implications for Pakistan's energy sector, including incorporating data uncertainty and success likelihood into energy policymaking, developing capacity building programs for renewable energy resources, and emphasizing indigenous energy resources. The results of this study can help Pakistan's policymakers make informed decisions about the future supply of the country's energy sources and provide a foundation for further research in the field of demand and supply modelling.

## Article 2: Electricity Price Forecasting Using Monte Carlo Simulation: The Case of Lithuania

Tat (2018) presented a forecasting model for electricity prices in the Lithuanian day-ahead power market. The model is based on the Mean-Reverting Jump-Diffusion process introduced by R. Weron (2005), with a modification to account for price spikes. Price spikes in electricity prices refer to sudden and significant increases or decreases in spot prices that occur due to changes in generation capacity, consumption, or transmission grid outages.

Tat (2018) expressed a function that defines price spikes and their sizes based on changes in forecasted consumption, generation, export, and import. This was used to improve the accuracy of the model, which was tested with Monte Carlo simulation techniques. The mean-reversion model with the price spikes factor was able to accurately predict the general trend as well as capture most of the price spikes. The correlation coefficient of 0.81 between the model's predictions and actual electricity prices indicates a strong relationship between the two.

In conclusion, the use of Monte Carlo simulation techniques in forecasting electricity prices can be a useful tool for decision-making in the electricity market. The achieved results suggest that the mean-reversion model with the price spikes factor is effective in predicting both the general trend and price spikes in the Lithuanian day-ahead power market.

## Article 3: ElecSim: Monte-Carlo Open-Source Agent-Based Model to Inform Policy for Long-Term Electricity Planning

Agent-based modelling (ABMs) has been used to model electricity markets in the early 2000s, when researchers first developed ABMs. Since then, numerous studies have explored the suitability of agent-based models for modelling liberalized electricity markets. The key advantage of agent-based modelling is that it incorporates imperfect information and heterogeneous actors; thus, it is well suited for modelling liberalized electricity markets.

The ElecSim model, developed in the late 1980s by David Clark, is an agent-based simulation of the UK electricity market that incorporates imperfect information through forecasting of electricity demand and future fuel and electricity prices. This allows agents to take risk on their investments. The model leads to more realistic market conditions than traditional methods, which have been criticized for not reflecting reality well enough.

The results of the ElecSim model demonstrate that increasing carbon taxes can induce investment in low-carbon technologies, leading to an increase in the use of such technologies. The simulation also shows that early decisions on carbon tax have a long-term impact on energy mix. The results also show that stochastic implementation of the model leads to a significant improvement in accuracy compared to the non-stochastic case.

In conclusion, agent-based modelling of electricity markets is a promising tool for predicting the impact of carbon taxes on investment in low-carbon technologies. The use of ElecSim has provided valuable insights into the impact of carbon taxes on investment in low-carbon technologies and the importance of early decisions on carbon taxes. Future work in this field will involve comparing agent-learning techniques and integrating higher temporal and spatial resolution to model changes in daily demand, capacity factors by region and transmission effects.

# Data Understanding and Preparation

The first step in the modelling process is to clean and pre-process the data to ensure that it is ready for analysis. This involves removing any missing values, outliers, or other data anomalies. Data preparation is an important step in any data analysis project as it helps to ensure that the data is in a suitable form for analysis. In this project, the data preparation was done using IBM SPSS Modeller's Auto Data Prep node.

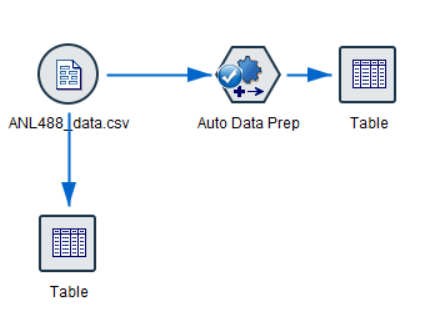


Figure 1: Using the Auto Data Prep Node in IBM SPSS Modeller

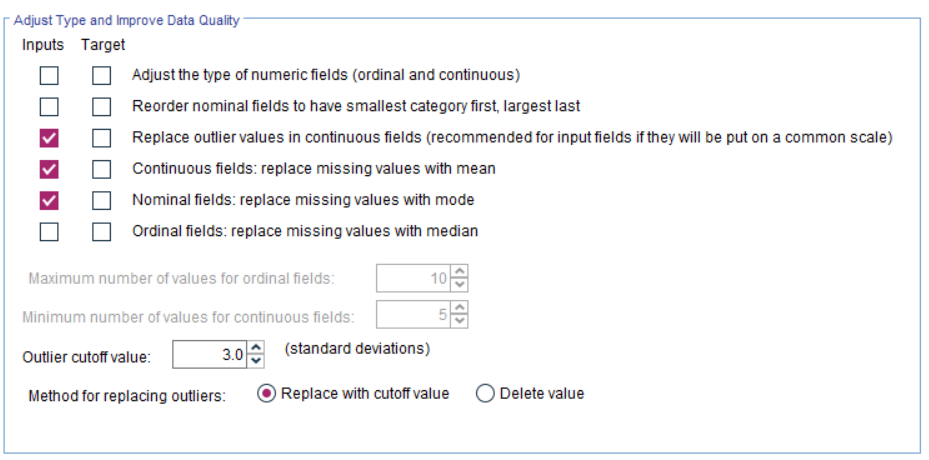


Figure 2: Settings on the Auto Data Prep Node

## a. Outlier Replacement:

The first step in the data preparation process was to replace outlier values in the continuous fields. Outliers are values that are significantly different from the rest of the data and can have a large impact on the results of the analysis. In IBM SPSS Modeller's Auto Data Prep node, the outlier values were replaced using a standard deviation approach. The outlier cutoff values were set to 3.0 standard deviations, which means that any value that is more than 3 standard deviations away from the mean was considered an outlier and was replaced.

## b. Missing Value Replacement:

The next step in the data preparation process was to replace missing values in the continuous fields. Missing values can occur for a variety of reasons, such as data entry errors or missing data. In IBM SPSS Modeller's Auto Data Prep node, the missing values were replaced with the mean of the field. This was done to ensure that the data was complete and ready for analysis.

# Proposed Modelling and Evaluation

In this project, we aim to model the energy mix and average monthly household consumption in Singapore from July 2022 to December 2023. The energy mix will be modelled based on the historical data for Steam Plant, Combined-Cycle Plants, Solar PV, and Others, while the average monthly household consumption will be modelled using a time series approach. The results of both models will be combined to simulate the future energy mix and consumption levels in Singapore.

## Modelling the Energy Mix:

The energy mix in a country is the proportion of energy produced from different sources such as coal, natural gas, renewable energy, and others. Understanding the future energy mix is important for energy planning and decision-making. In this project, the focus is on modelling the future energy mix in Singapore from July 2022 to December 2032.

### a. Probability Distribution Estimation:

To model the energy mix, the first step is to estimate the probability distribution for each energy source (Steam Plant, Combined-Cycle Plants, Solar PV, and Others). This is done using the historical data and statistical techniques such as maximum likelihood estimation. Maximum likelihood estimation is a statistical method that is used to estimate the parameters of a probability distribution that best fit the data.

The estimation of the probability distribution involves calculating the mean and variance of each energy source based on the historical data. The mean and variance are used to describe the shape and spread of the distribution, respectively. For example, if the mean and variance of the Steam Plant energy source are higher than those of the Combined-Cycle Plants, it suggests that the Steam Plant energy source is likely to produce more energy in the future.

### b. Random Sample Generation:

Once the probability distributions for each energy source have been estimated, the next step is to generate random samples of the energy mix for future periods (July 2022 to December 2023). This is done by randomly sampling from the estimated distributions for each energy source. The process is repeated several times (1000 times) to get a large number of samples of the future energy mix.

Random sampling is an important step in Monte Carlo simulation, which is a statistical method that is used to model complex systems by generating random samples of possible outcomes. In this case, the random samples of the energy mix represent different scenarios for the future energy mix in Singapore.

### c. Probability Distribution of the Future Energy Mix Estimation:

The final step in modelling the energy mix is to estimate the probability distribution of the future energy mix. This is done by aggregating the samples and calculating the frequency of each energy mix scenario. The frequency of each energy mix scenario represents the probability of that scenario occurring in the future.

For example, if the frequency of a scenario where the Steam Plant energy source produces the most energy is higher than that of a scenario where the Combined-Cycle Plants energy source produces the most energy, it suggests that the Steam Plant energy source is more likely to produce the most energy in the future.

## Modelling Average Monthly Household Consumption:

Average monthly household consumption refers to the average amount of electricity consumed by households in a particular region or country over a given period. Understanding the future trend and patterns in average monthly household consumption is important for energy planning and decision-making. In this project, the focus is on modelling average monthly household consumption in Singapore from July 2022 to December 2023.

### a. Stationary Time Series Creation:

The first step in modelling average monthly household consumption is to create a stationary time series. A time series is considered stationary if its statistical properties such as mean, and variance do not change over time. Non-stationary time series can be difficult to model as the underlying trend and patterns in the data are not constant.

To make the time series stationary, any trend and seasonality in the data must be removed. This can be done by differencing the data or by using a transformation such as taking the log of the data. The choice of transformation depends on the properties of the data and the type of model that is being used.

### b. ARIMA or SARIMA Model Identification:

Once the time series has been made stationary, the next step is to identify the appropriate ARIMA or SARIMA model for the data. ARIMA (AutoRegressive Integrated Moving Average) models are commonly used to model time series data with a stationary mean and variance. SARIMA (Seasonal ARIMA) models are similar to ARIMA models but also incorporate seasonality in the data.

The identification of the appropriate ARIMA or SARIMA model is done using the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots. These plots provide information about the relationship between the data and its lagged values, which is useful in selecting the appropriate model.

### c. ARIMA or SARIMA Model Fitting:

Once the appropriate ARIMA or SARIMA model has been identified, the next step is to fit the model to the stationary time series data. This involves estimating the parameters of the model that best fit the data. The parameters are estimated using optimization algorithms such as gradient descent or least squares.

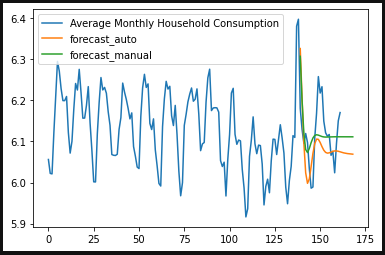


Figure 3: A trial ARIMA test forecast

### d. Random Sample Generation:

The final step in modelling average monthly household consumption is to generate random samples of future consumption levels. This is done by using the fitted ARIMA or SARIMA model to simulate the time series. The model is used to generate a large number of random samples of future consumption levels, which represent different scenarios for the future trend and patterns in average monthly household consumption.

## Integrating the Two Models:

The final step in this project is to integrate the results of the energy mix and average monthly household consumption models. This integration provides a comprehensive view of the future energy mix and consumption levels in Singapore from July 2022 to December 2032.

### a. Results Integration:

The results of the energy mix and average monthly household consumption models are combined to simulate the future energy mix and consumption levels in Singapore. This is done by considering the different scenarios for the future energy mix and consumption levels generated by the two models.

For example, if the energy mix model generates a scenario where the Steam Plant energy source produces the most energy and the average monthly household consumption model generates a scenario where the consumption levels are high, then the integration of the two models would suggest that the future energy mix in Singapore is likely to be dominated by the Steam Plant energy source and the consumption levels are likely to be high.

### b. Probability Calculation:

The final step in integrating the two models is to calculate the probability of each scenario. The probability of each scenario is calculated by aggregating the samples from both models and calculating the frequency of each scenario. The frequency of each scenario represents the probability of that scenario occurring in the future.

For example, if a scenario where the Steam Plant energy source produces the most energy and the consumption levels are high occur frequently in the aggregated samples, then it is considered to be a likely scenario with a high probability.

In summary, integrating the two models involves combining the results of the energy mix and average monthly household consumption models and calculating the probability of each scenario. The results of this integration will allow us to gain valuable insights into the future energy mix and consumption levels in Singapore and may inform decision-making and planning for the energy sector.

# Proposed Schedule

This is the project milestones:

|  |  |  |
| --- | --- | --- |
| Start Date | End Date | Description |
| 6/12/2022 | 14/12/2022 | Submit project intention survey |
| 23/1/2023 | 23/1/2023 | 1st Seminar and Briefing |
| 23/1/2023 | 13/2/2023 | Proposal |
| 20/3/2023 | 24/3/2023 | Oral Presentation |
| 24/3/2023 | 8/5/2023 | Final Report Submission |

This is the proposed project timeline:

|  |  |  |  |
| --- | --- | --- | --- |
| Start Date | End Date | Description | Duration |
| 6/12/2022 | 14/12/2022 | Project topic and Identification of Project Supervisor | 8 |
| 14/12/2022 | 23/12/2022 | Project Topic Research | 9 |
| 23/12/2022 | 23/12/2022 | Briefing with Project Supervisor | 1 |
| 23/12/2022 | 2/2/2023 | Sourcing for Data and working on Proposal | 41 |
| 2/2/2023 | 2/2/2023 | 1st Meeting with Project Supervisor | 1 |
| 2/2/2023 | 13/2/2023 | Revision of proposal based on feedback from Project Supervisor | 11 |
| 13/2/2023 | 20/2/2023 | Data Exploration and Preparation | 7 |
| 21/2/2023 | 7/3/2023 | Modelling and Evaluation | 14 |
| 8/3/2023 | 19/3/2023 | Preparation for Oral Presentation | 11 |
| 20/3/2023 | 24/3/2023 | Oral Presentation | 4 |
| 25/3/2023 | 1/4/2023 | Revision of Project | 7 |
| 2/4/2023 | 22/4/2023 | Preparation for Final Report | 20 |
| 23/4/2023 | 8/5/2023 | Finalisation of Final Report | 15 |

# References

Sajid, Z., Javaid, A., Khan, M. K., Sadiq, H., & Hamid, U. (2021). Integration of Regression Analysis and Monte Carlo Simulation for Probabilistic Energy Policy Guidelines in Pakistan. *Resources*, *10*(9), 88. <https://doi.org/10.3390/resources10090088>

Tat, A. N. (2018). Electricity Price Forecasting Using Monte Carlo Simulation: The Case of Lithuania. *Ekonomika (Economics)*, *97*(1), 76–86. <https://doi.org/10.15388/ekon.2018.1.11780>

Kell, A., Forshaw, M., & McGough, A. S. (2019). ElecSim. *Proceedings of the Tenth ACM International Conference on Future Energy Systems*. <https://doi.org/10.1145/3307772.3335321>

Dimoulkas, I., & Amelin, M. (2017). Monte Carlo simulation of district heating system short-term operation in electricity markets. *Energetika*, *63*(3). <https://doi.org/10.6001/energetika.v63i3.3560>

‌ *Rules for identifying ARIMA models*. (n.d.). People.duke.edu. <https://people.duke.edu/~rnau/arimrule.htm#:~:text=Rules%20for%20identifying%20ARIMA%20models&text=General%20seasonal%20models%3A%20ARIMA%20(0>

‌ *Plant Mix for Electricity Generation*. (n.d.). www.ema.gov.sg. Retrieved February 13, 2023, from <https://www.ema.gov.sg/TemStatistic.aspx?pagesid=20140926wbNYp2Yh8iqy&pagemode=live&sta_sid=20140802NEeM2zyMguvz>

‌ *Average Monthly Household Electricity Consumption by Dwelling Type*. (n.d.). www.ema.gov.sg. <https://www.ema.gov.sg/statistic.aspx?sta_sid=20140617E32XNb1d0Iqa>

*EMA*. (n.d.). Www.ema.gov.sg. Retrieved February 13, 2023, from <https://www.ema.gov.sg/statistic.aspx%3Fsta_sid%3D20140802apItNJRIa9Pa>

‌