Developing Hybrid Forecasting Models for Large-Scale Heavy Manufacturing Industries

Thesis for the Degree of Doctor of Philosophy (PhD)

by Herry Kartika Gandhi

Supervisor:
Prof. Dr. Márton Ispány

UNIVERSITY OF DEBRECEN
Doctoral School of Informatics
Debrecen, 2025

Hereby I declare that I prepared this thesis within the Doctoral Council of Natural Sciences and Information Technology, Doctoral School of Informatics, University of Debrecen in order to obtain a PhD Degree in Informatics at Debrecen University.

The results published in the thesis are not reported in any other PhD theses.

Debrecen, 202. . . . . .

signature of the candidate

Hereby I confirm that Herry Kartika Gandhi candidate conducted his studies with my supervision within the Applied Information Technology and its Theoretical Background Doctoral Program of the Doctoral School of Informatics between 2020 and 2024. The independent studies and research work of the candidate significantly contributed to the results published in the thesis.
I also declare that the results published in the thesis are not reported in any other PhD theses.
I support the acceptance of the thesis.

Debrecen, 202. . . . . .

signature of the supervisor

Developing Hybrid Forecasting Models for Large-Scale Heavy Manufacturing Industries

Dissertation submitted in partial fulfilment of the requirements
for the doctoral (PhD) degree in Informatics

Written by Herry Kartika Gandhi, certified Industrial Engineering scientist

Prepared in the framework of the informatics doctoral school of the
University of Debrecen

(Applied Information Technology and its Theoretical Background)

Dissertation advisor: Prof. Dr. Márton Ispány

The official opponents of the dissertation:

Dr. ....
Dr. ....
Dr. ....

The evaluation committee:

chairperson: Dr. ....
members: Dr. ....
Dr. ....
Dr. ....
Dr. ....

The date of the dissertation defence: ..... 20...

List of abbreviations

ACF . . . . . Autocorrelation Function
AIC . . . . . Akaike Information Criterion
ANN . . . . . Artificial Neural Network
AR . . . . . Autoregressive
ARCH . . . . . Autoregressive Conditional Heteroskedasticity
ARIMA . . . . . Autoregressive Integrated Moving Average
ARMA . . . . . Autoregressive Moving Average
BCOP . . . . . Brent Crude Oil Price
BIC . . . . . Bayesian Information Criterion
BPNN . . . . . Back Propagation Neural Network
BTUs . . . . . British Thermal Units
BW plot . . . . Box and Whisker plot
CEEMD . . . . Complete Ensemble Empirical Mode Decomposition
CEEMDAN . . Complete Ensemble Empirical Mode Decomposition Adaptive Noise
CNN . . . . . Convolutional Neural Network
CO . . . . . Crude Oil
Conv LSTM . . Convolutional Long Short-Term Memory
DL . . . . . Deep Learning
DM test . . . . . Diebold-Mariano test
EC . . . . . Electricity Consumption

EEMD . . . . . Ensemble Empirical Mode Decomposition
EARCH test . . Engle ARCH test
EMD . . . . . Empirical Mode Decomposition
ENC . . . . . Energy Consumption
ENN . . . . . Elman Neural Network
ENPS . . . . . Energy Price Shocks
ENS . . . . . Energy Sources
EPI . . . . . Electric Power Industry
ES . . . . . Exponential Smoothing
FNN . . . . . Feedforward Neural Network
GARCH . . . . Generalized Autoregressive Conditional Heteroskedasticity
GPR . . . . . Gaussian Process Regression
HB . . . . . Higher-Better
HDF . . . . . Hybrid Decomposition Forecasting
HLNF . . . . . Hybrid Linear-Nonlinear Forecasting
HO . . . . . Heating Oil
HOP . . . . . Heating Oil Price
HPD . . . . . Hodrick-Prescott Decomposition
HWES . . . . . Holt-Winters Exponential Smoothing
IMF . . . . . Intrinsic Mode Function
INAR . . . . . Integer-valued Autoregressive
IQR . . . . . Interquartile Range
JB test . . . . . Jarque-Bera test
KNN . . . . . K-nearest Neighbor
LB . . . . . Lower-Better
LBQ test . . . . Ljung-Box Q-test
LLC . . . . . Limited Liability Company
LNG . . . . . Liquefied Natural Gas
LSTM . . . . . Long Short-Term Memory

MA . . . . . Moving Average
MAE . . . . . Mean Absolute Error
MAPE . . . . . Mean Absolute Percent Error
ML . . . . . Machine Learning
MLP . . . . . Multi-layer Perceptron
MM . . . . . Method of Moments
MSE . . . . . Mean Squared Error
NB . . . . . Negative Binomial
NDFTS . . . . . Non-negative Discrete Forecasting Time Series
NG . . . . . Natural Gas
NGP . . . . . Natural Gas Price
NN . . . . . Neural Network
OTC . . . . . Over The Counter
PACF . . . . . Partial Autocorrelation Function
PDUs . . . . . Power Distribution Units
PINAR . . . . . Poisson Integer-valued Autoregressive
PIT . . . . . Probability Integral Transform
PJM . . . . . Pennsylvania-New Jersey-Maryland
PPIC . . . . . Production and Planning Inventory Control
PSO . . . . . Particle Swarm Optimization
QC . . . . . Quality Control
R2 . . . . . R Squared
RBF . . . . . Radial Basis Function
RMSE . . . . . Root Mean Squared Error
RNN . . . . . Recurrent Neural Network
RW . . . . . Random Walk
SAR . . . . . Seasonal Autoregressive
SARIMA . . . . Seasonal Autoregressive Integrated Moving Average
SARMA . . . . Seasonal Autoregressive Moving Average

SC . . . . . Supply Chain
SKULs . . . . . Stock Keeping Unit Locations
SMA . . . . . Seasonal Moving Average
SR . . . . . Sparse Representation
SVM . . . . . Support Vector Machine
SVR . . . . . Support Vector Regression
SW test . . . . . Shapiro-Wilk test
TLSAR . . . . . Two-Level Seasonal Autoregressive
VARIMA . . . . Vector Autoregressive Integrated Moving Average
WCOP . . . . . WTI Crude Oil Price
WD . . . . . Wavelet Decomposition

Contents

List of Figures

List of Tables

Chapter 1

Introduction

As the backbone of global economic growth, large-scale heavy manufacturing industries must ensure that products are distributed to all customers at the desired time and quantity. Upstream industries distribute raw materials to various downstream industries and store them in various warehouses at different points in the region before they reach end users such as individuals, communities, public institutions, offices, and government sectors. This relationship creates strong interrelationships called the supply chain (SC). On the one hand, each link is responsible for ensuring that the next receives the goods it needs at the exact time. On the other hand, industries must avoid overproduction and massively produce goods, resulting in wasteful quantity, and the quality of the remaining products could be degraded due to prolonged storage in the warehouse (Svetunkov and Boylan, 2020; Kilimci et al., 2019). This problem is even more critical for goods with short saleable, such as food or medicines. This phenomenon represents a significant challenge for the production and planning inventory control (PPIC) department to accurately analyse the demand for goods and project future needs using forecasting techniques. The prediction then determines the daily production to allocate to other related departments, such as production, logistics, purchasing and quality control (QC) (Alberg and Last, 2018).

Forecasting models are frequently used as the first step in production planning (Bouktif et al., 2020). Forecasting forms the foundation of the production planning hierarchy, which provides several important decisions, including purchasing raw materials, scheduling machines, bills of materials, maintaining production machinery, and scheduling delivery of goods. It is crucial to ensure that the forecasting results are as accurate as possible unless it could lead to

a shortage of demand and additional production time. On the contrary, excessive product yields overstocking in the warehouse. Rescheduling production may lead to delivery delays, which can affect customer satisfaction levels, reduce orders, and impact revenue.

In addition, it would be beneficial for the purchasing department to use forecasting techniques to predict the prices of routine materials for the production process. One such item is energy and heating resources, including coal, natural gas, heating oil and crude oil. Purchasing should monitor the daily movement of energy prices (Yu et al., 2016a) to avoid future significant energy price shocks (ENPS). They may also postpone purchases if future prices are predicted to decrease. Therefore, forecasting results also serve as a reference for the right time to purchase goods (Bollapragada et al., 2021) and consider supplier new contracts for the following long term.

The next department in the industry to be considered is the quality control (QC) department, which is responsible for preventing product defects from entering the next process. They also have the critical task of determining the additional number of products to be produced (Liu et al., 2016). At this stage, the QC employs various forecasting models to predict the number of product defects produced by each batch of products. This information is then shared with the production and planning inventory control (PPIC) department so that potential adjustments to the daily production can be considered. In the case of industries that produce various products, it may be beneficial to consider varying the additional number for each type. The other departments in industrial areas also employ forecasting, such as marketing (Abolghasemi et al., 2020; Pavlyshenko, 2019) and finance (Streimikiene et al., 2018).

Some articles used forecasting in various cases within the manufacturing industry. Abolghasemi et al. (2020) explored the potential of demand prediction in the food industry, aiming to fulfil more than 800 stock-keeping unit locations (SKULs). This technique could enhance managerial decision-making on the flow of SC to distributors. Komalawati et al. (2019) noted that price fluctuations may affect beef consumers emotionally. She observed that the slow response of beef meat producers to high demand shocks led to high prices and a shift in meat sourcing to overseas producers (imports). This could be anticipated by local producers with more accurate forecasting techniques. Jaipuria and Mahapatra (2021) conducted a study on the cement industry in eastern India and found that poor production planning may have contributed to overstock in storage. This can result in the bullwhip effect, where the demand level may contribute to significant fluctuations at the manufacturer. This case of the bullwhip effect represents a challenge in the pharmaceutical industry, as demonstrated by

Siddiqui et al. (2022). With numerous product variants, ensuring that the data analysis between retailers and industries is well-connected is crucial. Rosienkiewicz (2021) studied three vital sectors in the industry that use forecasting: PPIC, maintenance department, and QC. The accuracy of forecasting in these three sectors could potentially contribute to reducing inventory and increasing productivity. Some other articles which show the benefits of forecasting in the manufacturing industry are shown in Table 1.1.

The dataset in some articles in Table 1.1 have different characteristics. Forecasting methods require sufficient historical datasets to recognise data characteristics (such as trend and seasonality) to predict the next value (Kilimci et al., 2019; Gandhi, 2022). However, in several cases, big historical data is needed for SC cases because the demand is influenced by internal factors and many other external factors, such as promotions, competitors, and product innovation (Svetunkov and Boylan, 2020). The movement of product demand time series values are indeterminant and has high volatility (Bouktif et al., 2020; Yu et al., 2022); this uncertainties may cause single forecasting techniques to have difficulty in providing accurate predictions.

Although no single forecasting model provides the best results in all conditions, the development of hybrid forecasting has provided some more satisfactory results than single forecasting (Zhou et al., 2019; Shi et al., 2012; Shan et al., 2015). Shi et al. (2012) combined linear model of autoregressive integrated moving average (ARIMA) and non-linear model (combination of artificial neural network (ANN) and support vector machine (SVM)), which showed more accurate results than single forecasting models for wind speed forecasting. To confirm the advantages of hybrid forecasting, Wang et al. (2015) used the Ljung-Box Q-test (LBQ test) on hybrid seasonal autoregressive integrated moving average (SARIMA) model and nonlinear model, which shows satisfaction. Other articles on hybrid linear-nonlinear forecasting (HLNF) are Song et al. (2014) for municipal solid waste forecasting and Seo and Kim (2020) in the case of the volatility of bitcoin.

Another hybrid forecasting model is hybrid decomposition forecasting (HDF). This technique utilises decomposition techniques to perform time-scale signal decomposition and feature extraction and then enter the forecasting models (Srinath and Gayathri, 2022). Several forecasters use this concept. Cheng et al. (2021) used wavelet decomposition (WD) in the case of ground-level ozone as an air pollutant. Empirical mode decomposition (EMD) technique is used by Krishna and Singh (2023) with deep learning (DL) model, long short-term memory (LSTM) model, in the case of day-ahead electricity price forecasting. Several articles on hybrid forecasting development become a reference for us to

Table 1.1: Studies related to industrial forecasting

No	Author	Type of Industry	Dataset	Analysis
1	Bhatnagar et al. (2020)	Pump industry	Product demand	Forecasting products can be complicated because it can be related to many external factors.
2	Liu et al. (2016)	Common industry	Quality Control	Big data analysis can analyse quality problems and provide prevention for defective products.
3	Sharma et al. (2021)	Sugarcane industry	Sugarcane production	Demand, productivity, and product prices are related. Forecasting can equalise demand, increase productivity, and stabilise prices.
4	Kilimci et al. (2019)	Multi products industry	Product demand	Forecasting models can help reduce excess stock levels, which can result in revenue loss, insurance costs, and quality loss.
5	Ramyar and Kianfar (2019)	Energy industry	Crude oil price	Many industries are affected by the volatility of crude oil price, forecasting techniques can understand the direction of oil price movement and provide better decisions and policies.
6	Sigauke and Chikobvu (2011)	Electric power industry (EPI)	Daily demand in South Africa	The forecasting technique can analyse the next best planning of each daily production so that it becomes a reference for strategic planning in production capacity expansion.
7	Jaiswal et al. (2022)	Potato markets of India	Agriculture commodity price	This forecasting study is to determine the regional price benchmark in Asia and see the export potential in the global market.

apply to several cases in the industrial sector.

To summarise, here are the benefits of using forecasting techniques that have had a positive impact on the industrial sector:

a. Forecasting helps anticipate demand changes and reduce uncertainty

   Forecasting will help business and industry anticipate quickly the effects of uncertainty from internal and external factors, reduce risks, and create a stable SC balance (SAFIO Solutions, 2020).

b. Provide cooperation between several departments

   Forecasting for various cases in the industrial area will help strengthen the cooperation between related departments. This technique utilises established scientific analyses to ensure comprehensible and complementary results.

c. Able to quickly see the opportunities, chances, and potentials that exist

   Business managements have an opportunity to increase the industry's potential to reach a higher demand and market based on forecasting results. The industry can expand by adding production machines or expanding branches in other regions to capture demand gaps that competitors have not filled.

d. Guides scheduling for the rest of the company

   Machine scheduling planning, the number of raw materials, and employee working hours will follow daily demand, which forecasting has determined. This concept becomes the standard for the overall planning of daily production activities (Bollapragada et al., 2021).

e. Saving money by reducing inventories

   With the help of forecasting, the amount of production will be close to the demand value without overproduction (Bouktif et al., 2020). The flow of goods in and out is more stable, so there is no more excessive accumulation of products, which can save money.

f. Increases knowledge of the market and provides insights for new product development

   The predicted decline in demand value does not always negatively impact the industry. The predictions can provide an analysis of the possibility of changes in market or other external factors. Therefore, the industry can

plan innovation or new development decisions, such as market expansion or diversification, before the decline significantly reduces revenue.

1.1 Scope of dissertation

The dissertation will take cases in large-scale manufacturing and heavy industries. Heavy industries involve huge, heavy, and complex production machinery and equipment, requiring vast product facility areas. Therefore, heavy industries involve higher capital and also a lot of employment. Heavy industries sell products to companies, other industries, and governments, such as petroleum, chemicals, textiles, mining, paper rolls, vehicles, natural gas, and energy (Ittianath, 2024; CFI Team, 2022).

From the various examples above, I provide two scope objects related to my studies as follows:

1. Scope based on the product of the heavy industry

   I use three categories of heavy industries: (1) the paper industry, (2) the electric power industry (EPI), and (3) the energy industries.

2. Scope of the datasets

   I use three types of datasets from each thesis that I describe in each chapter: (1) the number of defective products, (2) demand products and (3) product prices.

1.2 Contributions of dissertation

In this dissertation, I develop hybrid forecasting to the related datasets. Some contribution points are

1. Thesis 1

   I focus on industrial quality problems using Poisson and negative binomial first-order integer-valued autoregressive (INAR(1)) models to fit and predict the product defect dataset from paper machines containing discrete numbers. I provide some statistical theories to detect which model performs better than others. I also perform comparison tests of Poisson and negative binomial models with and without INAR(1) model, where adding INAR(1) model brings lower goodness-of-fit.

2. Thesis 2

   I draw attention to industrial demand forecasting by introducing a novel SARIMA-SVR model, which belongs to the hybrid linear and nonlinear model (HLNF). After preprocessing the daily electrical consumption dataset, I leverage the SARIMA model to incorporate the seasonal relationship inside the model. To address the residual SARIMA's non-normality and heteroscedasticity properties, I enhance the model with SVR, which better captures nonlinear relationships.

3. Thesis 3

   I am concerned about price forecasting and have successfully applied a novel combination of decomposition techniques and deep learning models to predict four datasets of daily energy prices. The decompositions could split the datasets into several different frequencies. I compare four decomposition techniques and three deep learning models. I also perform mid-level multi-step forecasting, where the prediction ranges from 30 to 90 days ahead.

1.3 Arrangement of the dissertation

I divide the three studies into the following chapters (2, 3, and 4). Each chapter will explain the literature review, theory, and research methodology. The forecasting model's results and conclusions will be discussed at the end of each chapter. Furthermore, the last chapter, chapter 5, summarises all chapters of the dissertation.

Chapter 2

Prediction of product defects using Poisson and negative binomial INAR(1) models

2.1 Introduction

The first study in this chapter focuses on defective products, one of the pressing issues in the manufacturing industry. If not managed effectively, this issue can cause delays in delivering goods to customers. Defective products are categorised into three treatments: [a] repair, [b] discard, or [c] downgrade (Palkhe, 2020). Repair requires time and cost for further handling, and at the end of the process, it also requires retesting to determine if the product has met the standard. Discard is throwing the product into scrap, while the downgraded product can be sold to different customers at a lower price. Another impact is rescheduling production to compensate for the shortage of product quantities that customers want; this also increases production costs and wastes time. This results in delayed deliveries to customers, which can give negative sentiment to customers (Sharma et al., 2019; Naikan et al., 2012).

All product quality problems arise from two main categories: controllable and noise factors. Controllable factors are problems caused by industrial assets

such as machines, machine operators, equipment, or raw materials. This factor can be addressed directly by repairing and maintaining machines, retraining machine operators, and maintaining the quality of raw materials. In other words, it can be solved immediately. Other factors are noise or uncontrollable factors such as temperature, dirt, and humidity, which come from the environment and are difficult to avoid (Montgomery, 2017). This factor is also difficult to identify, but these defects can be analysed so that there is a product tolerance to avoid shortages at the end of the process.

The constantly developing method of time series data analysis can help understand many occurrences of problems to make predictions with a high degree of accuracy (Qin et al., 2022; Zhang et al., 2017a). Industry management usually underestimates this data analysis, which mainly uses intuition to determine product tolerance, resulting in another problem, such as overproduction. In addition, computerisation development is now increasingly sophisticated. Likewise, database processing is rapidly increasing, so processing large amounts of data is no longer a problem; it helps in data analysis (Li, 2020). This dissertation aims to use historical defect data with the first method of forecasting data. Next, I will find the best parameters for various forecasting methods and compare them using error measurements to determine the best forecasting model.

The particular case in this study compared to other common studies is that the product defect data is in the form of integer or discrete values. This is mainly for industries that produce countable products such as vehicles, paper rolls, electronic devices, etc. Therefore, I use non-negative discrete forecasting time series (NDFTS) analysis, which is also widely used for integer data such as the number of infectious diseases (Bourguignon et al., 2016), number of tourists (Atmanegara et al., 2019), population studies and others. NDFTS differs from the solutions in the following study (chapters 3 and 4), which have other dataset types.

2.2 Literature review

The development of forecasting models goes back to the concept of autoregressive (AR) model by Yule (1927) and Walker (1931) in the 1920s and moving average (MA) model by Yule and Slutsky, also in the 1920s and later. Moran and Whittle (1951) introduced the combination of these two methods into the autoregressive moving average (ARMA) model, which was popularised by Box et al. (1974). These three models became one of the most important devel-

opments in discrete forecasting techniques, by introducing the Integer-Valued Autoregressive (INAR) model, introduced by McKenzie (1985). Subsequent development is the addition of the Poisson distribution to the innovation distribution. Poisson model has equidispersion characteristics, while the negative binomial model is used for overdispersion in innovation dispersion (Weiß, 2018). In this first case study, I will use the INAR model and its two other extensions, Poisson INAR (PINAR) model and negative binomial INAR (NBINAR) model.

Atmanegara et al. (2019) used NDFTS to predict the number of foreign tourists visiting Indonesia from all countries. Compared to ARIMA model, the prediction of tourists from Singapore and Bahrain using PINAR model gives more accurate results. Yang et al. (2021) used PINAR model to observe the increase in the number of accidents caused by elderly drivers in one of the cities in Japan. This model can observe the monthly increase and the variability of the case. In addition, Bourguignon et al. (2019) used this method on the number of syphilis cases and its relationship with the number of domestic violence cases. For studies with NBINAR model, Olakorede and Olanrewaju (2023) compared this method with the first order INAR model in the case of COVID-19 in Nigeria, where NBINAR model gave better short-term predictions.

Wamwea et al. (2023) also conducted COVID-19 forecasting in Nigeria with NBINAR model, using the multi-stage daily forecasting technique with input data on cumulative infected victims. Some of the above articles suggest that developing INAR(1) model is needed for forecasting because there is still an innovation distribution that can be captured by adding Poisson model or negative binomial distribution properties. To our knowledge, NDFTS research on cases in heavy manufacturing industries has yet to be done. Hence, the novelty of this study is the use of NDFTS methods, namely PINAR and NBINAR models, in the case of product defects in manufacturing production.

In this study, let $(x_t)_{t\in\mathbb{N}_0}$ be a discrete, i.e., $\mathbb{N}_0$-valued time series where $\mathbb{N}_0$ denotes the non-negative integers. Alternatively, we may call (x_t) as count process or count time series.

2.2.1 Autocorrelation function (ACF) and partial autocorrelation function (PACF)

The orders of the autoregressive (AR(1)) model can be inferred from ACF and PACF. The autocorrelation of stationary process (x_t) at lag k ($\rho(k)$) is defined as

$$\rho(k) := Corr[x_t, x_{t-k}] = \frac{\gamma(k)}{\gamma(0)} = \frac{Cov[x_t, x_{t-k}]}{\sigma(t)\cdot\sigma(t-k)}, \tag{2.1}$$

where $\sigma^2(t)$ is variance of x_t and $\gamma(k)$ is the autocovariance at lag k that $\gamma(k) = Cov[x_t, x_{t-k}]$ which does not depend on t. The range of $\rho(k)$ is $-1 \leq \rho(k) \leq 1$.

In addition, PACF of stationary process x_t (ϕ_{kk}) is the autocorrelation between x_t and x_{t-k} after the linear dependence of $\{x_{t-1}, x_{t-2}, \ldots, x_{t-k+1}\}$ removed. The partial autocorrelation can be calculated from

$$\phi_{11} = Corr[x_t, x_{t-1}] = \rho(1) \tag{2.2}$$

and

$$\phi_{kk} = Corr[x_t - \hat{x}_t, x_{t-k} - \hat{x}_{t-k}], \quad k \geq 2, \tag{2.3}$$

where $\hat{x}_t$, for $k \geq 2$, is the regression of x_{t-k} on $\{x_{t-1}, x_{t-2}, \ldots, x_{t-k+1}\}$.

2.2.2 First order autoregressive model (AR(1))

The AR(1) model is a special case of the autoregressive model of order p or AR(p). AR(p) is a technique for understanding the value of x_t based on the previous past values of the process. The AR(p) formula is

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + w_t, \tag{2.4}$$

where $\phi_1, \phi_2, \ldots, \phi_p$ are constants with $\phi_p \neq 0$ and $w_t \backsim \mathrm{WN}(0, \sigma_w^2)$ is weak white noise.

In case p=1, the formula (2.4) leads to AR(1) model equation:

$$x_t = \phi_1 x_{t-1} + w_t. \tag{2.5}$$

AR(1) is a stationary model when $|\phi_1| < 1$.

2.2.3 First order integer-valued AR model (INAR(1))

Let the innovations $(\epsilon_t)_{t \in \mathbb{N}_0}$ be an i.i.d. sequence of non-negative integer-valued random variables, such that

$$E[\epsilon_t] = \mu_\epsilon \quad \text{and} \quad Var[\epsilon_t] = \sigma_\epsilon^2 \tag{2.6}$$

and $\alpha \in (0; 1)$. INAR (1) model follows the recursion

$$x_t = \alpha \circ x_{t-1} + \epsilon_t, \tag{2.7}$$

where $\circ$ denotes the thinning operator and α is the inflation parameter.

I use binomial thinning for the thinning operation. The random variable

$$\alpha \circ X := \sum_{i=1}^{X} Z_i \tag{2.8}$$

arises from X using binomial thinning, where (Z_i) denotes a counting sequence of Bernoulli random variables with mean α. Since $Z_i \sim Bin(1, \alpha)$ and the binomial distribution satisfies the addition property, $\alpha \circ X$ has conditional binomial distribution from X; $\alpha \circ X | X \sim Bin(X, \alpha)$.

The equation (2.7) is said as INAR(1) model if all thinning operations are independent of each other and of ϵ_t; and if each t, all thinning operations and ϵ_t are independent of x_s, where $s < t$.

If we have innovations' distribution, the mean and variance of x_t using the INAR(1) model are

$$\mu = \frac{\mu_\epsilon}{1 - \alpha} \tag{2.9}$$

and

$$\sigma^2 = \frac{\sigma_\epsilon^2 + \alpha\mu_\epsilon}{1 - \alpha^2} \tag{2.10}$$

with the assumption that the model is even stationary.

2.2.4 First order Poisson INAR model (PINAR(1))

The main characteristic of the PINAR model is the innovations (ϵ_t) follow the Poisson(λ) distribution; $\mu_\epsilon = \sigma_\epsilon^2 = \lambda$. In this case, the stationary distribution follows Poisson distribution as well. As Poisson distribution's characteristic, the index of dispersion (I) is equal to 1, called equidispersion. We use I to analyse equidispersion of a distribution using the formula of

$$I := \frac{\sigma^2}{\mu} \in (0, \infty), \tag{2.11}$$

where $I \in (0; \infty)$. Since $\mu = \sigma^2 = \lambda$, the index of dispersion (I) of PINAR model is 1.

Using equations (2.9) and (2.10), we have the index of dispersion of x_t as

$$I = \frac{I_\epsilon + \alpha}{1 + \alpha}, \tag{2.12}$$

or we can write

$$I_\epsilon = I(1 + \alpha) - \alpha. \tag{2.13}$$

The dispersion behaviour of x_t can refer to the value of I_ϵ.

One method for parameter estimation is using the method of moments (MM). I replace true moments with the corresponding sample moments. From the theoretical autoregressive coefficient $\rho(1)$, the MM estimators for PINAR(1) model with parameters α and λ are as follows

$$\hat{\alpha}_{MM} := \hat{\rho}(1), \tag{2.14}$$

$$\hat{\mu}_{MM} := \bar{x}, \tag{2.15}$$

and

$$\hat{\lambda}_{MM} := \hat{\mu}_\epsilon = \bar{x}(1 - \hat{\alpha}_{MM}), \tag{2.16}$$

where $\bar{x}$ is sample mean of data.

Another Poisson distribution's characteristic is zero value probability (p_0) defined by

$$p_0 := P(X = 0) = \exp(-\lambda) \tag{2.17}$$

and the zero index (I_{zero}) (Puig and Valero, 2007):

$$I_{zero} := I_{zero}(\lambda, p_0) := 1 + \frac{\ln p_0}{\lambda} \in (-\infty; 1). \tag{2.18}$$

The value $I_{zero} < 0$ shows zero deflation (less zeros than Poison distribution), while $I_{zero} > 0$ refer to zero inflation.

2.2.5 First order negative binomial INAR model (NBINAR(1))

The data indicate NBINAR(α, n, p) model if the innovations (ϵ_t) follow the negative binomial NB(n, p) distribution. The formula of mean (μ_ϵ) and variance (σ_ϵ^2) of ϵ_t are

$$\mu_\epsilon = \frac{n \cdot p}{q} \tag{2.19}$$

and

$$\sigma_\epsilon^2 = \frac{n \cdot p}{q^2}, \tag{2.20}$$

where $q = 1 - p$.

The NB model show overdispersion property ($I > 1$) from equation (2.11), because I_ϵ is given by

$$I_\epsilon = \frac{1}{p} \tag{2.21}$$

which cause the equation (2.12) becomes

$$I = \frac{\frac{1}{p} + \alpha}{1 + \alpha}. \tag{2.22}$$

The *MM* estimators for NBINAR(1) model with parameters α, n and p are

$$\hat{\alpha}_{MM} := \hat{\rho}(1), \tag{2.23}$$

$$\hat{\mu}_{\epsilon} := \bar{x}(1 - \hat{\alpha}_{MM}), \tag{2.24}$$

$$\hat{\sigma}^2_{MM} := \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2, \tag{2.25}$$

$$\hat{I}_{MM} := \frac{\hat{\sigma}^2_{MM}}{\hat{\mu}_{\epsilon}}, \tag{2.26}$$

$$\hat{I}_{\epsilon} := \hat{I}_{MM}(1 + \hat{\alpha}_{MM}) - \hat{\alpha}_{MM}, \tag{2.27}$$

$$\hat{q}_{MM} := \frac{1}{\hat{I}_{\epsilon}}, \tag{2.28}$$

$$\hat{n}_{MM} := \frac{\hat{\mu}_{\epsilon} \cdot \hat{q}_{MM}}{1 - \hat{q}_{MM}}, \tag{2.29}$$

and

$$\hat{p}_{MM} := 1 - \hat{q}_{MM}. \tag{2.30}$$

2.2.6 Goodness-of-fit

In order to facilitate a comparative analysis of the models in question, I employ the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) as means of assessing the goodness of fit, utilising the following formulae:

$$AIC = -2\delta_{max} + 2p \tag{2.31}$$

$$BIC = -2\delta_{max} + p \ln N, \tag{2.32}$$

where δ_{max} is the maximized value of the log-likelihood function of the model, N is the number of observations, and p is the number of parameters.

2.3 The dataset

2.3.1 The dataset description

I take the object dataset from a paper industry in Banten, Indonesia. The factory has operated since 1991 with several product types: brown paper packaging, white and glossy paper, printed corrugating paper boxes, and printed white paper boxes. The product area I will discuss is brown corrugating paper, where the output is a jumbo paper roll with an average weight of 1.5 - 2 tonnes. Raw material in the form of waste paper obtained from magazines, newspapers, waste brown carton boxes or white paper, then mixed with 50% pulp. The raw materials first enter through the headbox of the paper machine. After that, the chemical and grinding process is carried out until it becomes a paper layer and then dried in the machine until it comes out as dried paper. The final process is to rewind to become a jumbo paper roll.

The type of quality I observe is sheet brake, which arises when small dirt, pebbles or sharp objects enter the machine, which causes damage to the paper layer during the process. The problem creates sheet break in the paper roll that requires rework. The dirt that caused the damage is difficult to identify; it could have been wind, gravel mixed inside the raw material, or small parts of the machine that came loose during the process. This incident does not cause the machine to stop; it is just that the product in progress will be marginalised and categorised as rejected.

Sheet break data is taken daily on three different paper machines with the exact specifications and product types. The QC department collects the data, which describes the number of paper roll defects caused by sheet break events. Therefore, the data in this study is discrete because it contains the count of defective paper rolls. The reject value can be zero if no sheet break occurs in one day. The type of cause of this defect is the noise factor because the cause of this material is most likely from environmental factors that are difficult to detect.

2.3.2 The plot of the dataset

Data was taken daily by eliminating the paper machine's maintenance schedule (Gandhi and Ispány, 2025). All three machines operate non-stop, 24 hours and seven days a week. The line plot of the dataset for the three machines can be seen in Figure 2.1, 2.2, and 2.3.

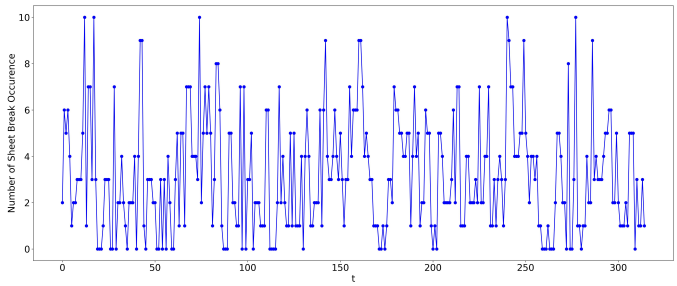


Figure 2.1: Line plot of daily defect product from machine 1

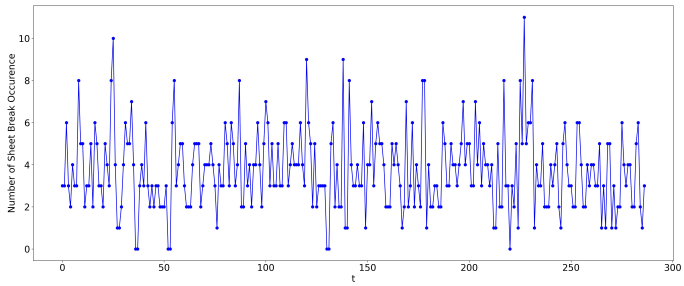


Figure 2.2: Line plot of daily defect product from machine 2

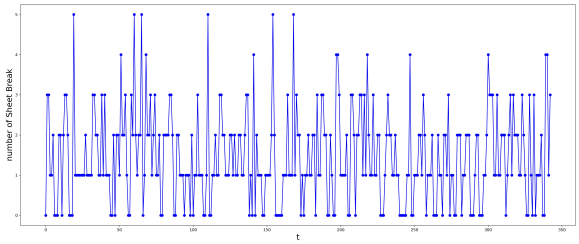


Figure 2.3: Line plot of daily defect product from machine 3

2.3.3 PACF analysis

The PACFs are illustrated in the Figure 2.4, 2.5, and 2.6.

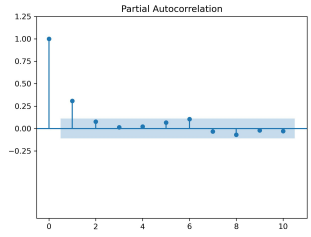


Figure 2.4: PACF plot of machine 1

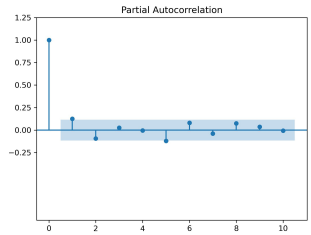


Figure 2.5: PACF plot of machine 2

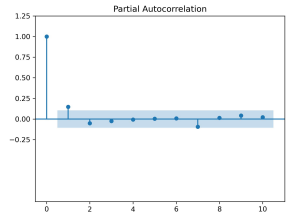


Figure 2.6: PACF plot of machine 3

Machine 1 and 3 show significant spike in lag 1, while machine 2 shows weak autocorrelation in both lags 1 and 5. Only the partial autocorrelation values at lag $k = 1$ with $|\theta_{kk}|$ are greater than the critical values. Therefore, the three datasets can be categorised as first-order autoregressive models.

2.4 Results and discussion

From Figure 2.1, 2.2, and 2.3, the three datasets are count processes ranging from 0 to 11. The sizes of datasets have different values of $n_1 = 315$, $n_2 = 287$, and $n_3 = 343$. The initial data statistic can be seen in Table 2.1.

Table 2.1: Data preliminary statistics

	machine 1	machine 2	machine 3
$\bar{x}$	3.2	3.746	1.437
s^2	6.097	3.575	1.34
ρ_1	0.307	0.125	0.148

From Table 2.1, it can be seen that machine 1 ($\bar{x} = 3.2$) and machine 2 (3.746) indicate larger numbers of product defects than machine 3 (1,437). In addition, the variance (s^2) of machine 1 data has a very large value (6.097) compared to other machines (3.575 and 1.34). Comparison $\bar{x}$ and s^2 between machine 2 (3.746 and 3.575) and machine 3 (1.437 and 1.34) indicate equidispersion in these two datasets. Meanwhile, machine 1 (3.2 and 6.097) allows for overdispersion in the dataset.

The value ρ_1 estimates the value $\hat{\alpha}$ according to the formula 2.14 and 2.23. The inflation parameter is used in the INAR(1) model in formula 2.7. All calculations and visualisations in this study were done using Python software via Jupyter notebook.

2.4.1 Analysis of Index of Dispersion

Table 2.2 shows the index of dispersion (I) values and other related parameters. Based on a normal distribution approximation, I construct upper and lower bounds at the 95% significance level: $\hat{I} - z_{\alpha/2} S_{\hat{I}} \leq I \leq \hat{I} + z_{\alpha/2} S_{\hat{I}}$. I see that I in machine 1 is 1.899 ($I > 1$), which indicates overdispersion, while in machine 2 and machine 3 ($I \approx 1$) indicates equidispersion. Determination of overdispersion

can also be seen from the value of I, which is outside the upper and lower bound [0 - 1.138]. The value of p_0 on machine 1 is much larger than the Poisson Value ($0.133 >> 0.041$), so the model on machine 1 cannot be categorised as a Poisson model; instead, this dataset can be modelled by negative binomial model.

Table 2.2: Index of dispersion preliminary parameter

	machine 1	machine 2	machine 3
I	1.899	0.951	0.930
Upper Bound	1.138	1.135	1.125
Lower Bound	0	0.7	0.802
p_0	0.133	0.024	0.224
Poisson Value	0.041	0.024	0.237

Machine 2 and 3 have I values of 0.951 and 0.930, where these two values are within the range bound of [0.7 - 1.135] and [0.802 - 1.125], so this can strengthen the equidispersion property. Likewise, machine 2 and machine 3 have values of p_0 close to the Poisson value. Machines 2 and 3 can be modelled by Poisson model.

2.4.2 Model results

After I get the distribution indication from the index of dispersion analysis above, I can estimate the parameters of the PINAR(1) and NBINAR(1) models using MM estimators as in the formulas (2.14-2.16) and (2.23-2.30). Likewise, I can compare these INAR(1) models to the i.i.d Poisson and negative binomial models. These results are shown in table 2.3.

For machine 1, the negative binomial n and p are 3.559 and 0.526, while NBINAR(1) parameters n_{est}, p_{est}, and $\hat{\alpha}$ are 2.033, 0.46, and 0.254. Machines 2 and 3 use Poisson's $\mu = 3.746$ and 1.437. Using PINAR(1) model, I have estimation of mean $(\acute{\lambda})$ and inflation $(\hat{\alpha})$ = [0.133, 0.125] for machine 2 and [0.133, 0.148] for machine 3. This is the μ estimation for PINAR(1) model.

Table 2.3: Model parameters

machine 1	machine 2	machine 3
Model: Neg Binomial	Model: Poisson	Model: Poisson
$n = 3.559$	$\mu = 3.746$	$\mu = 1.437$
$p = 0.526$		
Model: NBINAR(1)	Model: PINAR(1)	Model: PINAR(1)
$n_{est} = 2.033$	$\acute{\lambda} = 0.133$	$\acute{\lambda} = 0.133$
$p_{est} = 0.46$	$\hat{\alpha} = 0.125$	$\hat{\alpha} = 0.148$
$\hat{\alpha} = 0.254$		

2.4.3 Probability integral transform (PIT) model comparison

In this section, I compare the model fit according to the PIT shapes of PINAR(1) and NBINAR(1) models for all machines. Ronald Fisher introduced the PIT in 1932. In this dissertation, the PIT establishes the basis for evaluating whether the dataset can be reasonably modelled as a specified model. In particular, the PIT is used to create an equivalent set of values and a test is subsequently conducted to ascertain the suitability of a uniform distribution for the constructed dataset (David and Johnson, 1948).

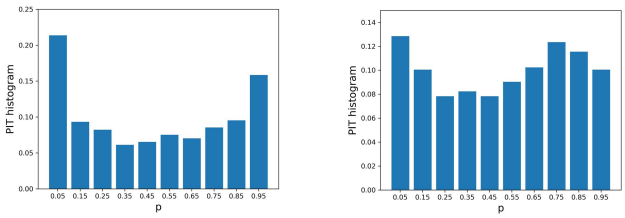


Figure 2.7: PIT on machine 1 (a) PINAR(1) model and (b) NBINAR(1) model

PINAR(1) model in Figure 2.7(a) has a probability that it is far from uniformity and gives u shape, thus showing a far result for the PINAR(1) model

fit. The PIT for the NBINAR(1) model gives close to a flat shape (Figure 2.7(b)) and shows that NBINAR(1) model of machine 1 gives a better fit than PINAR(1) model.

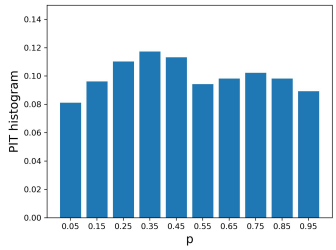


Figure 2.8: PIT of PINAR(1) model on machine 2

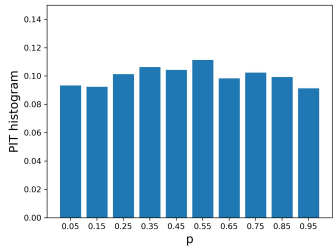


Figure 2.9: PIT of PINAR(1) model on machine 3

The figures 2.8 and 2.9 show flat shape which we can conclude that the PINAR(1) model is fit for machines 2 and 3.

2.4.4 Model selection

I use AIC and BIC as goodness-of-fit with lower better (LB) properties. I have classified the machines using the recommended models discussed above. It can be seen in Table 2.4 that for machine 1, NBINAR(1) model gives a lower error rate than the ordinary negative binomial distribution, while for machines 2 and 3, PINAR(1) models give better results than the original Poisson distribution.

The table shows that changing the models into the PINAR(1) and NBINAR(1) models than Poisson and negative binomial models provide higher accuracies.

Table 2.4: Goodness-of-fit

	Machine 1		Machine 2		Machine 3	
	Neg Binomial	NBINAR(1)	Poisson	PINAR(1)	Poisson	PINAR(1)
AIC	1401	1377	1163	1160	1027	1019
BIC	1409	1388	1167	1162	1031	1026

2.4.5 Model fitting

At this stage, I fit the selection models with the proportion data from the relative frequency distribution of the sample. In Figure 2.10, NBINAR(1) model has a peak proportion at event 2 (18.89%), whereas the actual data shows the highest percentage at event 1 (17.14%).

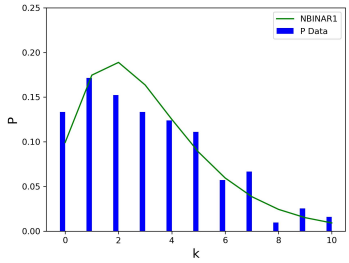


Figure 2.10: Model fitting of machine 1

Overall, the movement of the NBINAR(1) model graph with the proportion data is similar, indicating positive skewness, where all percentage values are at less than 20%.

In machine 2 (Figure 2.11), the highest percentage is in event 3, with 24.04% in actual data and 20.7% in PINAR(1) model. All events have close numbers except events 3 and 7. Likewise, machine 3 has close percentages in all events.

The event with the largest percentage between PINAR(1) model and the bar chart is in the same event, event 1. The largest event has a percentage of 0.35%.

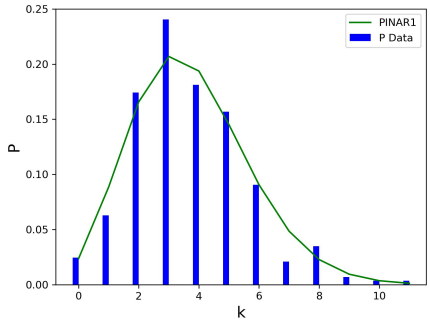


Figure 2.11: Model fitting of machine 2

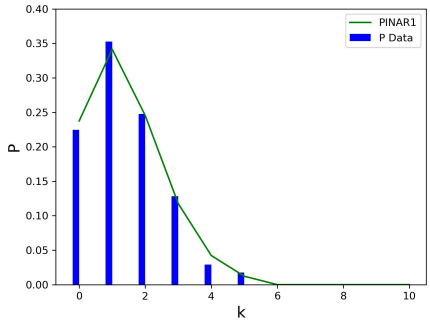


Figure 2.12: Model fitting of machine 3

Figure 2.13 compares three models. The green line chart represents machine 1 (NBINAR(1) model), the blue line chart machine 2 (PINAR(1) model), and the red bar chart machine 3 (PINAR(1) model). The figure shows that machine 3 has a lower probability of defect occurrence. The comparison is also supported by the probability of zero occurrences (p_0) = 23.75%, the highest compared to the other machines (9.93% and 2.36%).

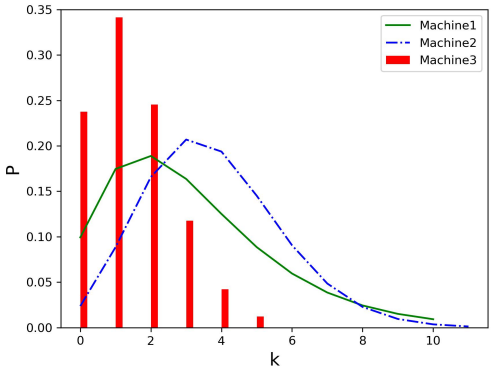


Figure 2.13: Forecasting comparison models of 3 machines

The probabilities of 5 product defects or more ($p \geq 5$) per day for machines 1, 2, and 3 are 23.51%, 32.09%, and 17.2%, respectively. The percentages of less than or equal to 2 ($p \leq 2$) defects daily from machines 1, 2 and 3 are 46.28%, 27.79%, and 57.9%.

2.5 Conclusion

This study performs model fitting on the product defect dataset in 3 paper machines producing end-product jumbo paper rolls. Given the dataset's use of discrete number, I estimate the order of the INAR models using PACF indications. The PACF value categorises the three datasets as first-order integer-valued autoregressive (INAR(1)) models.

I use four analysis tools to indicate the suitable PINAR(1) model or NBINAR(1) model of the three datasets using [1] $\bar{x}$ and s^2 analysis, [2] index of dispersion, [3] p_0 and Poisson value, and [4] PIT analysis. All methods show that the suitable model for machine 1 is NBINAR(1) model, while for machines 2 and 3 are PINAR(1) model.

In comparing INAR(1) model and the original distribution model, I use AIC and BIC. The NBINAR(1) model accuracy level for machine 1 provides higher accuracy than the negative binomial distribution. The PINAR(1) models on machines 2 and 3 are also better than the Poisson distribution. In the forecasting

analysis, machine 1 shows that the probability of less than or equal to 2 defects per day is 46.28%, machine 2 is 27.79%, and machine 3 is 57.9%. The results show that the performances of machines 1 and 3 are better than that of machine 2. The methods can illustrate how often product defects occur in each machine per day.

Thesis 1

I focus on industrial quality problems using Poisson and negative binomial first-order integer-valued autoregressive (INAR(1)) models to fit and predict the product defect dataset from paper machines containing discrete numbers. I provide some statistical theories to detect which model performs better than others. I also perform comparison tests of Poisson and negative binomial models with and without INAR(1) model, where adding INAR(1) model brings lower goodness-of-fit.

Related publications:

- Gandhi, Herry Kartika, and Ispány Márton. Analyzing uncontrollable factors that cause defective products by Poisson and negative binomial INAR(1) for fitting model. Proceedings on Engineering. doi:10.36055/jiss.v5i1.6494. (SJR: Q4) [Status: Accepted]

Chapter 3

Forecasting of daily energy consumption using hybrid linear nonlinear model: SARIMA-SVR

3.1 Introduction

Chapter three takes the common industrial problem of determining the daily quantity of products under consumer demand. Demand forecasting is also frequently discussed in industry areas (Rennie et al., 2021; Kilimci et al., 2019). The object is the electric power industry (EPI), which is vital in providing electricity consumption (EC) to all housing, offices, public facilities, and other manufacturing industries. In determining the daily electrical supply for the next few days, the planning manager must predict the appropriate amount of energy supply so that it is close to the amount of EC from consumers. Balancing between supply and demand must be maintained. A high supply of energy to consumers will cause waste energy that can reduce the efficiency of the generator supply, waste raw materials, and increase production costs (Du et al., 2018; Mirasgedis et al., 2006). Meanwhile, lack of electricity supply can cause blackouts. Although the primary energy source has a voltage control system to automatically correct the electrical torque, the system is imperfect. Additional

manual processes are still needed to maintain stability (Parihar and Bhaskar, 2018).

Many blackouts are caused by load shedding when part of the power system is disabled to avoid stress caused by overloading. Load shedding is caused by high EC and lack of planning for the availability of supplies from other power supply systems and power distribution units (PDUs). Load shedding is different from power outages in that it is done with prior planning (Barney and Courtemanche, 2023). Therefore, load shedding should be carried out with careful planning, accompanied by sufficient backup power to cover the supply shortage from main sources. This planning is based on accurate EC prediction so that good prediction and planning can avoid blackouts even if load shedding must be carried out (Bhatnagar et al., 2020).

The consequences of blackouts harm the economy and productivity of communities, businesses and offices. Halper (2022), an energy analyst, reported that every 30-minute blackout in California, United States, causes a loss of USD 15,709 for one medium-level business. Similarly, Hartsfield-Jackson Airport in Atlanta, United States, suffered a shutdown in 2018, which caused several flight cancellations with losses of up to USD 50 million (Bhattar, 2018). This incident financially impacts all public sectors in cities, such as tourism, education, government, trading, banking and financial offices.

Based on the above problems, using accurate forecasting techniques can help EPI managers balance EC without incurring significant losses. EC data is very volatile and complex to understand (Zheng et al., 2023; Chodakowska et al., 2021), but with the development of current forecasting methods using advanced mathematical methods and computational algorithms, prediction accuracy has also increased for many cases. Modern forecasting can capture trends and seasonality with characteristics of hourly, daily, weekly and monthly demand values (Sigauke and Chikobvu, 2011; Mirasgedis et al., 2006). Therefore, in this study, I want to develop hybrid forecasting from statistical and machine learning (ML) methods to predict the daily EC. I use different time horizons to determine how the model can maintain accuracy for longer time horizons.

The results of this forecasting technique are useful not only in planning the daily balance of supply and demand but also in providing several important information: [1] as a consideration in increasing electricity capacity in the long-term plan, [2] planning maintenance schedules for power generators to PDUs, [3] scheduling raw material to fulfil electricity supply, and [4] planning work schedules and employee overtime if needed.

The demand forecasting study is similar to the general demand forecasting products for common industries. They also urgently manage production to

meet customer demand and avoid a lack of raw material supplies and product overstock. Therefore, this study represents one of solutions to the problems of manufacturing industries that require reliable and accurate forecasting techniques in projecting demand for the next term. The reasons I take EPI are [1] EPI is a vital industry sector that has an impact on the surrounding economy, and [2] EC has high volatility and is strongly influenced by many external factors (such as weather, natural disasters, conflicts between countries and export and import policies) (Sigauke and Chikobvu, 2011). Hence, the complexity of the dataset is very high.

This study's novelty is the development of the SARIMA-SVR hybrid forecasting model for daily EC, which, to the author's knowledge, has never been applied in previous literature. This model benefits from combining the different characteristics of SARIMA and SVR models. SARIMA model uses a linear equation for non-stationary datasets with seasonal indication, while SVR model uses kernel tricks as nonlinear features. This study uses several statistical tests to show several indications among the predictions' residuals.

This chapter has the following sub-chapter arrangement: Section 3.1 introduces the problem, especially in EPI. Section 3.2 contains related works on forecasting methods applied to the same or related objects, the corresponding theory, formulas for statistical tests and error measurements. Section 3.3 describes dataset retrieval and the initial forecasting process. Section 3.4 contains graphs, tables of results, and analyses. Finally, section 3.5 is the conclusion.

3.2 Literature review

EC dataset is one of the most interesting objects for forecasters; some studies use other terms, such as power consumption or load consumption. The use of a single statistical model became the initial forecasting model applied in several demand product cases. Chodakowska et al. (2021) applied ARIMA model to Polish EC. The dataset tends to have random disturbance, so using a single ARIMA model is unsatisfactory. Voronin and Partanen (2014) used a hybrid combination of ARIMA and Artificial Neural Network (ANN) models in predicting the Finnish energy market. The results provide better accuracy than a single forecasting model. The combination of ARIMA model and ANN belongs to the hybrid linear and nonlinear forecasting (HLNF) category, whereas ANN model belongs to the deep learning (DL) model. Sigauke and Chikobvu (2011) used SARIMA model, an extension of ARIMA model with an indication of seasonality, with a combination of generalized autoregressive conditional

heteroscedasticity (GARCH) model on electricity demand in South Africa. In this study, several statistical tests were also conducted to see any indication of heteroscedasticity in the residuals after SARIMA model. Valipour (2015) also used SARIMA model in the case of hydrological forecasting, where SARIMA model outperformed the ARIMA model in the case of the United States. In this dataset, there is an indication of trend and seasonality, which gives the SARIMA model a more suitable predictive model. Likewise, Ampountolas (2021) compared several single methods on daily hotel demand where SARIMA model gave better results than ANN and multi-layer perceptron (MLP) models. From some of the studies above, the indication of seasonality in the dataset gives the possibility of using the SARIMA model better. The autocorrelation function (ACF) can help indicate the dataset's seasonality. The emergence of seasonality is the basis for using Seasonal ARIMA model rather than ARIMA model. Wang et al. (2015) used the Ljung-Box Q-test (LBQ test) to show that a single model of ARIMA did not bring satisfactory results compared to the results in HLNF.

In addition to statistical forecasting, the ML development in forecasting methods is also growing. The concept of support vector machine (SVM) model, one of the ML methods, is a supervised learning model that became the basic background of support vector regression (SVR). Piotrowski et al. (2021) developed a hybrid econometrics model from an ensemble of 4 popular ML methods: Long short-term memory (LSTM), MLP, SVR and K-nearest Neighbor (KNN) on energy production from wind turbine. In addition, Wang et al. (2020) also used SVR combined with LSTM for natural gas price forecasting, where the results of the proposed model outpowered the ability of single models. Ouahilal et al. (2017) used SVR in market price prediction, where SVR has advantages over regression algorithms. The advantages of SVR over other forecasting techniques are discussed in many articles. It is an extension of regression theory that can detect complex data volatility and provide better prediction output. Also, SVR model can recognise nonlinear patterns in the data movement scheme by adding kernel tricks (Sudheer et al., 2014). SVR's applications in EC forecasting have been applied in various studies (Singh and Mohapatra, 2021; Zhang et al., 2017b; Dong et al., 2022; Dai et al., 2018). Some of the studies used a combination of SVR model with other models, such as optimisation algorithm with particle swarm optimization (PSO) (Sudheer et al., 2014) and with a cuckoo algorithm (Zhang et al., 2017b). Dong et al. (2022) combined two machine learning techniques: SVR and k-means. In this study, we use HLNF, SARIMA and SVR models. To the best of the author's knowledge, this method has never been applied to the case of EC forecasting.

HLNF is a forecasting model that combines linear regression characteristics

Table 3.1: Summary article related to EC forecasting

No	Authors	Cases	Forecasting	Analysis
1	Chodakowska et al. (2021)	Data from Polish power system	ARIMA model	ARIMA model result was relatively weak
2	Voronin and Partanen (2014)	Finnish energy market	Combination of ARIMA and NN model	The combination model provided significant improves than single model
3	Ghosh (2008)	EC in northern India	SARIMA and HWES	SARIMA model performed better than ES
4	Ismail et al. (2009)	Daily EC in Malaysia	SARIMA model and the rule-based of fuzzy systems	The rule-based showed a significant improvement in forecast accuracy
5	Mirasgedis et al. (2006)	Greek power system	AR model to estimate effects of meteorological on EC	EC depends on economic variables and climatic conditions
6	Soares and Medeiros (2008)	Hourly EC in the southeast of Brazil	Two-level seasonal AR (TLSAR) model	The model is useful for EC forecasting in tropical environments
7	Jahanshahi et al. (2019)	The residential EC in the Euro area	ARIMA model	The amount predicted to be dropped due to the energy efficiency

in one model and the nonlinear ability from another model. Linear models are widely found in statistical models such as exponential smoothing (ES), Holt-Winters exponential smoothing (HWES), ARIMA, SARIMA, Vector ARIMA (VARIMA), and GARCH. Examples of models that have nonlinear abilities are SVR, DL, ANN, and LSTM. Cheng et al. (2022) combined ARIMA and SVM models to predict crude oil prices. The model detected fluctuations in dataset features that impact certain events. Shi et al. (2012) used hybrid ARIMA-ANN

model for wind power generation, where the proposed model outperformed the individual ARIMA and ANN models. The article also used multiple forecasting horizons. HLNF has the properties of a straightforward and time-saving model (Wang et al., 2015). Therefore, in this study, the combination of SARIMA and SVR models is expected to provide good results. For comparison, I also use other single and hybrid models.

I also used several statistical tests to assess the presence of normality and heteroscedasticity in residuals. These tests show whether the residuals have random, normal and homoscedasticity features or not (Shumway and Stoffer, 2019; Montgomery et al., 2015). For the model comparison test, I use the Diebold-Mariano test (DM test), a common benchmark method used by previous studies (Yu et al., 2022; Ahmad et al., 2021; Du et al., 2018). The DM test is a tool to measure the quality of the forecasting model with benchmark models using confidence levels. Finally, I also perform multi-step-ahead on the best forecasting model to assess how well the model can be applied to longer-term forecasting.

3.2.1 ARIMA and SARIMA models

The autoregressive moving average (ARMA) model combines autoregression (AR) and moving average (MA) models for stationary time series. Let $\{x_t\}$ be time series data. I have already introduced the AR(p) model in subsection 2.2.2. The MA(q) model, if $\theta_q \neq 0$, is defined by

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \ldots + \theta_q w_{t-q}, \tag{3.1}$$

where q is moving average orders and $\theta_1, \theta_2, \ldots, \theta_q$ are moving average coefficients.

Combining equations (2.4) and (3.1), we get the equation of ARMA(p, q) model for stationary time series dataset:

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q} \tag{3.2}$$

or it can be shorten as

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + w_t + \sum_{i=1}^{q} \theta_i w_{t-i}, \tag{3.3}$$

where $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregression coefficients.

ARIMA(p, d, q) model uses an additional d parameter as the order of differencing and $d \in \mathbb{Z}_+$. By using the backshift operator (B), the difference of order

d is defined as

$$\nabla^d x_t = (1-B)^d x_t. \tag{3.4}$$

The ARIMA(p,d,q) model can be written as

$$\phi(B)(1-B)^d x_t = \theta(B) w_t, \tag{3.5}$$

where ϕ and θ are the autoregressive and moving average polynomials, respectively.

The pure seasonal ARMA$(P,Q)_s$ model has formula

$$\Phi_P(B^s) x_t = \Theta_Q(B^s) w_t, \tag{3.6}$$

where s = seasonal period, Φ is the seasonal autoregressive polynomial, and Θ is the seasonal moving average polynomial. The polynomials Φ and Θ formulas are

$$\Phi(x) = 1 - \Phi_1 x - \Phi_2 x^2 - \ldots - \Phi_P x^P, \tag{3.7}$$

and

$$\Theta(x) = 1 + \Theta_1 x + \Theta_2 x^2 + \ldots + \Theta_Q x^Q, \tag{3.8}$$

where P and Q are seasonal autoregressive and moving average orders, respectively.

Multiplicative seasonal ARIMA model can be written using a seasonal difference of order D as

$$\nabla_s^D x_t = (1-B^s)^D x_t, \tag{3.9}$$

where D and $s \in \mathbb{Z}_+$. The multiplicative SARIMA$(p,d,q)\times(P,D,Q)_s$ model is

$$\Phi(B^s)\phi(B)\nabla_s^D \nabla^d x_t = \Theta(B^s)\theta(B) w_t. \tag{3.10}$$

To identify the order of SARIMA model, we can use the ACF and PACF (see 2.2.1). The ACF feature in the pure SAR model behaves similarly to the PACF in the pure SMA model, while the PACF in the pure SAR model behaves similarly to the ACF in the pure SMA model. We can see details in Table 3.2, where k=1,2,....

Table 3.2: Behaviour ACF and PACF based on pure SARMA models

	AR$(P)_s$	MA$(Q)_s$	ARMA$(P,Q)_s$
ACF	Tails off at lags $k \cdot s$	Cuts off after lag Qs	Tails off at lags $k \cdot s$
PACF	Cuts off after lag Ps	Tails off at lags $k \cdot s$	Tails off at lags $k \cdot s$

3.2.2 Support vector regression (SVR)

Boser et al. (1992) and Vapnik (1999) popularised one of the classification methods called support vector machine (SVM). Ortiz-García et al. (2010) modified the SVM model for the case of forecasting time series, named support vector regression (SVR) model, and applied it to hourly O3 concentrations case. SVR model has a similar concept to SVM model; it has high dimensions of data-solving ability and can be applied to a small amount of data (Brownlee, 2019). In addition, the combination of SVR model and nonlinear functions using kernels increases the model's capability on the hyper plane instead of linear forms (Liu, 2011). The kernels include linear, polynomial, and radial basis functions (RBF).

Let the time series dataset be $\{x_1, x_2, \ldots, x_T\}$, shortly $\{x_t\}$. The initial process of SVR model is to convert the dataset into a k-dimensional format. The k value corresponds to how much prior information we will use as information in predictions. The dataset conversion takes place in the training dataset, as follows

$$\{((x_{t-k}, x_{t-k+1}, x_{t-k+2}, \ldots, x_{t-1}), x_t), ((x_{t-k+1}, \ldots, x_t), x_{t+1}), \ldots, ((x_{T-k}, x_{T-k+1}, \ldots, x_{T-1}), x_T)\}. \quad (3.11)$$

The conversion of training dataset above consists of 2 components, input k-dimensional vector of $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$ and output value (y_i) or next prediction. Both come from the original time series. The space of the input vector is $\boldsymbol{x}_i \subseteq \mathbb{R}^k$ and the output is $y_i \in \mathbb{R}$. SVR uses weight vectors of $\boldsymbol{w}_i = (w_1, w_2, \ldots, w_k) \in \mathbb{R}^k$ and bias (b) so that the k-dimensional hyperplane equation is

$$f(\boldsymbol{x}) = \langle \boldsymbol{w} \cdot \boldsymbol{x} \rangle + b \quad (3.12)$$

$$f(x_1, x_2, \ldots, x_k) = w_1 x_1 + w_2 x_2 + \ldots + w_k x_k + b, \quad (3.13)$$

where $\langle \boldsymbol{w} \cdot \boldsymbol{x} \rangle$ is the Euclidean inner product.

With the equation 3.12, SVR tries to create the hyperplane with ϵ as max allowed error margin as shown in Figure 3.1 and forms ϵ-insensitive tube. The ϵ-insensitive tube is between ϵ^+ as the positive max error margin and ϵ^- as the negative max error margin.

The objective of SVR for dual problem is

$$\min_{\alpha, \alpha^*} \frac{1}{2}(\alpha - \alpha^*)^T Q(\alpha - \alpha^*) + \varepsilon e^T(\alpha + \alpha^*) - y^T(\alpha - \alpha^*) \quad (3.14)$$

subject to

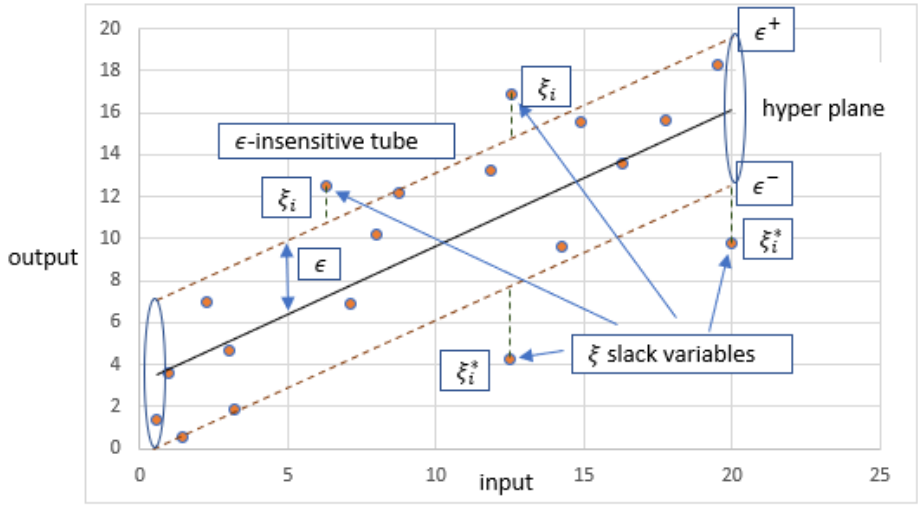


Figure 3.1: Representation of SVR (Gandhi, 2024a)

$$e^T(\alpha - \alpha^*) = 0,$$
$$0 \leq \alpha_i, \alpha_i^* \leq C,$$
$$i = 1, \ldots, n,$$

where Q is an n positive semidefinite matrix based on n, e is the vector of all ones, and $Q_{ij} \equiv K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i)^T\phi(\boldsymbol{x}_j)$ is the kernel. Here the training vectors are mapped into higher dimensional space using function ϕ.

The prediction is:

$$\sum_{i \in SV} (\alpha_i - \alpha_i^*) K(\boldsymbol{x}_i, \boldsymbol{x}) + b. \tag{3.15}$$

3.2.3 Hybrid forecasting by SARIMA-SVR model

SARIMA-SVR model belongs to the HLNF. This proposed model uses SVR to predict the value of ε_t, where ε_t are the innovations or residuals of the SARIMA predictions, $\varepsilon_t = x_t - \hat{x}_t$. The proposed model utilises the nonlinear SVR model to determine ε_t where the final prediction ($\hat{x}_{t+1}$) is the addition between output prediction of SARIMA ($\hat{x}_{t+1}$) and SVR ($\hat{\varepsilon}_{t+1}$).

I describe the input of SVR model as a k-dimensional vectors of (ε_1, ε_2,

$\ldots,\varepsilon_t)$ as

$$\{((\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_k),\varepsilon_{k+1}),((\varepsilon_2,\varepsilon_3,\ldots,\varepsilon_{k+1}),\varepsilon_{k+2}),\ldots,((\ldots,\varepsilon_{t-1}),\varepsilon_t)\}. \quad (3.16)$$

Hence, the next prediction of ε_{t+1} comes from input vector $(\varepsilon_{t-k+1},\ldots,\varepsilon_t)$.

I use recursive multi-step forecasting for multi-step analysis, as described in Fig. 3.2. I perform recursion where the output prediction of SVR $\hat{\varepsilon}_{t+1}$ becomes the input of two processes: [1] final prediction of the proposed model, $\acute{x}_{t+1}$ and [2] the next SVR model for predicting next $\hat{\varepsilon}$. I also explain the recursive model using algorithm 1 for n-step forecasting.

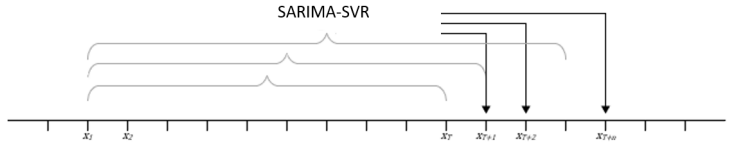


Figure 3.2: Recursive multi-step SARIMA-SVR model (Gandhi, 2024a)

Algorithm 1 Recursive multi-step SARIMA-SVR model

Input: $\{x_t\}$, $\{\varepsilon_t\}$, n_{step}, k
Output: $prediction = [\hat{x}_{t+1},\hat{x}_{t+2},\ldots,\hat{x}_{t+n_{step}}]$

$i \leftarrow 1$
$X \leftarrow \{x_t\}$
$\varepsilon \leftarrow \{\varepsilon_t\}$
$prediction \leftarrow [\,]$
while $i \leq n_{step}$ **do**
  $\boldsymbol{X}_{svr}, Y_{svr} \leftarrow$ to-SVR-input(ε, k)
  SVR$_{model} \leftarrow$ SVR.$fit(\boldsymbol{X}_{svr}, Y_{svr})$
  $\varepsilon\_hat \leftarrow$ SVR$_{model}$.$predict(\varepsilon[:-k])$
  SARIMA$_{model} \leftarrow$ SARIMA.$fit(X)$
  $x\_hat \leftarrow$ SARIMA$_{model}$.$predict(X)$
  $X.append(x\_hat + \varepsilon\_hat)$
  $\varepsilon.append(\varepsilon\_hat)$
  $prediction.append(x\_hat + \varepsilon\_hat)$
  $i \leftarrow i + 1$
end while

3.2.4 SARIMA-ARCH and SARIMA-GARCH models

Engle (1982) first introduced the ARCH model. It is constructed by identifying alterations in volatility, defined as a change in variance. Many studies use ARCH model commonly in financial cases, such as rate of return, risk management, or demand shocks.

In this model, r_t as the input of the ARCH and GARCH model is calibrated using SARIMA's innovations ε_t as

$$\varepsilon_t = x_t - \hat{x}_t, \tag{3.17}$$

$$r_t = \frac{\varepsilon_t - \varepsilon_{t-1}}{\varepsilon_t} \approx \nabla \log(\varepsilon_t), \tag{3.18}$$

where $\hat{x}_t$ is prediction from SARIMA model and ε_t is assumed to be normally distributed.

The ARCH(p) model equation is

$$\sigma_t^2 = \alpha_0 + \sum_{k=1}^{p} \alpha_k \cdot r_{t-k}^2, \tag{3.19}$$

where σ_t^2 is the variance of r_t, and the following conditions should be held: $\alpha_0 > 0$ and $\alpha_k \geq 0$. The estimation of ARCH(p) is under the assumption of normality of r_t, given by

$$r_t|(r_{t-p}, \ldots, r_{t-1}) \sim \mathrm{N}(0, \alpha_0 + \sum_{k=1}^{p} \alpha_k \cdot r_{t-k}^2). \tag{3.20}$$

The generalized ARCH or GARCH (p, q) model is the generalization of ARCH (p) by Bollerslev (1986), where the GARCH (p, q) extends equation 3.19 to

$$\sigma_t^2 = \alpha_0 + \sum_{k=1}^{p} \alpha_k \cdot r_{t-k}^2 + \sum_{k=1}^{q} \beta_k \cdot \sigma_{t-k}^2, \tag{3.21}$$

the following conditions should be held: $\alpha_0 > 0$, $\alpha_k \geq 0$ for $k = 1, \ldots, p$, $\beta_k \geq 0$ for $k = 1, \ldots, q$, and $\sum_{k=1}^{q} \beta_k + \sum_{k=1}^{p} \alpha_k < 1$.

The SARIMA-ARCH and SARIMA-GARCH models use x_t as input of SARIMA and r_t as input of ARCH or GARCH. The final prediction $\acute{x}_t$ is obtained from the sum of the two outputs from both models. The final prediction ($\acute{x}_t$) comes from

$$\acute{x}_t = \hat{x}_t + \hat{\varepsilon}_t. \tag{3.22}$$

3.2.5 Shapiro-Wilk test (SW test)

The null hypothesis is that the data has been generated from the normal distribution. The test statistic is

$$W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{3.23}$$

where n is the number of data, $x_{(i)}$ are ordered samples, and a_i are constants from Shapiro-Wilk table (Shapiro and Wilk, 1965). We have $0 \leq W \leq 1$.

The p-value for the test is from the closest W value in the Shapiro-Wilk table (Shapiro and Wilk, 1965), interpolating if necessary. If the p-value is lower than the significance level (α), we can reject the null hypothesis with the conclusion that the data has not been generated from a normal distribution.

3.2.6 Jarque-Bera test (JB test)

Apart from the SW test, the normality can also be checked with the JB test, where testing on the data looks at the skewness and kurtosis values adjusted for normality (Jarque and Bera, 1980). The null hypothesis is a joint hypothesis where the skewness and excess kurtosis values are close to zero; this indicates that the data has a normal distribution. The statistical test is

$$JB = \frac{n}{6}(S^2 + \frac{1}{4}(K-3)^2), \tag{3.24}$$

where S is the sample skewness, K is the sample kurtosis, and n is the sample size. Skewness (S) can be calculated as

$$S = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3} \tag{3.25}$$

and the formula of kurtosis (K) is

$$K = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4}, \tag{3.26}$$

where $\bar{x}$ is the sample mean, and s is the sample's standard deviation. In this study, I use the test to check skewness and kurtosis of residual data check.

3.2.7 Ljung-Box Q test (LBQ test)

The LBQ test was introduced by Greta M. Ljung and George E.P. Box as a statistical test to identify the presence of residual autocorrelation (Box and Pierce, 1970). The test investigates any autocorrelation exhibits in residuals for a number of lags. The null hypothesis is the residuals are independent distributed. The LBQ test is defined as follows

$$Q = T(T+2)\sum_{k=1}^{L}\frac{\rho(k)^2}{T-k}, \tag{3.27}$$

where T is the sample size, L is the number of tested lags of autocorrelation, and $\rho(k)$ is the sample autocorrelation at lag k.

The Q test statistic follows the asymptotic chi-square distribution with h degrees of freedom. The critical value of the null hypothesis is $Q > \chi^2_{1-\alpha,h}$

3.2.8 Engle ARCH test (EARCH test)

Engle introduced the Engle ARCH test (1982) to identify whether squared residuals are white noise and homoscedastic (Engle et al., 1985). The null hypothesis is that there is no correlation, which means the series of residuals is homoscedastic and no ARCH effect. Engle proposed that we can apply the ARCH test to the dataset in the auxiliary regression:

$$\hat{\varepsilon}_t^2 = a_0 + a_1\hat{\varepsilon}_{t-1}^2 + \ldots + a_h\hat{\varepsilon}_{t-h}^2 + u_t, \tag{3.28}$$

where $\hat{\boldsymbol{\varepsilon}}_t$ denote the least square (LS) residuals from the model.

The test statistic is the F statistic typically utilised in regression analysis on squared residuals. The null hypothesis (H_0): $a_1 = \ldots = a_h = 0$, while H_1: $a_j \neq 0$ for at least one $j \in 1, \ldots, h$ in the equation (3.28).

3.2.9 Diebold-Mariano test (DM test)

The DM test is a test to show whether the proposed forecasting model is significantly better than the benchmark model using the significance level (α). The H_0 is that the proposed model has the same accuracy as benchmark models, and the H_1 is that the proposed model is significantly better than the benchmark model using α significance level (Diebold and Mariano, 1995).

The DM test is calculated using the formula

$$DM = \frac{\frac{1}{2} + \sum_{i=1}^{n} d_i}{\sqrt{\frac{(\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k)}{n}}}, \tag{3.29}$$

where $DM \sim \mathrm{N}(0,1)$ under the H_0, d_i is loss-differential, γ_k is autocovariance at lag k, n is the sample size and $h = n^{\frac{1}{3}} + 1$.

3.2.10 Error measurement

Data forecasting requires measurement errors as (i) a measure of the accuracy of predicted data, (ii) a comparison with other forecasting models, and (iii) a reference for further model development. I use five popular tools for error measurement, which are as follows

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|x_t - \hat{x}_t| \tag{3.30}$$

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{x_t - \hat{x}_t}{x_t}\right| \tag{3.31}$$

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(x_t - \hat{x}_t)^2 \tag{3.32}$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t - \hat{x}_t)^2} \tag{3.33}$$

$$R2 = \frac{\sum_{t=1}^{T}(\hat{x}_t - \overline{x})^2}{\sum_{t=1}^{T}(x_t - \overline{x})^2} \tag{3.34}$$

MAE and MAPE give equal weight to the deviations between predicted data and actual, while MSE and RMSE are sensitive to outliers. R2 is Higher-Better (HB), while MAE, MSE, MAPE, and RMSE are Lower-Better (LB).

3.3 Dataset

The dataset was obtained from PJM Interconnection LLC (`www.pjm.com`) and is available at `https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption?resource=download&select=AEP_hourly.csv`. This data is electrical consumption in the eastern part of the United States, covering Virginia, West Virginia, Indiana, Maryland, Kentucky, North Carolina, Pennsylvania, Illinois, Michigan, New Jersey, Delaware, Tennessee, and Ohio. The region covers 60 million people with a revenue of around $42 billion.

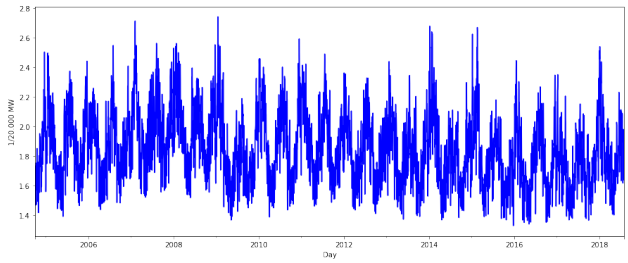


Figure 3.3: Line plot of EC dataset

The data is per day from '2004-10-01' to '2018-08-02' and has a size of 5054. To shorten the nominal data, we use 1/20000 MW. The number of testing data is 30. Before the data processing, we will look at the descriptive statistics in Table 3.3.

Table 3.3: Descriptive statistics for EC dataset

No	Feature	Value
1	Count	5054
2	mean ($\bar{x}$)	1.8596
3	standard deviation (σ)	0.2367
4	min	1.3294
5	Q1	1.6864
6	median	1.8327
7	Q3	2.0161
8	max	2.7417

As shown in Table 3.3, the dataset is in the range [1.329, 2.7417] with $\bar{x} = 1.859$ and $\sigma = 0.2367$. In Figure 3.3, it can be seen that the EC values indicate seasonality. With the comparison of values (x_{min}= 1.329 and $Q1$= 1.6864) and (x_{max}= 2.7417 and $Q3$= 2.0161), it can be seen that the upper value sometimes has an extreme value high above 75% (Q3), compared to the lower value of the dataset (Gandhi, 2024a).

3.4 Results and discussion

In this section, I describe step by step until the proposed model is combined to get the final predictions. A comparison of the proposed model is made with four other benchmark models: (i) SARIMA, (ii) SARIMA-ARCH, (iii) SARIMA-GARCH, and (iv) SVR. In model building, I conduct several statistical tests to indicate the features of the residual. For the comparison test, I use DM test. I use the multistep forecasting test to see how robust the model can be applied for longer-step forecasting.

3.4.1 SARIMA model

I determine the ARIMA/SARIMA model based on the ACF plot in Figure 3.4.

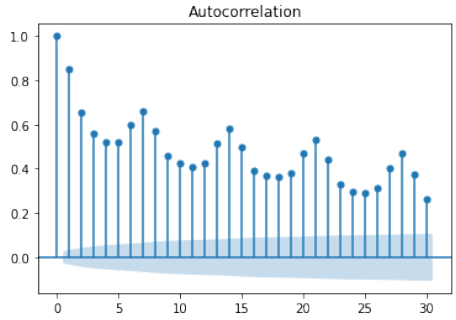


Figure 3.4: ACF plot of the dataset

Figure 3.4 shows that the ACF value decreases until lag 5, rises, and reaches the second peak in lag 7. The ACF values are still far from the value 0 even until lag 30. After seasonal differencing = 7, in Figure 3.5, the ACF value drops

to 0 at lag 4. The seasonal differencing can be a recommendation for using the SARIMA model rather than the ARIMA model. After using autoSARIMA in Python, the recommended model is (p=1, d=0, q=2) (P=3, D=1, Q=2) S=7.

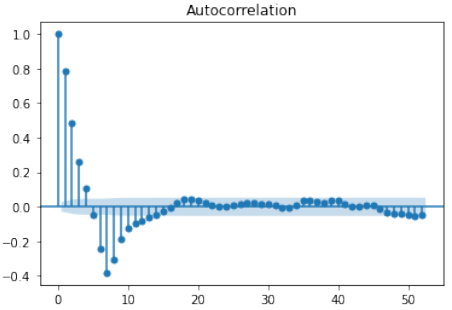


Figure 3.5: ACF plot after seasonal decomposition

3.4.2 Normality check

The results of the autoSARIMA has residuals (ε). The residual SARIMA model line plot is in Figure 3.6, it shows that the values range from [-0.4, 0.4].

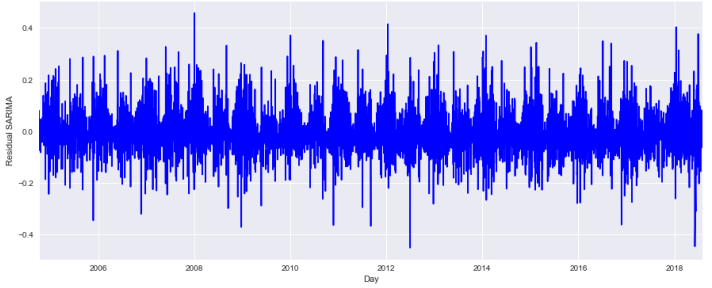


Figure 3.6: Residual plot after fitting SARIMA model

Table 3.4 shows the normality test on the residuals of all benchmark models. Using $\alpha = 0.05$, only the SVR model indicates normal distribution, where SW and JB p-values are higher than α. For the SARIMA models, two statistical tests, the SW-test and JB-test, show that the residuals do not indicate normal

distribution, using $\alpha = 0.05$. Other comparison models, SARIMA-ARCH and SARIMA-GARCH, show residuals that also do not indicate normal distribution.

Table 3.4: Normality check using SW and JB tests

Model	SW stat	SW p-value	JB stat	JB p-value
SARIMA	0.9515	0.001	18.8466	8.082E-05
SARIMA-ARCH	0.9554	0.0019	17.6552	0.0001
SARIMA-GARCH	0.9572	0.0025	18.4483	9.86E-05
SVR	0.9942	0.3059	1.7564	0.4155

3.4.3 Homoscedasticity check

In addition to the normality check above, testing homoscedasticity in the residual variance is also essential. In Table 3.5, SARIMA model has an EARCH p-value less than alpha, so I can conclude that the SARIMA residuals do not have constant residual variance. Of the four benchmark models in Table 3.5, only SVR model has a constant variance. The other two comparison models, SARIMA-ARCH and SARIMA-GARCH, do not indicate homoscedasticity in the residual value.

Table 3.5: Homoscedasticity check using EARCH tests

Model	EARCH statistic	EARCH p-value	EARCH F-Test statistic
SARIMA	32.0846	0.0004	4.3765
SARIMA-ARCH	31.8654	0.0004	4.3302
SARIMA-GARCH	31.0137	0.0006	4.1536
SVR	11.9212	0.2904	1.194

3.4.4 Residual dependency test

The Figures 3.7 and 3.8 are LBQ Tests on the residuals of single SARIMA and single SVR models. With $\alpha = 0.05$ (shown as a straight line), the LBQ test for the SARIMA model shows that of the first 20 lags, the first six lags indicate

residual dependencies. Unlike the SVR model in Figure 3.8, where all p-values are above the significance level (α), this indicates that the residuals of the SVR model have independence. Further p-values for the LBQ test in quantitative values for all models are shown in Appendix A.

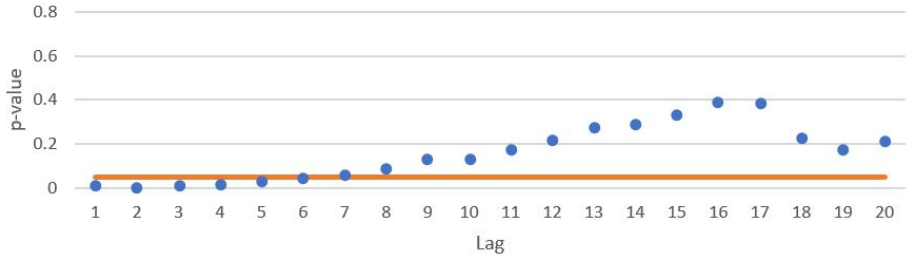


Figure 3.7: LBQ test for the residual SARIMA model

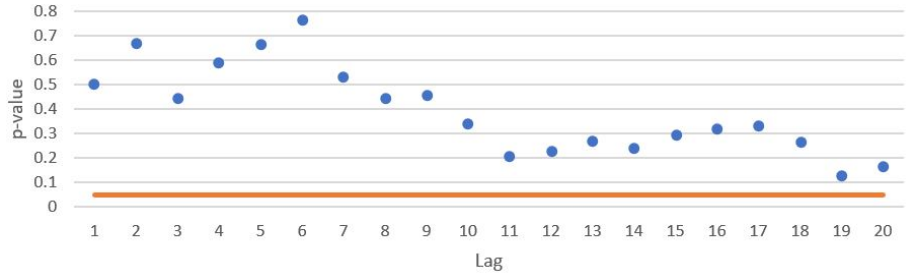


Figure 3.8: LBQ test for the residual SVR model

3.4.5 Applying SVR after SARIMA

I used Python 3.8 with the scikit.learn library in the SVR application. The application allows a grid search function with two critical parameters: $C = [0.5, 1, 5, 10, 15]$ and kernel = ['linear', 'poly', 'rbf']. The results show that the recommended parameters are [C=10; kernel='rbf'].

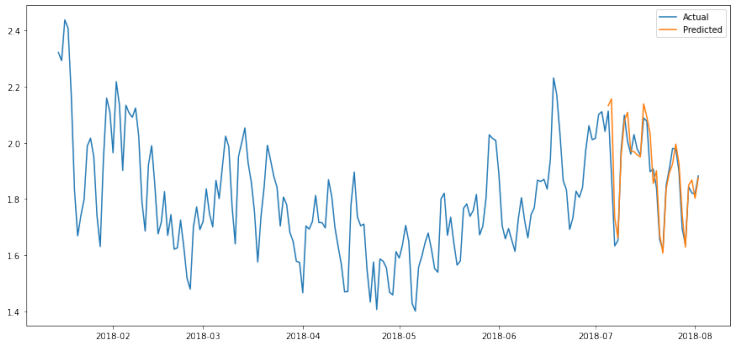


Figure 3.9: SARIMA–SVR prediction values plot

Figure 3.9 describes the result of the proposed model forecasting, SARIMA-SVR model, for 30 days predictions according to the testing data. We see that the predictions follow the actual data fairly well. The next section will discuss the comparison of the testing model result.

Next, I use statistical tests to indicate normality and homoscedasticity in the residual model. For the normality test, the SW-test and JB-test show p-values greater than α, indicating a normal distribution in the residual model. The EARCH test indicates constant variance or homoscedasticity in the proposed model's residuals, where the p-value of EARCH test greater than $\alpha = 0.05$.

Table 3.6: Summary tests of proposed model's residual

No	Statistical Test	Value
1	SW-stat	0.9792
2	SW p-value	0.1144
3	JB-stat	3.9277
4	JB p-value	0.1516
5	EARCH-stat	7.876
6	EARCH p-value	0.6409
7	EARCH F-test	0.7577

Figure 3.10 plots the p-values for the first 20 lags with the LBQ test. The JB p-values are above the significance level (α=0.05) in all lags, indicating that

the proposed model's residuals are independent in each of the lags.

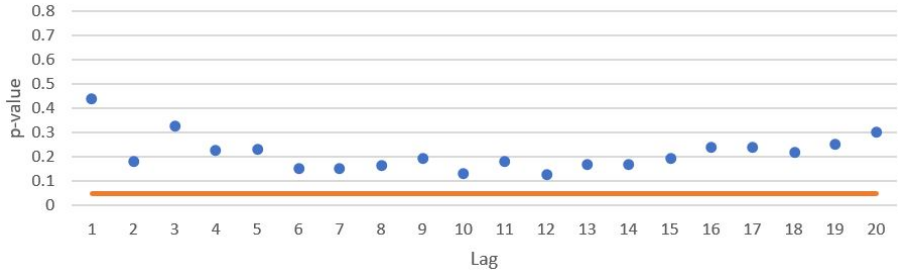


Figure 3.10: LBQ test for the residual SARIMA-SVR model

3.4.6 Error measurement test

Table 3.7 shows the five types of error measurement that have been formulated (3.30-3.34) where the model with the best accuracy shows lower values in MSE, MAE, RMSE, and MAPE and larger values in R2. SARIMA-SVR model becomes the 1st rank of the other four models because it has bigger values, although the R2 value still has a close difference with the SARIMA model. The SARIMA model has the 2nd-highest rank, whereas this single model is better than the SARIMA-GARCH and SARIMA-ARCH models, which rank 3 and 4. At the same time, the single SVR model does not show good results compared to other models.

Table 3.7: Benchmark error and measurement

Model	MSE	MAE	RMSE	MAPE	R2
SARIMA	0.0051	0.0545	0.0713	2.8546	0.7588
SARIMA-ARCH	0.0059	0.0611	0.0772	3.1911	0.7173
SARIMA-GARCH	0.0055	0.0578	0.0739	3.0121	0.7443
SVR	0.0082	0.0718	0.0906	3.8209	0.6109
SARIMA-SVR	0.0048	0.0436	0.0694	2.3214	0.75715

3.4.7 DM test

In this section, I use the Diebold-Mariano (DM) test to show whether forecasting models have significance better using confidence level $\alpha = 0.05$. In Table 3.8, I compare the models in the left-hand column with the benchmark models in the top row, where a p-value smaller than α indicates a significantly better model using α=0.05.

The SARIMA-SVR model is shown to have p-values that are all smaller than α compared with all benchmark models. The SARIMA-SVR model is significantly the best model for this case. Meanwhile, SARIMA model is seen to have a much smaller p-value than SARIMA-ARCH model but is not good enough when compared to SARIMA-GARCH model. Likewise, SARIMA-GARCH model also has a better prediction than SARIMA-ARCH model. Finally, the SVR model is the worst compared to the other four models.

Table 3.8: DM test between the models (p-value)

Models	SARIMA	SARIMA-GARCH	SARIMA-ARCH	SVR
SARIMA-SVR	0.0325	0.0119	0.014	0.024
SARIMA		0.2868	0.0055	0.098
SARIMA-GARCH			0.0159	0.0934
SARIMA-ARCH				0.0515

3.4.8 N-step horizon forecasting

I use five forecasting models (SARIMA-SVR, SARIMA, SARIMA-GARCH, SARIMA-ARCH, and SVM) with [1, 2, 3, 5, 10, 15]-steps to see how good the model is when longer steps are applied. As shown in Figures 3.11 (RMSE) and 3.12 (MAPE), SARIMA-SVR model still shows the smallest measurement error at [1, 2, 3]-steps compared to other models. However, for [5, 10, 15]-steps, SARIMA-GARCH model outperforms other models, including the SARIMA-SVR model. The SVR model does not perform as well as the other models. However, its role in improving prediction accuracy in the SARIMA-SVR model is quite significant. More details on the error measurement values are in Appendix B.

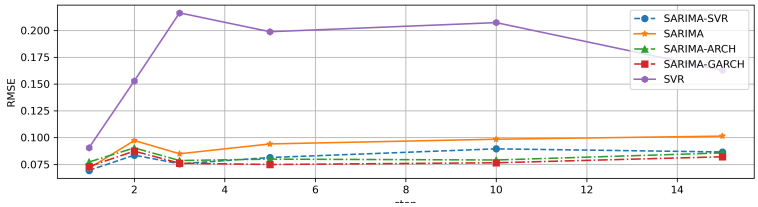


Figure 3.11: Line plot RMSE for n-steps forecasting models

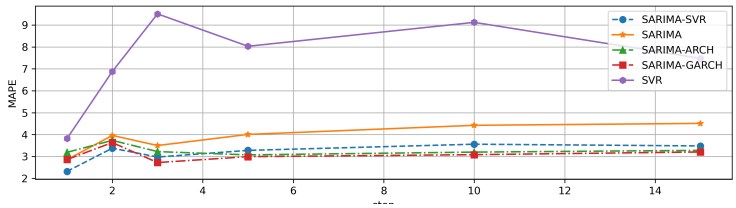


Figure 3.12: Line plot MAPE for n-steps forecasting models

3.5 Conclusion

The second study takes the scope of demand forecasting, and the object is electricity demand. As a vital daily problem, solving it with a high degree of accuracy can reduce the negative impact of blackouts, low machine efficiency, and production costs. This study combines two models with different characteristics: SARIMA with a linear model and SVR as a non-linear prediction.

The use of the proposed SARIMA-SVR model has several reasons obtained in this study: [1] There is a seasonal indication on the ACF plot, [2] Residuals from the SARIMA model still have non-normality and heteroscedasticity properties, and [3] LB test indicates the presence of dependent SARIMA's residuals between lags. The proposed model provides better error measurement than the other single models, and the DM test shows that SARIMA-SVR model significantly outperforms the four benchmark models with $\alpha = 0.05$. The SARIMA-SVR residuals also showed independent residuals and normality. Finally, using the n-step forecasting analysis, the proposed model provides consistently low measurement error values for [1, 2, 3]-steps, while SARIMA-GARCH model shows better results than other models for [5, 10, 15]-steps.

Thesis 2

I draw attention to industrial demand forecasting by introducing a novel SARIMA-SVR model, which belongs to the hybrid linear and nonlinear model (HLNF). After preprocessing the daily electrical consumption dataset, I leverage the SARIMA model to incorporate the seasonal relationship inside the model. To address the residual SARIMA's non-normality and heteroscedasticity properties, I enhance the model with SVR, which better captures nonlinear relationships.

Related publications:

- Gandhi, Herry Kartika. Applying hybrid forecasting model SARIMA-SVR for daily energy consumption data. The 2024 IEEE 3rd Conference on Information Technology and Data Science (CITDS 2024), Debrecen, Hungary, 2024, doi: 10.1109/CITDS62610.2024.10791394.

Chapter 4

Energy price forecasting using hybrid decomposition and deep learning models

4.1 Introduction

Energy sources (ENS) are vital for everyone, operating their homes, vehicles, schools, hospitals, offices, industries and all facilities worldwide. Industries absorb up to 35% of total energy consumption (ENC) (Energy Information Administration, 2023), including petroleum, natural gas (NG), coal and heating oil (HO). Some industries produce their own ENS through generators. They generate electricity totally or partially, mainly using coal or crude oil (CO). Alternatively, several industries use their material residue as internal ENS. For example, the paper industry uses black liquor, a residue from pulping wood, mixed with coal to generate electricity and heat (Wang, 2018). The majority of industries use partial electricity to anticipate external energy supply problems. Besides that, industries need ENS to produce steam or hot water for industrial processes. Most industries need steam to generate heat for their production process, which can be obtained from natural gas or heating oil. Industries also use ENS for their raw material. The petrochemical industries require CO to produce plastics or other chemical products.

As illustrated in Figure 4.1, the heavy manufacturing and mining industries

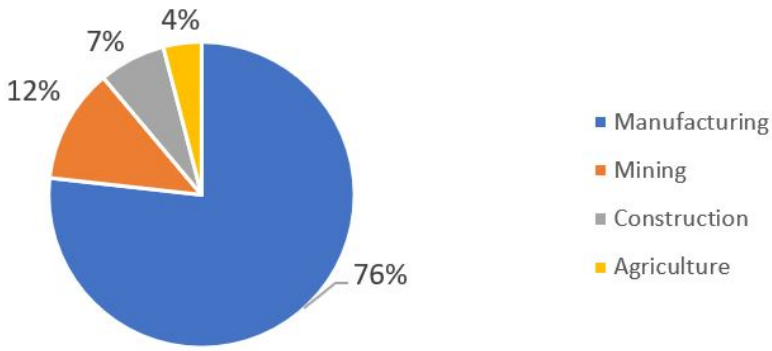


Figure 4.1: Percentage shares of industrial energy consumption (Energy Information Administration, 2023)

represent the largest ENC sectors, collectively amounting to 88% of total energy usage for industries. Therefore, the energy spent by the U.S. manufacturing annually reaches USD 230 billion and uses 25% of the nation's total energy use (Gray Solutions, 2024). Meanwhile, the average energy cost in the industry is 4.1% of total company revenues (Ondina and Fuster, 2022). In addition, ENC is also growing very fast; liquid fuel (1.8%/year), natural gas (2.2%/year), electricity (2.0%/year), and coal (1.8%/year) (Energy Information Administration, 2016). With such a significant value, the energy price is important for the industrial sector to pay attention to because, besides energy being the only material for the manufacturing process, the price increase puts great pressure on the industrial sector (energy price shock (ENPS)). Ondina and Fuster (2022) classifies several industry categories based on their impact on ENPS. The most 'extreme' impact categories are the oil refining industry and construction, which have a share of 12% and 14% of ENC. The 'very high' category is metallurgy and chemical industries, with percentage shares of 8.5% and 7.2%. The 'high' category is the paper and wood industries, while the plastic, metal, electrical equipment, and food industries are in the 'medium' category with percentage shares between 2 and 4%. All the 'extreme' to 'medium' level categories are heavy industries with large and complicated production systems and machinery. These categories have a significant impact on energy price changes. Therefore, this section focuses on predicting energy prices by developing hybrid decomposition forecasting (HDF) for good accuracy results. The objective of this study

is for the industry management to monitor the movement and indication of significant energy price changes that can affect the company's revenue. The study will discuss three ENS: crude oil (CO), natural gas (NG) and heating oil (HO).

4.1.1 Crude Oil (CO)

Crude oil is one of the vital non-renewable ENS in various sectors such as offices, businesses, public facilities and industries. CO is commonly referred to as petroleum, or in short as oil. The majority of CO is used in [1] transportation energy. CO is refined into gasoline, diesel, and kerosene and used for various means of transport. American transportation in 2021 uses 90% of fuel sources from CO processing (Tara Energy, 2018). [2] CO can be used for electric generation. Burning oil can heat water, which becomes steam that drives turbines. However, the use of oil in electricity generation is very expensive, so it is rarely done. In addition, the use of burning oil can cause air pollution, which has an impact on high carbon dioxide and methane, which triggers climate change. [3] Oil is an industrial feedstock for certain products, such as waxes, plastics, petrochemicals, and polyurethane.

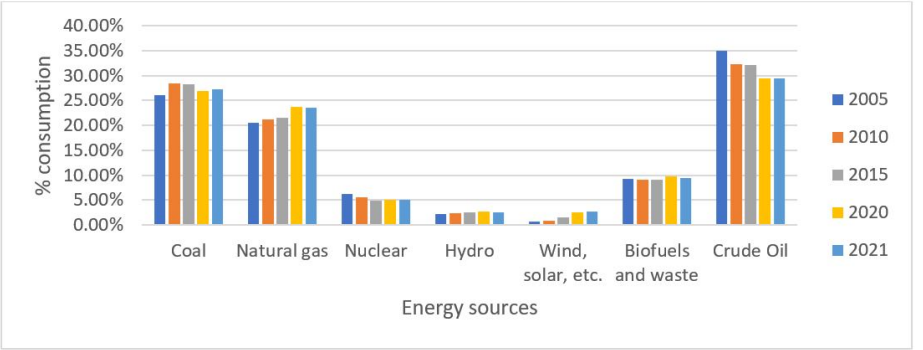


Figure 4.2: World energy balances (International Energy Agency, 2023)

As shown in Figure 4.2, CO still provides the highest percentage of worldwide ENC compared to other ENS. Likewise, the industry's most considerable dependence on CO is transportation energy. Trucks need gasoline or diesel (distillate fuels) to transport goods by land; container ships need diesel for shipping by sea; and kerosene or jet oil for aircraft. These impact the industry's stability, affecting logistics costs, especially for exporting industries. In addition, several

industries with the most significant impact on oil are [1] Oil refining industries. Increasing oil prices will create government regulations to reduce oil imports and promote public transport, thereby reducing revenue. These conditions refer to the study conducted by Farzanegan and Markwardt (2009) on Iran's economy as an oil-exporting country where oil price shocks reduce industrial income due to high import prices. [2] Petrochemical industries, as investigated by Wang and Zhang (2014) on global oil price shocks in China, petrochemicals suffer huge losses because some of their customers withhold purchasing supplies until prices stabilise again. [3] For furniture manufacturers, with the majority of raw materials from petrochemicals products, oil shocks increase plastic manufacturers' prices sharply, and many consumers are looking for other competitor manufacturers using cheaper materials such as particle board.

4.1.2 Natural Gas (NG)

Natural Gas is known as a smart source of heat energy, which is economical and efficient. NG has the following advantages in the combustion process: [1] NG emits much less carbon dioxide, sulphur, nitrogen and other particulates than other fossil fuels. [2] NG can be an energy 'bridge', to develop other ENS such as solar or wind power. [3] NG works efficiently with current technological advances; an example is the use of modern gas furnaces that can reduce the amount of NG to 40% less than older furnaces. [4] NG is safer. The delivery of NG to consumers uses underground pipelines, which has a lower leakage effect on neighbouring communities. [5] NG has an affordable price. With the more efficient nature of NG, the cost of using NG also becomes more efficient and also with good heating quality. [6] NG has increasingly widespread use. Currently, the use of NG is not limited to residential heating but also fuels transport and power generation (UniSource Energy Services, 2020). Therefore, the development of NG exploitation in various countries is now increasingly massive. This exploitation is an effective economic driver for a country and supports increasing employment and gross domestic product.

With the various advantages that NG has, NG consumption has increased dramatically, according to Figure 4.3 above. This condition also happens in the industrial area. NG is included in the top three world commodities contracts among South America, Asia, and Europe industries. NG is starting to replace coal in heating production manufacturing processes (United Nations Environment Programme, 2023), such as the iron and steel industry, with one of the benefits of reducing carbon emissions. Industrial areas that are not connected to NG pipelines can get it in the form of liquefied natural gas (LNG); this form

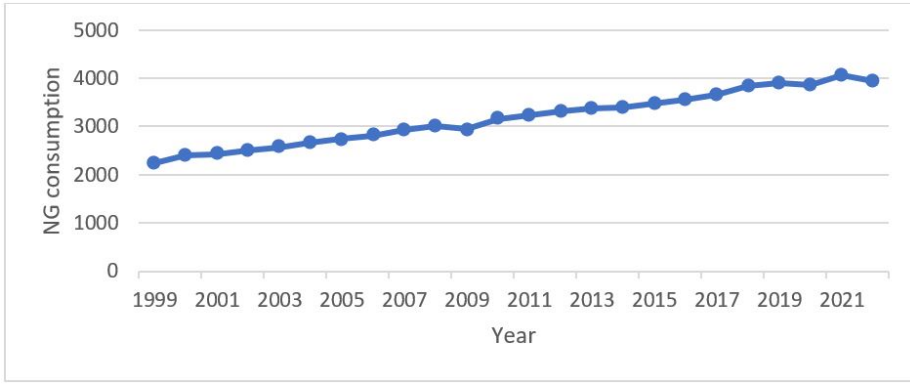


Figure 4.3: NG consumption worldwide (in billion cubic metres) (Statista Search Department, 2023)

is compact - 600 times smaller than gas form - and can be delivered by sea or land. NG purchase agreements by the industry follow a long-term contract system with NG suppliers. However, the industry must continuously monitor NG daily spot prices to plan production and prepare for long-term NG storage. Therefore, industries must complement their monitoring with a reliable forecasting system to review the next contract with the NG supplier.

4.1.3 Heating Oil (HO)

HO is a safe, efficient and durable heating ENS. HO does not burn in a liquid state, HO tanks will not explode in flames, and even if there is a leak, HO will not burn even if touched by fire. In addition, HO is an ultra-low sulfur source, so its combustion produces low levels of nitrogen oxide, sulfur oxide, and other particulates. HO is an efficient energy source because it has 35% more British Thermal Units (BTUs) than other heating sources, a measure to raise the combustion temperature of water. So, the use of HO is faster and more efficient. With low sulphur content, the tank system's cleaning frequency is less. HO system has a long-lasting nature, with a lifetime between 15 to 25 years and regular maintenance (Heatable, 2018). Despite its advantages, HO also has disadvantages, such as [1] unstable long-term price and becoming higher, [2] the need for maintenance costs for tanks, and [3] oil furnaces that have separate systems with air conditioners.

HO is the second most important crude oil product (Stein, 2024). It is used in industry to heat industrial office areas and support the heating process in manufacturing. Currently, the use of HO is shifting to NG, even though 82% of industries and residential areas in the U.S. Northeast (Energy Information Administration, 2024) and 1.5 million homes in the U.K. still use HO. This study includes HO as an object energy price in the mid-term forecasting study.

The objective of this study is to use mid-term forecasting to provide a longer view of energy price prediction for up to 3 months or more at every step. This longer view clearly helps companies and industries adopt the right policies to avoid more significant losses.

I use the proposed hybrid forecasting decomposition models with deep learning (DL) models. I develop models that provide high accuracy for the longer prediction steps. I also use multi-step analysis to measure how reliable each forecasting model is in keeping the accuracy high.

4.2 Literature review

The study consists of 4 essential energy prices: crude oil price, including Brent crude oil price (BCOP) and WTI crude oil price (WCOP), natural gas price (NGP) and heating oil price (HOP). Zhang and Zhang (2018) and Yu et al. (2016a) stated that energy prices have complex, irregular, and nonlinear characteristics. The level of complexity in the volatility of the datasets is a challenge for forecasting models to produce reliable predictions (Sehgal and Pandey, 2015; Ghalayini, 2017). Energy prices are vulnerable to various external factors, such as natural disasters, extreme weather changes, wars, pandemic diseases, policies of exporting countries, and others. Therefore, energy price changes can be rapid and show very complex volatility. Some forecasters are interested in implementing new model innovations into these objects. One of the most widely developed innovation techniques is the hybrid model, which combines techniques into a new forecasting model.

One of the hybrid forecasting techniques is hybrid decomposition forecasting (HDF). HDF concept is 'decompose and ensemble' and has three main steps: data decomposition, individual forecasting, and ensemble method. Wang et al. (2018) used the wavelet decomposition (WD) technique with several DL models in solar irradiance forecasting. The combination of WD and convolutional LSTM model gives better results than WD with ANN or CNN model. Yu et al. (2016b) applied a hybrid ensemble empirical mode decomposition (EEMD) and back propagation neural network (BPNN) model to detect nonlinear and non-

stationary datasets. This hybrid model is proven to be better than other single models such as ARIMA, ES, and fuzzy logic-based approaches. Likewise, Yun et al. (2023) predicted carbon dioxide emission price with a hybrid of complete EEMD adaptive noise (CEEMDAN) with LSTM. The combination model provides a significant advantage of using decomposition techniques and is proven to separate datasets according to time-frequency features. Some of these studies prove some applications of HDF on other object datasets and have more satisfactory results.

Regarding energy prices, the development of forecasting models also continues to progress. As oil price prediction became a striking issue, Abdollahi (2020) performed CEEMD with SVM model and switching GARCH. With the DM test, HDF outpowered other single models. Lu et al. (2021) compared five single methods including random walk (RW), ARMA, Elman neural network, LSTM and GARCH. Of the five models, LSTM provides the lowest measurement error. Cheng et al. (2022) combined CEEMD-ARIMA-SVM model as HDF to forecast the volatility of BCOP and WCOP. In addition, Wang et al. (2020) investigated NGP and took a combination of two models, SVR and LSTM models, to see how well the weighted hybrid model performed on the NGP dataset. Comparison with other individual models, the proposed model has the highest prediction ability. I summarise some of the studies that have been done on the energy price dataset in Table 4.1, and to my knowledge, the application of HDF with DL models still needs to be done. With several successful HDFs conducted on other object datasets, I am motivated to apply them to the four energy prices.

HDF has various advantages over a single forecasting model in many ways: [1] Decomposition techniques can break down complex data sets into something that is easy to understand (Yu et al., 2017). [2] The decomposed dataset can significantly reduce the difficulty of the single forecasting model that follows it (Sun et al., 2022). [3] HDF has uncomplicated steps, with only three main processes: decomposing the data set, forecasting each decomposition, and ensemble into one prediction (Dong et al., 2019). [4] HDF can transform nonlinear complex datasets with multiscale and can substantially improve prediction accuracy (Dong et al., 2019). For the HDF that I perform here, I use four different decomposition methods, namely Hodrick-Prescott decomposition (HPD), wavelet decomposition (WD), empirical mode decomposition (EMD) and complete ensemble EMD (CEEMD) for comparison.

I use several DL models for the forecasting model, which follows decomposition. DL is the branch of machine learning methods analogous to human brain neurons' working system. DL is an attractive method nowadays because it is

Table 4.1: Summary articles related to energy price

No	Authors	Datasets	Forecasting	Results
1	Ramyar and Kianfar (2019)	BCOP and WCOP	ANN model	ANN model performs better than AR
2	Yu et al. (2017)	BCOP	Feedforward Neural Network (FNN)	FNN with sparse representation (SR) showed better accuracy than single FNN
3	Lu et al. (2021)	BCOP and WCOP	LSTM model	LSTM model performed better compared with random walk (RW), ARMA, and ENN models
4	Abdollahi (2020)	BCOP and WCOP	CEEMD-SVM-Markov switching GARCH	Proposed model is more accurate than benchmark models
5	Aamir et al. (2018)	Monthly BCOP	CEEMDAN-ARIMA model	Proposed model is better than using EMD and using ARIMA single model
6	Čeperić et al. (2017)	NGP	Neural Network (NN) and SVR model	Proposed model showed better forecasting than classical models (naïve, AR, and ARIMA) for short-term
7	Nguyen and Nabney (2010)	NGP	Wavelet and multi-layer perceptron (MLP)	The model overpowered wavelet and GARCH model
8	Su et al. (2019)	NGP	ANN model	ANN model showed better accuracy compare with SVM, gradient boosting machines, and Gaussian process regression (GPR) model
9	Fałdziński et al. (2021)	HOP	Comparison SVR and GARCH	SVR model gave lower error measurement than GARCH model
10	Yazdanfar (2015)	HOP	GARCH (1, 1) model	GARCH (1, 1) model had high performance in short-term horizons, for mid-term the performance becomes decreases

able to identify long-term dependencies in datasets with nonlinear approaches to create more accurate predictions (Bouktif et al., 2020). DL requires big data to train continuously on the model and perform self-improvement with back-propagation optimisers. This study uses three popular DL models: Artificial Neural Network (ANN), LSTM, and Convolution Neural Network-LSTM (CNN-LSTM). Each model will be fitted to each decomposition result and for every period-term forecasting. I use mid-terms along with multi-step analyses.

4.3 Decomposition methods

4.3.1 Hodrick-Prescott decomposition (HPD)

The Hodrick-Prescott Decomposition (HPD), also known as the Hodrick-Prescott Filter, was introduced by Hodrick and Prescott (1997). This theory was developed from Whittaker's concept in 1923 (Whittaker, 1922). This concept was first introduced in business theory, macroeconomics, where it can detect trends in raw data and separate the cyclic components. In addition, this theory is also used to smooth time series.

In this study, HPD is used to separate the time series raw data (x_t) into trend components (τ_t) and cyclical components (c_t). The model express this relationship through

$$x_t = \tau_t + c_t. \tag{4.1}$$

The objective function is to minimise the following equation:

$$\min_{\tau_1,\tau_2,\ldots,\tau_t} \left\{ \sum_{t=1}^{T} (x_t - \tau_t)^2 + \lambda \cdot \sum_{t=2}^{T-1} ((\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1}))^2 \right\}, \tag{4.2}$$

where T is the sample size and λ is the smoothing parameter. We can transform the data using a natural logarithm as $u_t = \ln x_t$.

From the equation (4.2), there are two parts: [1] The left part, $\sum_{t=1}^{T}(x_t - \tau_t)^2$, indicates how close the results are to the actual value. [2] The right part, $\lambda \cdot \sum_{t=2}^{T-1}((\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1}))^2$, can be defined as the change in the movement of the trend component. The λ is kind of regularization that gives how much influence the right component has on the left component. The right component will have a smaller value as the change in trend (τ_t) is close to constant. The value of the weighting factor (λ) serves as a trade-off consequence; increasing λ directs more attention to the right component. Conversely, lowering λ gives more focus on the trend value (τ_t).

4.3.2 Wavelet decomposition (WD)

Alfred Haar introduced the Haar wavelet method in 1911 (Haar, 1911). This method takes a sequence of numbers rescaled 'square-shaped' and transforms it into wavelet form. The Haar sequence is known as the first base wavelet, and later, other wavelet models, such as Fourier analysis, were developed. This study uses Haar wavelets because this technique is fast processing, has sparse representation, and can be applied to algorithms quickly.

The calculation of the wavelet transform consists of two types: scaling function $\varphi(t)$ and wavelet function $\psi(t)$ with function spaces of V and W. If $f(t) \in V_{j+1}$, then $f(t)$ can be described using $\varphi(t)$ from V_{j+1} as

$$f(t) = \sum_k c_k \varphi(2^{j+1}t - k) \tag{4.3}$$

or we can combine $\varphi(t)$ from V_j with $\psi(t)$ from W_j as

$$f(t) = V_j + W_j = \sum_k c_k \varphi(2^j t - k) + \sum_k d_{j.k} \psi(2^j t - k), \tag{4.4}$$

where j corresponds to scales, k is index which is vary from 0 to 2^{j-1}; $k, j \in \mathbb{Z}$, and $c_k, d_{j.k} \in \mathbb{R}$.

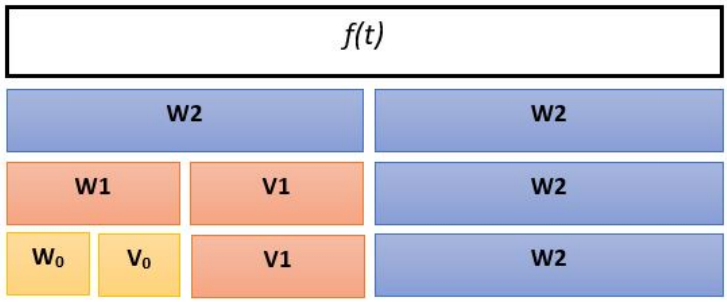


Figure 4.4: Wavelet decomposition representation

The Haar coefficients are

$$c_0, d_{0.0}, d_{1.0}, d_{1.1}, d_{2.0}, d_{2.1}, \ldots, d_{j.k}, \ldots, d_{(n-1).(2^{n-1})}, \tag{4.5}$$

which in Figure 4.4, comes from $V_0, W_0, W_1, \ldots, W_j$.

For the higher number of j, we can use recursion on V_{j+1}:

$$V_{j+1} = V_{j-1} + W_{j-1} + W_j = V_0 + W_0 + W_1 + W_2 + \ldots + W_j \tag{4.6}$$

and if $f(t) \in V_{j+1}$, then

$$f(t) = \sum_k c_k \varphi(t-k) + \sum_k \sum_j d_{j.k} \psi(2^j t - k), \tag{4.7}$$

where $V_0 = \sum_k c_k \varphi(t-k)$.

4.3.3 Empirical mode decomposition (EMD)

EMD was first introduced by Huang et al. (1998). It consists of three words: [1] Empirical, which is an adaptive system for processing decomposition; [2] Mode, which is part of the complete signal; and [3] Decomposition, which is the result of the process that shows insight into inherent features (Liu et al., 2018; Zeng et al., 2022). Thus, the objective of EMD is to break down or separate the dataset into components that can help analyse and understand its inherent features.

The output of EMD is intrinsic mode functions ($imfs$) and a residual. The larger index of $imfs$ indicates smaller scales or smaller frequencies. Original signals are the sum of all modes and residuals, as per the following formulation

$$x_t = \sum_i imf_i + res. \tag{4.8}$$

Let f be $f(t) = (x_1, x_2, \ldots, x_T)$. To get imf_i, there are several stages in EMD, as follows.

Step 1. Find local extrema and put them into envelopes

If $x_t > x_{t-1}$ and $x_t > x_{t+1}$, then $x_t \to x_{maxima}$ and $E_{up} \cup (t, x_t)$.
If $x_t < x_{t-1}$ and $x_t < x_{t+1}$, then $x_t \to x_{minima}$ and $E_{low} \cup (t, x_t)$, where E_{up} is upper envelope and E_{low} is lower envelope.

Step 2. Generate $E_{up}(t)$ and lower envelope $E_{low}(t)$ with cubic splines interpolation

Step 3. Generate $E_{mean}(t)$ by calculate point-by-point average of upper and lower envelopes

$$E_{mean}(t) = \frac{[E_{up}(t) + E_{low}(t)]}{2}. \tag{4.9}$$

Step 4. Residual ($res(t)$)

$$res(t) = f(t) - E_{mean}(t) \tag{4.10}$$

Step 5. If $N(E_{up}) + N(E_{low}) - N(zerocrossing) \in \{1, 0, -1\}$

or $mean(E_{up}, E_{low}) \cong 0$ then go to step 6

else $f(t) = res(t)$, and repeat steps 1-5

The $N(A)$ denotes the cardinality of set A. This step decides whether the difference between the number of zero crossings and extrema at most by one or whether $E_{mean}(t)$ is a zero mean function.

Step 6. Calculate

$imf_i = res(t)$

and

$f^{new}(t) = f(t) - res(t) = f(t) - imf_i$

Step 7. Repeat steps 1-6 while stopping criterion (ϵ) is not reached.

The stopping criterion is satisfied when

$$\sum_t \frac{(res(t) - f(t))^2}{f(t)^2} < \epsilon. \tag{4.11}$$

The EMD process is structured into an inner loop and an outer loop. The inner loop is composed of steps 1 to 5, where the condition at step 5 makes the process repeat to step 1 again. The outer loop is the process of steps 1 to 8, where the condition in step 5 occurs, and proceed to step 6.

The best output from EMD is where the $imfs$ are symmetrical to the x-axis, and the residual is very small. The EMD can extract data from the sifting process at high to low frequency.

The stopping criterion is based on the value of ϵ or by setting the maximum number of siftings (number of total iterations).

4.3.4 Complete ensemble EMD (CEEMD)

The critical issue that plagues EMD is mode mixing. This problem arises when EMD fails to execute the sifting process accurately, resulting in imf similar to the original data (figure 4.5).

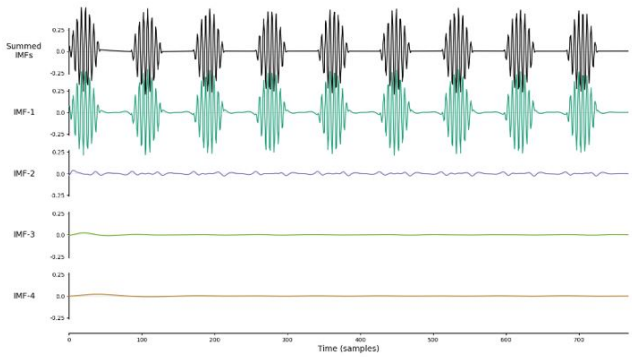


Figure 4.5: Sample of mode mixing (Gandhi and Ispány, 2024)

Wu and Huang (2009) addressed this issue by introducing ensemble EMD (EEMD). This method involves the dataset process multiple M times by adding white noise, $w_n(t) = \sigma \cdot N(0,1)$, as

$$x_t^i = x_t + w_t^i. \tag{4.12}$$

The formula (4.12) is used before step 1. In step 7, CEEMD use the following formulation

$$imf = \frac{1}{M} \sum_{i=1}^{M} imf_i. \tag{4.13}$$

This EEMD process solves the mode mixing problem, but there are shortcomings due to the addition of white noise (w_t^i) that leaves additional values in the dataset. So that there are differences between the original data and reconstructed signals. To overcome this problem, Yeh et al. (2010) introduced Complete EEMD or Complementary EEMD (CEEMD), which performs a critical process replacing equation 4.12. CEEMD generates two datasets by adding datasets with noise signals ($M_+(t)$) and subtracting datasets with noise signals ($M_-(t)$). As the process becomes

$$\begin{bmatrix} M_+(t) \\ M_-(t) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} f(t) \\ w(t) \end{bmatrix}. \tag{4.14}$$

At the end of step 7, we get imf_{+ji} and imf_{-ji}. After M repetition, we get

$$imf_i = \frac{1}{2M} \sum_{j=1}^{M} (imf_{+ji} + imf_{-ji}). \tag{4.15}$$

4.4 Deep learning (DL) models

4.4.1 Artificial neural network (ANN) model

Rochester et al. (1956) introduced artificial neural network (ANN) model, which was later developed by Rosenblatt (1958) with the invention of perceptron as part of ANN. Minsky and Papert (1969) developed multi-layer perceptron (MLP) with perceptron training inside and later applied it to the general case by Amari (1967) and Ivakhnenko and Lapa (1967). In 1986, Rumelhart et al. (1986) added back-propagation optimisers to ANN model. For later, ANN model is known as Neural Network (NN) or Multi-Layer Perceptron (MLP) model.

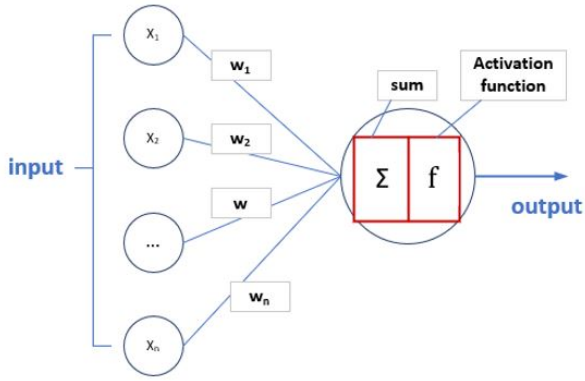


Figure 4.6: Single neuron/perceptron

Figure 4.6 is a representation of a single neuron. This single neuron, with other neurons, builds the ANN model. This single neuron will repeat the training process until it gets an acceptable weight vector. The initial weight vector ($\boldsymbol{w}$) is taken from random weights, and then linear combinations are calculated with the addition of bias (b) as follows

$$z = \sum_{i=1}^{n} w_i x_i + b_i = \boldsymbol{x}^T \boldsymbol{w} + b \tag{4.16}$$

then uses an activation function (ϕ) and the formula:

$$h(\boldsymbol{x}) = \phi(\boldsymbol{x}^T \boldsymbol{w} + b). \tag{4.17}$$

where the examples of ϕ are

a. Sigmoid function

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{4.18}$$

b. tanh function

$$tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{4.19}$$

c. Swish function

$$swish(x) = \frac{x}{1-e^{-x}} \tag{4.20}$$

The working principle of the ANN model is to combine several neurons in several layers and perform information transfer within them. This represents the human brain process, which consists of billions of neurons that work simultaneously to understand what they are dealing with. Figure 4.7 represents ANN model; input data will enter from the left side through the input layer, then all neurons in this layer will be connected to the hidden layer, and a linear process will occur. The hidden layers bring complex relationships to data sets, but excessive use of hidden layers will lead to overfitting. The number of neurons in the output layer for the forecasting case is determined based on the number of steps in time series, also called multi-output predictions.

Feed-forward is the process from the input to the output layer; there is no change in weights. After that, a back-propagation process will be carried out to correct each connection's weights ($\boldsymbol{w}$) and bias ($\boldsymbol{b}$) values due to the loss function $\mathcal{L}$ and learning rule.

This back-propagation process is also called an optimiser, where one that is often used is gradient descent:

$$w_{i,j}^{t+1} = w_{i,j} + \eta(\frac{\partial \mathcal{L}}{\partial w_t}), \tag{4.21}$$

where $w_{i,j}$ is the connection weight between the i^{th} neuron and j^{th} neuron, t is iteration step and η is learning rate. The w_t should depend on i and j.

Another optimiser using β as momentum rate is called Adamax. The equations are

$$w_{i,j}^{t+1} = w_{i,j} - \frac{\eta}{S_t}\hat{V}_t, \tag{4.22}$$

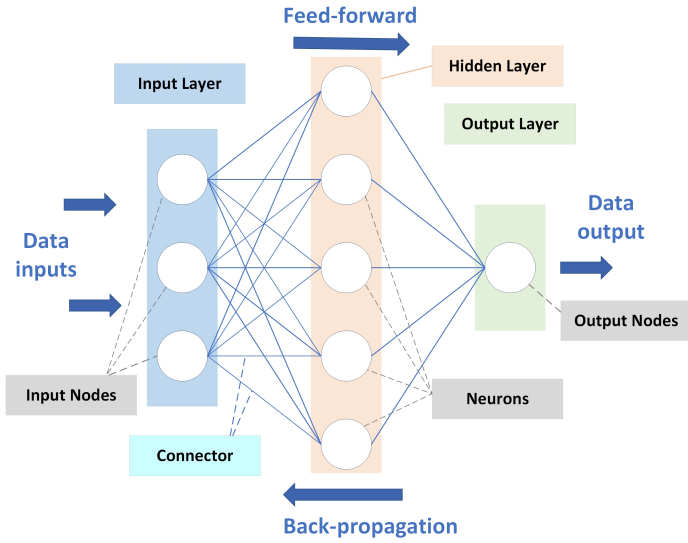


Figure 4.7: Artificial neural network

$$\hat{V}_t = \frac{V_t}{1-\beta^t}, \tag{4.23}$$

$$V_t = \beta V_{t-1} + (1-\beta)(\frac{\partial \mathcal{L}}{\partial w_t}), \tag{4.24}$$

and

$$S_t = \max\left(\beta S_{t-1}, \left|\frac{\partial \mathcal{L}}{\partial w_t}\right|)\right). \tag{4.25}$$

4.4.2 Long short-term memory (LSTM) model

Hochreiter and Schmidhuber first developed LSTM model in 1997. This model fixed the recurrent neural network (RNN) method, which has problems with exploding and vanishing gradients. LSTM model has long-term and short-term memory mechanisms that answer these two problems (Hochreiter and Schmidhuber, 1997).

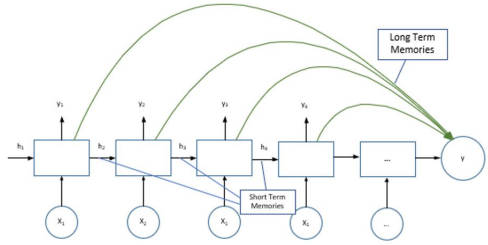


Figure 4.8: Long short-term memory idea

As Figure 4.8 shows, LSTM model simultaneously uses long-term dependencies and short-term memories. In addition, LSTM still has long-period information that influences the process's output. LSTM model has a memory that contains information about previous sequential data Brownlee (2018).

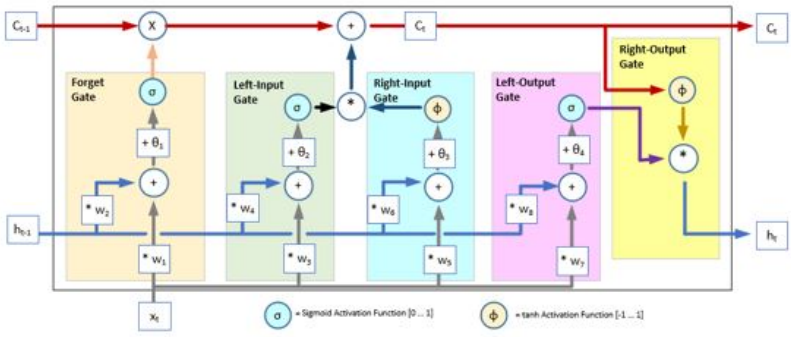


Figure 4.9: Long short-term memory model (Gandhi, 2024b)

Figure 4.9 shows a representation of an LSTM model. The red line at the top shows the cell state line or the long-term memory. The blue line at the bottom is the hidden state or short-term memory line. The first gate is the forget gate. As the first step in the LSTM model, this stage is to determine how much information will be given to the cell state (C) or what percentage of the long-term memory is remembered:

$$f_t = \sigma(\boldsymbol{w}_t \cdot [h_{t-1}, x_t] + \theta_1), \tag{4.26}$$

where h_{t-1} is the previous short-term memory, $\boldsymbol{w}_t$ is the weight vector of the

forget gate, which is a vector of $[w_1, w_2]$, and θ_1 is the bias at the forget gate. The σ is a sigmoid function (see equation 4.18).

The next gate is the input gate, which determines how much the potential memory (C) is to be kept:

$$i_t = \sigma(\boldsymbol{w}_i \cdot [h_{t-1}, x_t] + \theta_2), \tag{4.27}$$

$$\tilde{C}_t = tanh(\boldsymbol{w}_c \cdot [h_{t-1}, x_t] + \theta_3), \tag{4.28}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t. \tag{4.29}$$

where $\boldsymbol{w}_i$ is the weight of the left-input gate (w_3, w_4), θ_2 is the bias of the left-input gate. The C_{t-1} is the previous long-term memory, $\boldsymbol{w}_c$ is the weight of the right-input gate (w_5, w_6), and θ_3 is the bias of the right-input gate. The $tanh$ is described in equation 4.19.

Moreover, the output gate is the last gate used to determine the output value of the model. The formulas are

$$o_t = \sigma(\boldsymbol{w}_o \cdot [h_{t-1}, x_t] + \theta_4), \tag{4.30}$$

$$h_t = o_t * tanh(C_t). \tag{4.31}$$

where $\boldsymbol{w}_o$ is the weight of the left-output gate (w_7, w_8), and θ_4 is the bias of the left-output gate.

4.4.3 Convolutional neural network-LSTM (CNN-LSTM) model

CNN-LSTM model is the development of the LSTM model with a convolutional neural network (CNN) model before entering the LSTM. CNN model was developed by Fukushima (1980), which was previously taken from the thoughts of Hubel and Wiesel (1959) from research on visual neuron techniques from cats to vision on small parts. CNN model reduces the number of input data without reducing accuracy, but rather being able to facilitate reading the essential parts of the dataset. Some techniques used in CNN model are pooling, filtering, padding, shredding and flattening.

Forecasting uses a 1-dimensional CNN model, as shown in Figure 4.10. The benefit of the combination with LSTM model are [1] smaller input data. CNN-LSTM model reduces input dimension using n-dimension kernel tricks to become $\frac{T}{n}$ dimension. [2] CNN can identify features of each small shift of input data. By filtering and pooling the dataset, CNN tries to identify the small changes in every shift of the dataset sequence.

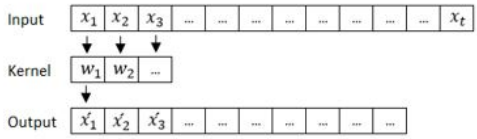


Figure 4.10: One-dimensional CNN model (Gandhi, 2024b)

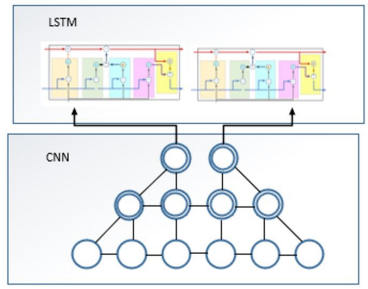


Figure 4.11: CNN-LSTM (Gandhi, 2024b)

The mechanism of CNN-LSTM model is as shown in Figure 4.11, where for the dataset $\{x_1, x_2, \ldots, x_T\}$, the number of datasets will be reduced to $\{\acute{x}_1, \acute{x}_2, \ldots, \acute{x}_n\}$ where $n < T$ by the convolutional process.

Then, it enters the n-models of LSTM model. The training process will continue with the backpropagation mechanism to improve the weights and biases in the model.

4.4.4 Proposed hybrid decomposition model

I do a split (80:20) where the dataset is split into 80% training and 20% testing. The dataset enters the decomposition process in Figure 4.12 and splits into multiple $imfs$. The imf_i enters the forecasting model, which is a deep learning (DL) model. After the forecasting model predicts, the results are combined into one number with simple addition. For more details, see the following Algorithm 2.

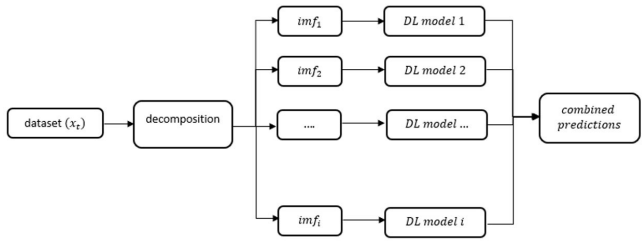


Figure 4.12: Hybrid decomposition (Gandhi and Ispány, 2024)

Algorithm 2 Hybrid decomposition

Input: $\{x_t\}$, n_{test}, n_{train}, n_{step}
Output: prediction = $[\hat{x}_{t+1}, \hat{x}_{t+2}, \ldots, \hat{x}_{t+n_{step}}]$

train $\leftarrow x_t[: -n_{test}]$
prediction $\leftarrow [\,]$
for h in range $(1, n_{test})$ **do**
    $imfs \leftarrow [\,]$
    $total\_xhat \leftarrow [n_{step}]$ of zeros
    If $h \ \% \ n_{step} \ != 0$ **then** continue
    else
    $imf_i \leftarrow$ decompose(train)
    $imfs.insert(imf_i)$
    for j in range $(1, \text{len}(imfs))$ **do**
        forecasting.fit$(imf_j[: -n_{step}], imf_j[-n_{step} :])$
        $xhat \leftarrow$ forecasting.predict(n_{step})
        $total\_xhat$ += xhat
    end for
    prediction.append$(total_xhat)$
    train = $x_t[: (-n_{test} + h)]$
end for

4.5 Datasets

4.5.1 Data description

The global energy market is significantly influenced by two pivotal sources of oil prices: Brent (BCOP) and West Texas Intermediate (WTI) crude oil price (WCOP). Brent crude oil, extracted from the North Sea in Europe, including Oseberg, Ekofisk, and Forties, and WTI crude oil, sourced from Texas, Louisiana, and North Dakota, United States. The dataset (USD/barrel) I utilise, obtained from Yahoo Finance using the yfinance library in Python 3.8, spans from 4th January 2000 to 17th June 2022 daily (working days), focusing on the close spot oil price category.

Line plots of both crude spot oil prices are shown in Figures 4.13 and 4.14. Both datasets show similarities in day-by-day movements. The beginning (2000 to 2003) is one of stability at the lowest state, and then until 2008, there is an increase to the highest value. After that, the value dived sharply to a low point again at the end of 2008. Then, the price rose again to around USD $120 in mid-2011. This price stayed in the range of USD $100-120 until 2014. At the end of the period (2014-2022), the price movement again showed fluctuations that were difficult to recognise. The movement of the dataset is complex, irregular, and highly volatile, presenting a significant challenge in its analysis. This irregularity and complexity call for advanced analysis techniques, as suggested by the references presented at the beginning of Chapter 2.

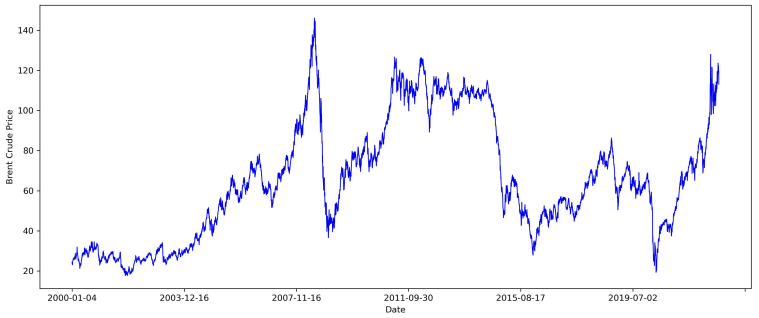


Figure 4.13: BCOP line graph

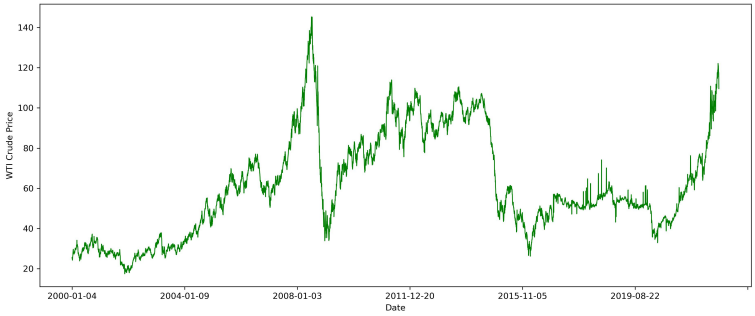


Figure 4.14: WCOP line graph

Natural gas price (NGP) data comes from Trading Economics (2023). This data is based on close prices from Henry Hub in Louisiana, United States, in 10,000 million British thermal units (USD/mmBtu). The price is interstate natural gas pipeline system contracts using a gas pipeline system traded. The dataset ranges from January 2000 to June 2022 using working days. The line plot can be seen in Figure 4.15.

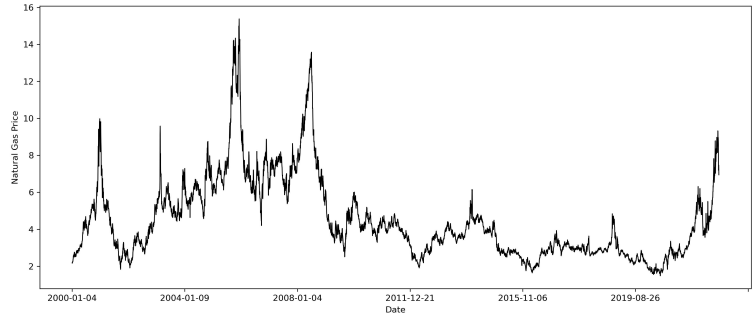


Figure 4.15: NGP line graph

A heating oil price (HOP) dataset based on two sources, financial contracts and over-the-counter (OTC) in USD/GAL units, comes from Yahoo Finance

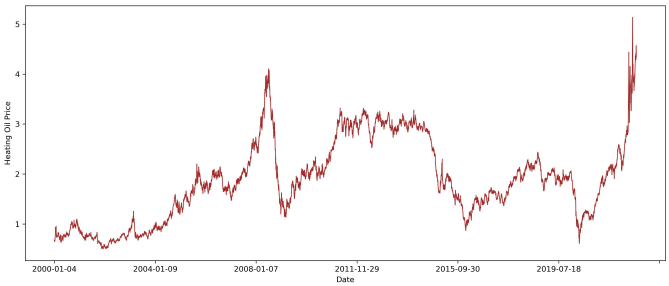


Figure 4.16: HOP line graph

(2023). However, the data can also be retrieved via the yfinance library in Python. The data is on working days from 4th January 2000 to 17th June 2022.

The movements of the two datasets (NGP and HOP) show inconsistencies, high volatility, and difficult-to-detect fluctuations.

4.5.2 Descriptive statistics

Table 4.2 collects all the descriptive statistics of the four datasets. BCOP and WCOP have close values each other, where the mean BCOP (65.32) is greater than WCOP (61.32). BCOP standard deviation (29.38) is also higher than WCOP (25.57).

Table 4.2: Descriptive statistics for energy prices

Attributes	BCOP	WCOP	NGP	HOP
Count	5768	5744	5742	5770
Mean	65.32	61.35	4.48	1.87
Std Dev	29.38	25.57	2.22	0.81
min	17.68	17.45	1.48	0.49
Q1	42.72	42.87	2.85	1.24
median	62.33	55.57	3.87	1.84
Q3	84.06	81.24	5.56	2.38
max	146.08	145.29	15.38	5.13

NGP has a smaller unit value than the oil price. The NGP dataset has a mean of 4.48 and a standard deviation of 2.22, but the highest value is almost three times the mean value of 15.38. Q3 of NGP is worth 5.56. The movement of NGP in Figure 4.15 is much different from that of BCOP, WCOP, and HOP.

Finally, HOP has the smallest data range among these four datasets: max=5.13 and min=0.49. Its mean is 1.87, and standard deviation=0.81.

4.5.3 Decomposition results

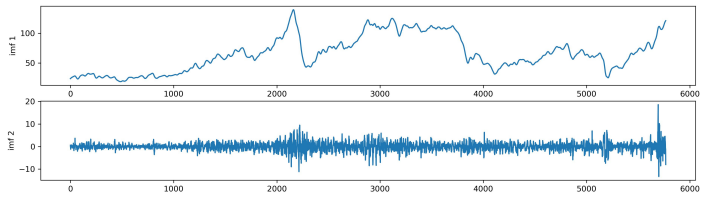


Figure 4.17: HP decomposition of BCOP

This chapter uses the BCOP for the decomposition representation. The other three datasets are in Appendix C. Hodrick-Prescott decomposition (HPD) gives two results commonly expressed as trend and cyclic. As shown in Figure 4.17, trend HPD is in $imf1$, and cyclic is in $imf2$.

For wavelet decomposition (WD), I initially transform the dataset as shown in Figure 4.18. I delete five transformed results with the lowest amplitude ($\cong$ 0) and then retransform them into WD, as shown in Figure 4.19. Finally, I get two $imfs$ from WD (Figure 4.19). $Imf1$ is the retransform process from Figure 4.18, while $imf2$ is the residual from WD ($imf1$).

The result of EMD decomposition is shown in Figure 4.20. The result of EMD gives ten $imfs$ in the case of BCOP. They start from high frequency ($imf1$) to low frequency ($imf10$). The number of $imfs$ can differ from other datasets; more details are in Appendix C. All of these $imfs$ will be input to the forecasting model, and EMD has no residuals because the number of each t in $imf_{i.t}$ corresponds to the value of x_t.

The result of CEEMD on BCOP is shown in Figure 4.21 with eight $imfs$. The sum of every $imf_{i.t}$ corresponds to the value of x_t.

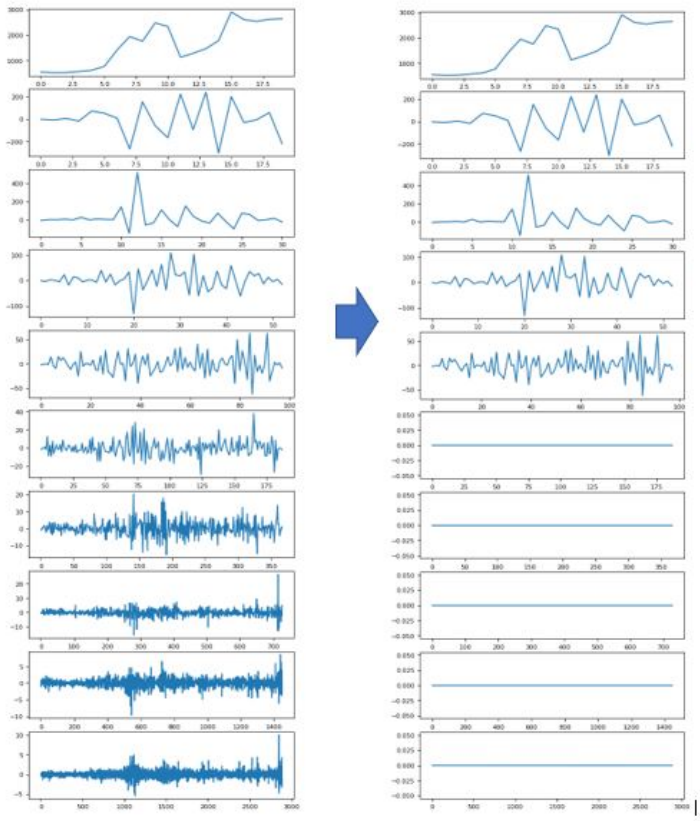


Figure 4.18: Breakdown BCOP using WD

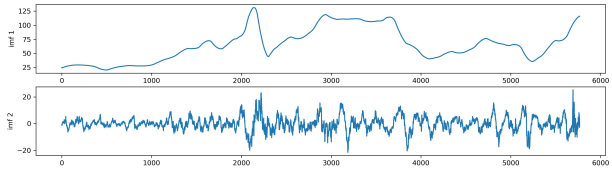


Figure 4.19: Reconstruct WD of BCOP

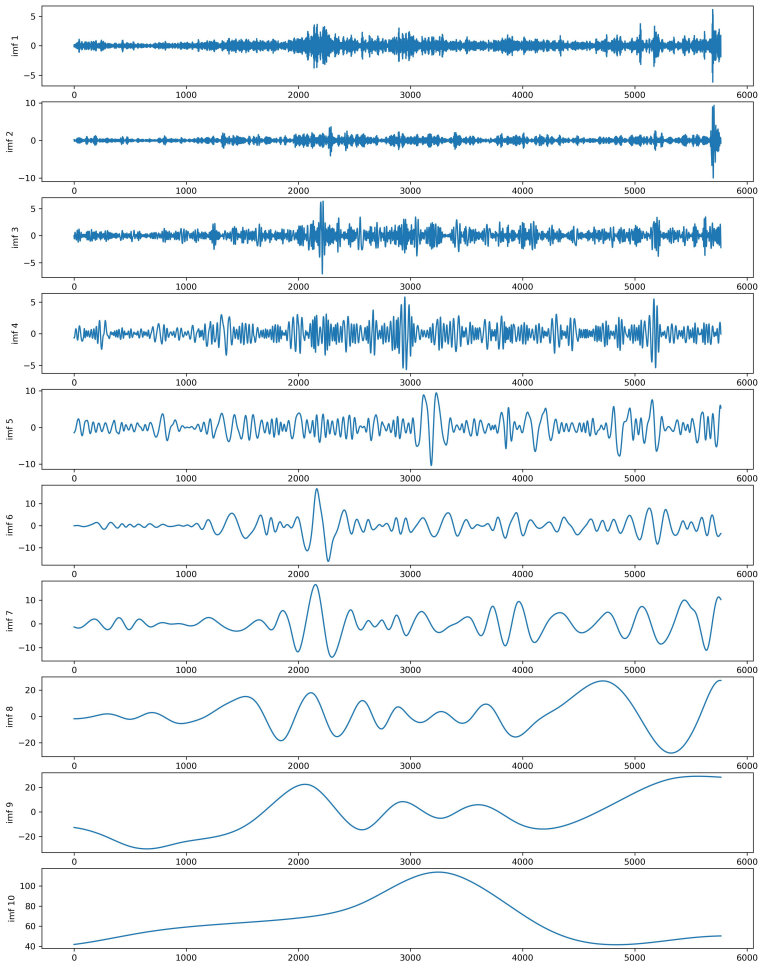


Figure 4.20: EMD decomposition of BCOP

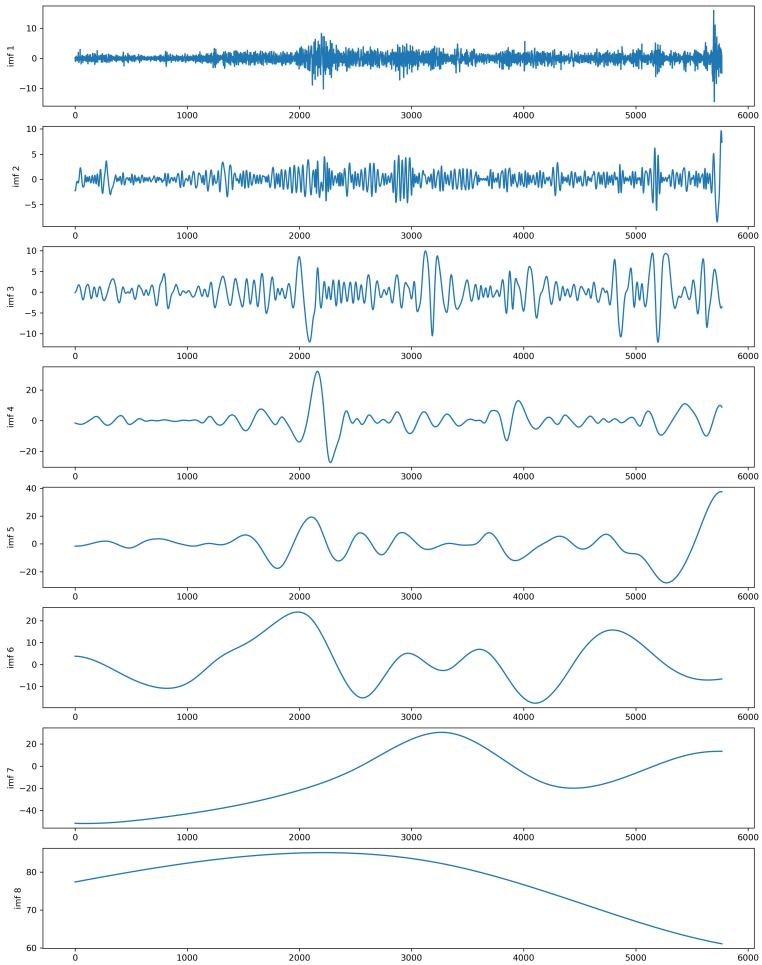


Figure 4.21: CEEMD decomposition of BCOP

4.5.4 Multi-step forecasting

As shown in Figure 4.22, I use multiple output forecasting (Lazzeri, 2020). With the training dataset $\{x_1, x_2, \ldots, x_T\}$, we get n outputs for one iteration, which is $\{\hat{x}_{T+1}, \hat{x}_{T+2}, \ldots, \hat{x}_{T+n}\}$. For the next step, we input n actual data into the

training dataset $\{x_1, x_2, \ldots, x_{T+n}\}$ for the next prediction. And so on until the end of the prediction. The process indicates that there are n neurons in the output layer of the neural network.

Figure 4.22: Multiple output forecasting

4.5.4.1 Measurement error calculations

I use the five error measurements (MAE, MAPE, RMSE, MSE and R2) described in section 3.2.10. Lewis (1981) made a classification based on the MAPE value, where (< 10) is highly forecast, ($10 \leq$ MAPE < 20) is good forecast, ($20 \leq$ MAPE < 50) is reasonable, and ($50 \leq$ MAPE) is an inaccurate forecast.

4.6 Results and discussion

The discussion in this chapter uses four tools to analyse the accuracy and robustness of the forecasting models for four datasets. [1] Line plot error measurement: This tool represents how low the parameter error is for each n-step forecast. [2] Bar chart decreasing error: The figures show how many percentages the best models MAE outperforms other decompositions. The higher the bar chart shows, the better the best model is than the benchmark model. [3] Box plot analysis, I simulated the best models with 30 repetitions. The repetitions are due to the deep learning model's trait, which gives different output values for every prediction.

4.6.1 Analysis BCOP

As seen in Figure 4.23, the MAE of CEEMD-CNN-LSTM model gives a lower line plot than the MAE of other models. Likewise, the plots of CEEMD-ANN and CEEMD-LSTM models have relatively low MAE but not lower than CEEMD-CNN-LSTM model. All MAE values of the models with no decomposition and with HPD show large gap values at the top of the line plot. In

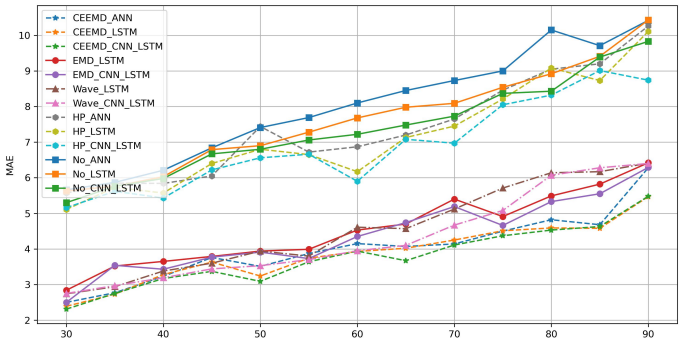


Figure 4.23: Benchmark models using MAE for BCOP

comparison, EMD and WD show slightly bigger values than CEEMD decomposition. Nevertheless, even so, CEEMD-CNN-LSTM provides the lowest MAE value.

For the RMSE comparison, CEEMD-CNN-LSTM model still shows the lowest measurement error values. However, in steps 40 and 45, WD-CNN-LSTM model provides a lower value than the best model. From Figure 4.24, 'no-decomposition' results in a much higher measurement error than the best model.

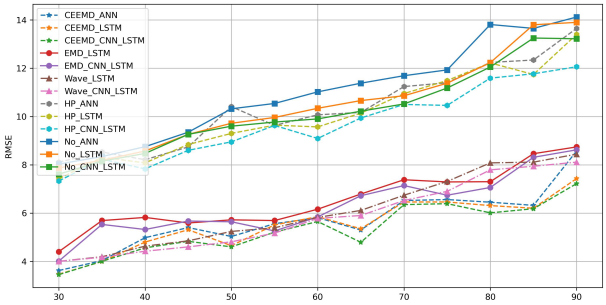


Figure 4.24: Benchmark models using RMSE for BCOP

Figure 4.25 compares the best model (CEEMD-CNN-LSTM) with the same model (CNN-LSTM) but with different decomposition techniques at each forecasting step. We see at all steps that CEEMD can reduce the MAE value to

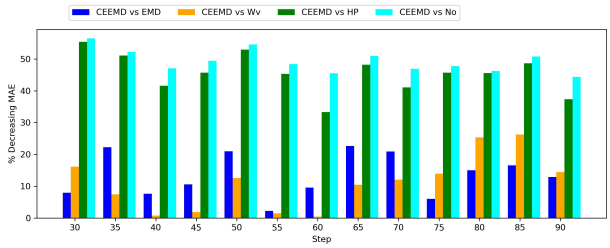


Figure 4.25: Percentage decreasing MAE of the Models for BCOP

around 50% against using 'no-decomposition'. Compared with HP decomposition, replacing with CEEMD results in a reduction of around 30-55%. For the comparison with EMD and WD, the MAE reduction at some steps gives values that fluctuate between 0-30%. The table of percentage decreasing MAE can be seen in Appendix D.

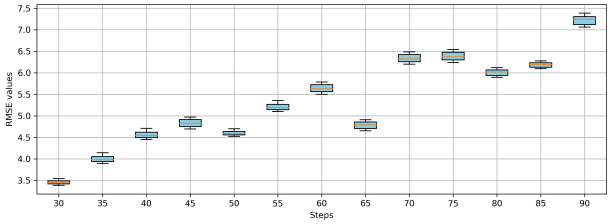


Figure 4.26: BoxPlot RMSE of CEEMD-CNN-LSTM for BCOP

Figure 4.26 shows the RMSE outputs of 30 simulations from the best model (CEEMD-CNN-LSTM) in a boxplot graph. I take the mean of these simulations as the value in Figure 4.24. The box plot of all steps in Figure 4.26 has an interquartile range (IQR) ranging from 0.07-0.18, where the IQR increases for every step between 0.1 and 0.18. Meanwhile, the difference between the lower and upper whiskers ranges from 0.21-0.6. With IQR and difference whisker values, the difference in output from each iteration of the forecasting model is still relatively low and tolerable. The percentage comparison between the difference whiskers and the mean is below 10%. Detailed values are in Appendix E.

Figure 4.27 shows the line plot of prediction vs. actual. The prediction

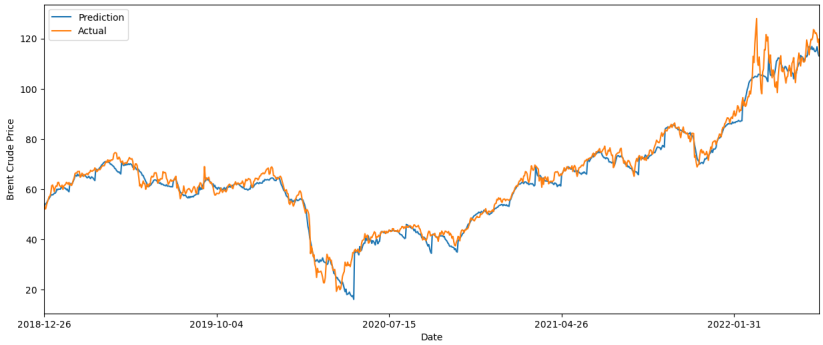


Figure 4.27: Line Plot of Actual vs Prediction (BCOP) using CEEMD-CNN-LSTM

(blue line) can follow the movement of the actual data (yellow line). The use of CEEMD-CNN-LSTM model can follow the movement even with multi-step forecasting.

4.6.2 Analysis WCOP

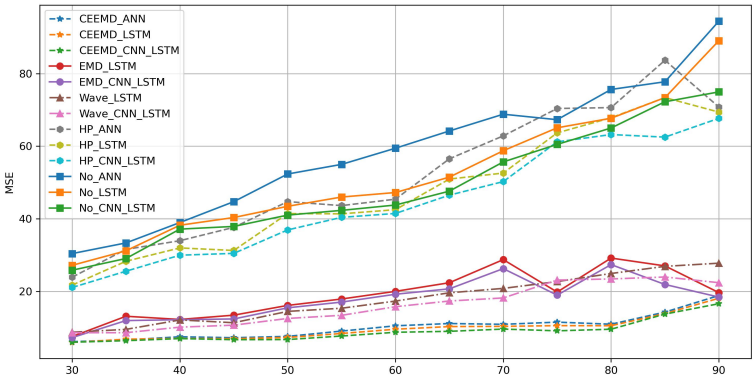


Figure 4.28: Benchmark models using MSE for WCOP

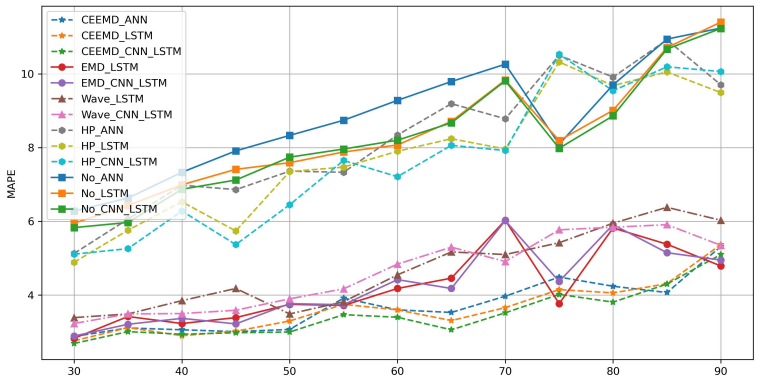


Figure 4.29: Benchmark models using MAPE for WCOP

For WCOP, Figure 4.28 shows that the three models with CEEMD decomposition (CEEMD-ANN, CEEMD-LSTM and CEEMD-CNN-LSTM models) show low MSE values compared to the other models. Like the previous dataset, CEEMD-CNN-LSTM model provides the lowest value. EMD and WD models have larger MSE values than the CEEMD model, while HP and No Decomposition have much larger MSE values.

As for Figure 4.29, according to the classification proposed by Lewis (1981) (see chapter 4.5.4.1), the MAPE of CEEMD, EMD and WD are below 10 for 30 to 90 steps. So, these three decompositions are classified as highly forecast. As for HP and No decomposition at 75 to 90 steps shows MAPE above 10, so this model is categorised as good forecasting.

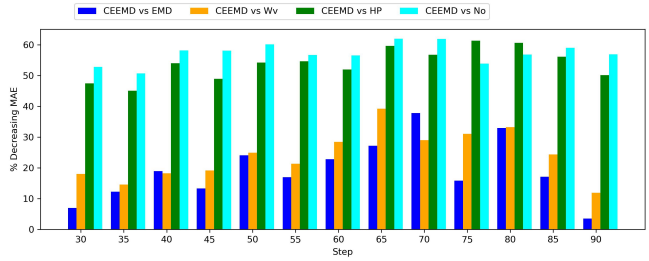


Figure 4.30: Percentage decreasing MAE of the models for WCOP

The percentage decreasing MAE between CEEMD-CNN-LSTM model and other decompositions (Figure 4.30) is quite similar to Figure 4.25. Decreasing MAE against using no Decomposition and HPD ranges from 45 - 60%, while against using WD is between 10-40%. For EMD, replacing with CEEMD outperforms by reducing MAE by 3-30%.

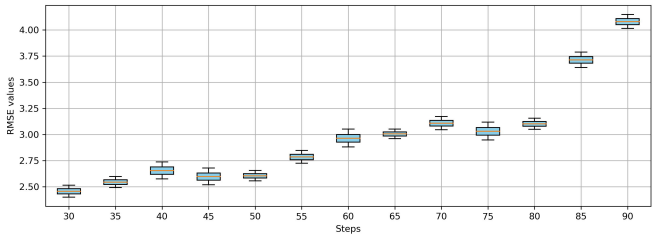


Figure 4.31: BoxPlot RMSE of CEEMD-CNN-LSTM for WCOP

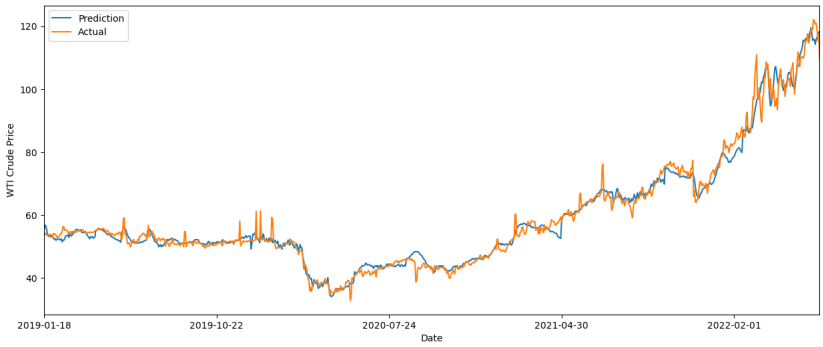


Figure 4.32: Line Plot of Actual vs Prediction (WCOP) using CEEMD-CNN-LSTM

The box plot analysis in Figure 4.31 shows an IQR between 0.04 and 0.08, with whiskers differences between 0.1 and 0.22. With the mean of each step ranging between 2.4 and 4.1, the percentage ratio between the whiskers differences and the means range from 5 to 8%. This range of percentage values is still relatively low. Figure 4.32 shows the plot of predictions from CEEMD-CNN-LSTM model and actual data.

4.6.3 Analysis NGP

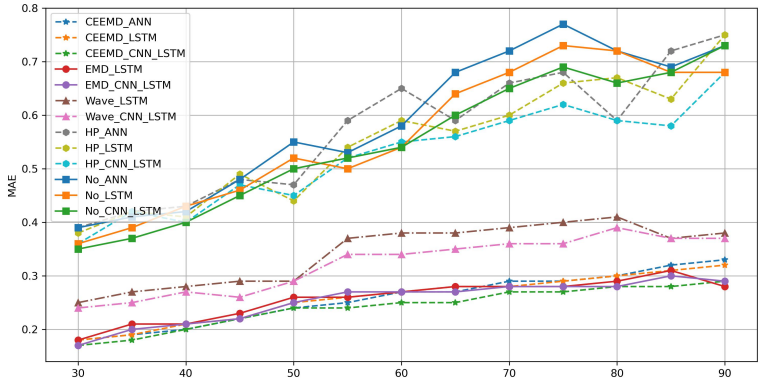


Figure 4.33: Benchmark models using MAE for NGP

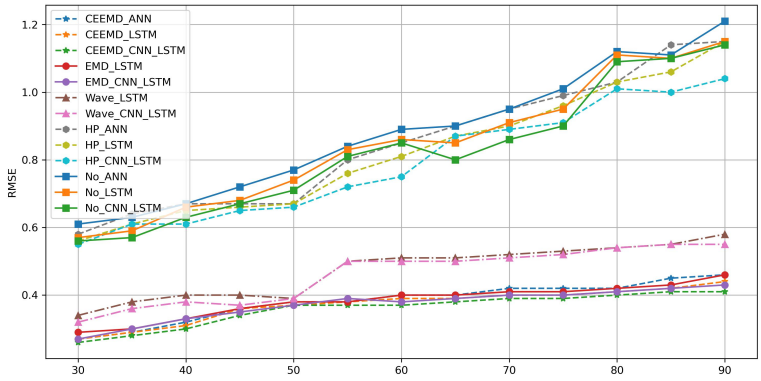


Figure 4.34: Benchmark models using RMSE for NGP

The MAE value of Natural Gas Price (Figure 4.33) shows that the CEEMD and EMD models have much lower values than other decompositions. WD has a slightly larger MAE value, while HPD and No Decomposition have much larger values. On this dataset, EMD can outperform CEEMD at some steps, but the CEEMD-CNN-LSTM value still provides more accurate values. The MAE results do not differ from the RMSE characteristics in this case.

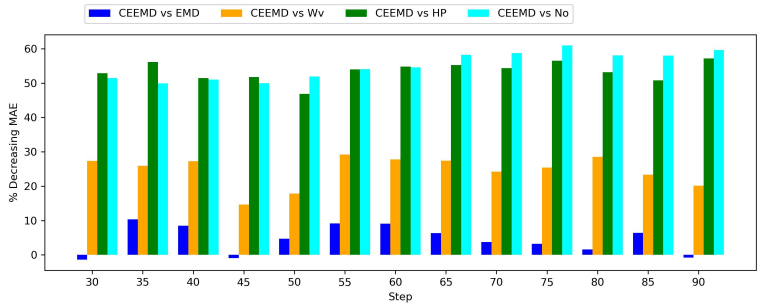


Figure 4.35: Percentage decreasing MAE of the models for NGP

In NGP, in some steps (30, 45, and 90), EMD gives better results than CEEMD, with decreased percentages in MAE below 2%. But for other steps, CEEMD outpowers EMD, with decreased percentages between 1-12%. Meanwhile, with WD, the percentages are between 15-30%. CEEMD outpowers HP and No-Decomposition with percentages above 50%.

Box plot analysis shows that the Q2 value moves from 0.26 (step 30) to 0.41 (step 90). It is also shown in Figure 4.36, where the RMSE's mean is increasing. The percentage ratio of the whisker's length to the mean RMSE ranges from 4 to 8%. The details of the boxplot values are in Appendix E. The prediction results of CEEMD-CNN-LSTM model are depicted in Figure 4.37.

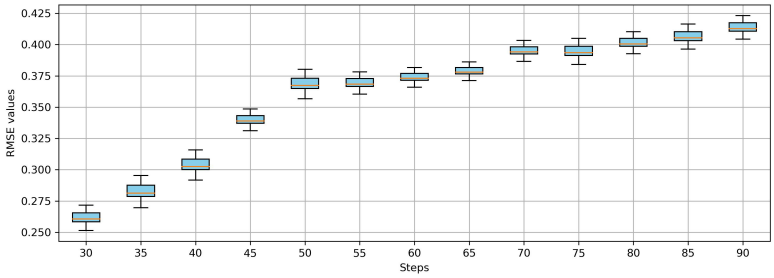


Figure 4.36: BoxPlot RMSE of CEEMD-CNN-LSTM for NGP

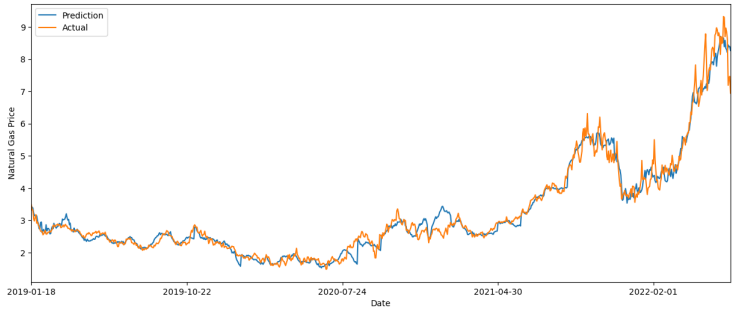


Figure 4.37: Line Plot of Actual vs Prediction (NGP) using CEEMD-CNN-LSTM

4.6.4 Analysis HOP

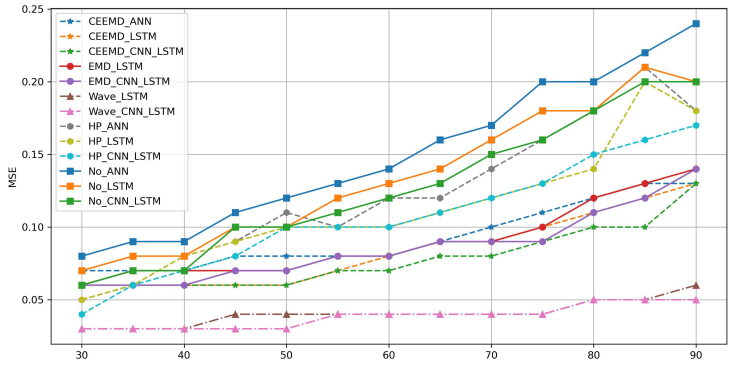


Figure 4.38: Benchmark models using MSE for HOP

The HOP results differ from those of the previous three datasets. Figure 4.38 shows that WD gives much better results than the other four decompositions in all error measurements. CEEMD, EMD, and HPD have results that are close to each other from various steps. In the MAPE value, Figure 4.39, only WD provides a value below 10, highly forecasting. The other four models have a majority of MAPE values above 10. It is less accurate than the other decomposition models for no decomposition.

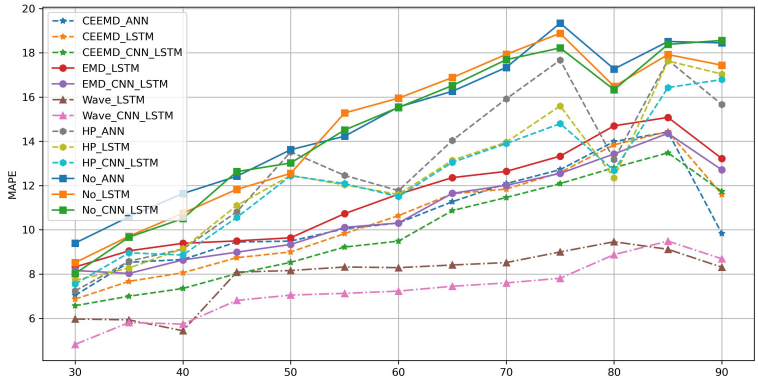


Figure 4.39: Benchmark models using MAPE for HOP

In HOP, where WD-CNN-LSTM model provides the best forecasting compared to other models, the percentages decreased in the MAE shown in the bar chart in Figure 4.40. WD decreases MAE by 25-40% compared to using EMD. As for CEEMD, using WD can provide a decrease between 15-35%. Likewise, compared to HPD and no-decomposition, replacement using WD provides a decrease between 30-55%.

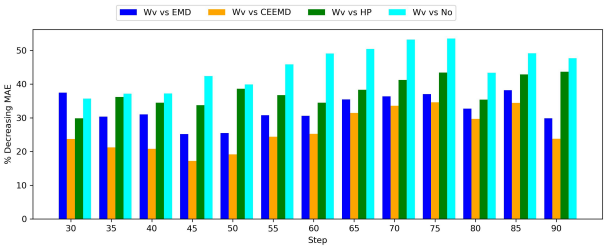


Figure 4.40: Percentage decreasing MAE of the models for HOP

In general, the simulation results of the best model (WD-CNN-LSTM) with boxplots show that the variation of values (shown in whiskers) still has a low range, ranging between 0.01 and 0.02. The means are between 0.16 and 0.23, and the percentage ratio of the whisker length to the mean is between 6 and 8%. Figure 4.42 shows the prediction results of the best model.

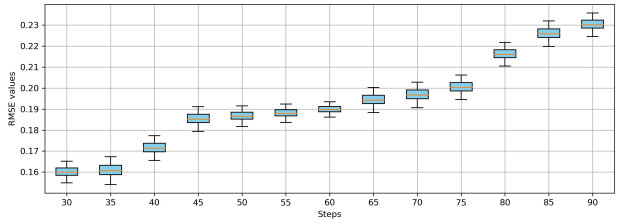


Figure 4.41: BoxPlot RMSE of WD-CNN-LSTM for HOP

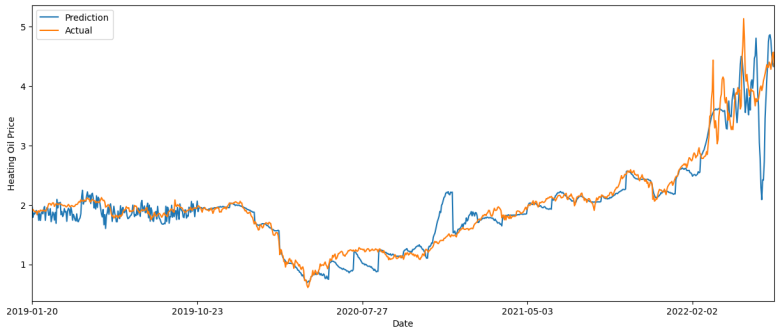


Figure 4.42: Line Plot of Actual vs Prediction (HOP) using WD-CNN-LSTM

4.7 Conclusion

I propose a hybrid decomposition forecasting (HDF) and deep learning (DL) models using four decomposition techniques and three popular DL models. Decomposition techniques can split the dataset's complexity into imf, making it easier to understand the movement of the dataset. The techniques are also relatively simple. For EMD and CEEMD decomposition, this technique can generate more than two $imfs$ with results ranging from low to high frequency.

For analysis, I use mid-term multi-step forecasting. Using the multiple output forecasting method, I take 30 to 90 steps with five intervals for each step. Three out of four datasets (Brent oil, WTI crude oil, and Natural Gas) show that the CEEMD-CNN-LSTM model is better than the other benchmark models. The CEEMD decomposition technique proves to be much better than other

decompositions. At the same time, using EMD and WD have better results than HP and no decomposition but not better than CEEMD decomposition. CEEMD can reduce the percentage of MAE values between 30 and 50% compared to HPD or no decomposition. Meanwhile, for WD or EMD, CEEMD can reduce MAE values by 0 to 20%. I repeated the best model to do 30 simulations. The best model provides a variation of the RMSE value that ranges from 5 to 8% of the mean RMSE value. These percentages indicate a low variance of the RMSE value for each simulation.

For the HOP dataset, WD-CNN-LSTM model shows better results than the three decomposition models for the entire n-step forecasting. CEEMD and EMD decomposition are the second-best models, HP decomposition is third, and no decomposition gives the worst results. The use of WD decreases MAE for CEEMD and EMD by 20 to 40%, while for HP and no decomposition, WD can decrease MAE between 30% and 50%. The best model prediction results for all datasets provide values close to the actual data.

Thesis 3

I am concerned about price forecasting and have successfully applied a novel combination of decomposition techniques and deep learning models to predict four datasets of daily energy prices. The decompositions could split the datasets into several $imfs$ with different frequencies. I compare using paired models of four decomposition techniques and three deep learning models. I also perform mid-level multi-step forecasting, where the prediction ranges from 30 to 90 days ahead.

Related publications:

- Gandhi, Herry Kartika, and Ispány Márton. Multi-step Natural Gas Price Forecasting using Ensemble Empirical Mode Decomposition and Long Short-Term Memory Hybrid Model. International Journal of Energy Economics and Policy. 14(4) (2024), 590–598. doi:10.32479/ijeep.16053. (SJR: Q2)
- Gandhi, Herry Kartika. Mid-term forecasting of crude oil prices using hybrid CEEMDAN and CNN-LSTM deep learning model. Polityka Energetyczna. 27(4) (2024), 19-38, doi: 10.33223/epj/190486. (SJR: Q3)

Chapter 5

Conclusions

This dissertation proposes novel variations of combining several methods into hybrid forecasting models to increase the forecasting accuracy and decrease the error of predictions within the scope of problems in heavy manufacturing industries. I classify the problems into three groups: [1] quality forecasting, [2] demand forecasting, and [3] price forecasting. For applied forecasting models, I group them into [1] non-negative discrete forecasting, [2] hybrid linear and nonlinear methods, and [3] hybrid decomposition methods. As forecasting techniques, I use popular methods derived from statistics and machine learning. This dissertation aims to provide various forecasting techniques beyond the single forecasting we often get today, with a strong focus on practical application in heavy manufacturing industries. This dissertation uses several statistical techniques, such as preprocessing, error measurements, multi-step forecasting, and benchmark testing. As dataset objects, I take sources from packaging brown paper, electrical, and crude oil industries with variant products, natural gas and heating oil.

The first study is quality forecasting, which I applied to paper roll defect products in the paper industry. This dataset is a non-negative integer-valued, which comes from the number of daily paper roll defects after leaving the paper machine and is caused by uncontrollable factors. I take data from three different paper machines. The results of INAR(1) will be adjusted to the characteristics of PINAR(1) model or NBINAR(1) model according to their corresponding characteristics. Through model fitting analysis, the PINAR(1) and NBINAR(1) models are shown to fit the actual probability dataset. Model fitting is done by adjusting the model characteristics to the data's index of dispersion (I), p_0,

and Poisson value parameter values. In addition, I also use Probability Integral Transform (PIT) visualisation to support the selection of the suitable model. The selected model is proven to have lower AIC and BIC error values than other models. The forecasting results can show the performance of each paper machine in terms of product quality and compare it with other machines.

The second study is demand forecasting, which I take from daily electrical consumption in the eastern region of the United States. The data shows continuous numbers of units of 1/20,000 Mega Watts. Pre-processing with the ACF plot shows that the data have an indication of seasonality, so using SARIMA model at the beginning of the process is an appropriate tool. I use a hybrid linear and nonlinear model, with SARIMA model as the first (linear method) and SVR as the second (nonlinear) model. This combination technique involves estimating innovations ($\hat{\varepsilon}_t$) by SVR model, which later will be added with the predictions of SARIMA model ($\hat{x}_t$). SARIMA-SVR model proved to outpower other models in (1, 2, 3)-step ahead forecasting compared to benchmark models using five error measurement parameters. The proposed model is significantly better than benchmark models using the DM test, with a significance level $\alpha = 0.05$. However, for longer steps (5, 10, 15), SARIMA-GARCH model provides better accuracy than SARIMA-SVR model. The linear and nonlinear hybrid model can handle the non-normality and heteroscedasticity properties of SARIMA residuals. Besides that, there is still an indication of dependency on SARIMA residuals using the Ljung Box Test. According to the line plot between SARIMA-SVR's prediction and the actual, it can be seen that the prediction value is very close to the actual data.

The third study is price forecasting, which highlights energy prices as one of the vital sources for all manufacturing industries. I use the four highest priority energy prices datasets: Brent and WTI crude oil, natural gas, and heating oil. Energy prices have high complexity, irregularity and volatility, so handling decomposition at the beginning of the process can reduce complexity before entering the forecasting model. The result of decomposition shows that the $imfs$ are formed from low to high frequency. For the forecasting model, I use three popular techniques from deep learning. Deep learning can understand both linear and nonlinear data. Hybrid decomposition has been proven to give much lower measurement errors than no decomposition for all forecasting models. In three cases, CEEMD decomposition gives good results compared to all decompositions, while in heating oil price, wavelet decomposition is the best among others. The best model can reduce MAE values ranging from 10 to 50% compared to using other decompositions. Although deep learning provides outputs with different results, with 30 simulations performed from each step,

the changes in RMSE values are only 5 to 8% around the mean RMSE, which is still relatively low. The prediction and actual in one line plot show that the prediction is close to the actual value.

In the future, hybrid forecasting could develop in other ways using different techniques and can be options for developing forecasting models other than single forecasting.

Acknowledgement

All praise to Allah the Almighty, and the Most Gracious for His blessing given to me strength to complete my Phd study and completing the dissertation. May Allah's blessing goes to His final Prophet Muhammad (peace be up on him), his family and his companions.

I would like to express my gratitude and sincere thanks to my supervisor Prof. Dr. Márton Ispány, who has given me invaluable knowledge, guidance, suggestions and encouragement so I could complete this study. I am extremely grateful to him for his belief in my research potential and for providing me with opportunities to present my work at conferences and publish in respected journals. I also wish to thank Prof. György Terdik, who was a major influence on my significant development during my early studies.

I want to express my heartfelt gratitude to my parents, Supriyanto Notohadisuwarno and Endang Wahyu Pertiwi, my sister and her family, who always support me with endless love and prayers. Your love and support have given me the strength to pursue my dreams. To my beloved wife, Meri Wahid, I really appreciate your unwavering support and love. To my dearest children, thanks to all of you, I have become better and stronger than I ever thought possible. I would like to extend my sincere thanks to my wife's family in Bukit Tinggi, words cannot express how thankful I am to have such a loving family.

I would like to thank all my colleagues at the University of Debrecen as well, you make the Phd rooms brighter and whole lot of fun. Also, I would like to recognize the assistance that I received from Indonesian students in Debrecen. Wishing you a lot of success in whatever you do. I cannot fail to mention my deep respect for the University of Banten Jaya for supporting me in pursuing a higher level of study at the University of Debrecen.

Finally, I express my deep gratitude for the privilege of attending the prestigious University of Debrecen. This journey would not have been attainable without the financial assistance from the Stipendium Hungaricum Scholarship.

List of publications

Journal papers

Gandhi, Herry Kartika, and Ispány, Márton. Analyzing uncontrollable factors that cause defective products by Poisson and negative binomial INAR(1) for fitting model. *Proceedings on Engineering*. doi:10.36055/jiss.v5i1.6494. (**SJR: Q4**) [Status: Accepted]

Gandhi, Herry Kartika, and Ispány, Márton. Multi-step Natural Gas Price Forecasting using Ensemble Empirical Mode Decomposition and Long Short-Term Memory Hybrid Model. *International Journal of Energy Economics and Policy*. **14(4)** (2024), 590–598. doi:10.32479/ijeep.16053. (**SJR: Q2**)

Gandhi, Herry Kartika. Mid-term forecasting of crude oil prices using hybrid CEEMDAN and CNN-LSTM deep learning model. *Polityka Energetyczna*. 27(4) (2024), 19-38, doi: 10.33223/epj/190486. (**SJR: Q3**)

Conferences

Gandhi, Herry Kartika. Application of time-series method in company revenue. *The 2022 IEEE 2nd Conference on Information Technology and Data Science (CITDS 2022)*, Debrecen, Hungary, May 16 – 18, 2022. (Online Conference) [Note: Presentation only]

Gandhi, Herry Kartika. Applying hybrid forecasting model SARIMA-SVR for daily energy consumption data. *The 2024 IEEE 3rd Conference on Information Technology and Data Science (CITDS 2024)*, Debrecen, Hungary, 2024, pp. 1-6, doi: 10.1109/CITDS62610.2024.10791394.

Bibliography

Aamir, M., Shabri, A., and Ishaq, M. (2018). Crude oil price forecasting by CEEMDAN based hybrid model of ARIMA and Kalman filter. *Jurnal Teknologi*, 80(4):67–79.

Abdollahi, H. (2020). A novel hybrid model for forecasting crude oil price based on time series decomposition. *Applied Energy*, 267:115035.

Abolghasemi, M., Beh, E., Tarr, G., and Gerlach, R. (2020). Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion. *Computers & Industrial Engineering*, 142:106380.

Ahmad, W., Aamir, M., Khalil, U., Ishaq, M., Iqbal, N., and Khan, M. (2021). A new approach for forecasting crude oil prices using median ensemble empirical mode decomposition and group method of data handling. *Mathematical Problems in Engineering*, 2021(1):5589717.

Alberg, D. and Last, M. (2018). Short-term load forecasting in smart meters with sliding window-based ARIMA algorithms. *Vietnam Journal of Computer Science*, 5(3-4):241–249.

Amari, S. (1967). A theory of adaptive pattern classifiers. *IEEE Transactions on Electronic Computers*, EC-16(3):299–307.

Ampountolas, A. (2021). Modeling and forecasting daily hotel demand: A comparison based on SARIMAX, neural networks, and GARCH models. *Forecasting*, 3(3):580–595.

Atmanegara, E., Suhartono, and Atok, R. M. (2019). Forecasting foreign tourist using intervention analysis on count time series. In *IOP Conference Series: Materials Science and Engineering*, page 052014. IOP Publishing.

Barney, N. and Courtemanche, M. (2023). Load shedding. https://corporatefinanceinstitute.com/resources/economics/heavy-industry/ [Accessed on 8 January 2024].

Bhatnagar, A., Gupta, R., and Thakar, G. D. (2020). Prediction of demand for supply chain using time series predictive models. *International Journal of Computer Applications*, 175(22):33–39.

Bhattar, P. (2018). Blackout economics. https://www.wartsila.com/insights/article/blackout-economics [Accessed on 13 May 2024].

Bollapragada, R., Mankude, A., and Udayabhanu, V. (2021). Forecasting the price of crude oil. *Decision*, 48(2):207–231.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). Training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pages 144–152.

Bouktif, S., Fiaz, A., Ouni, A., and Serhani, M. A. (2020). Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting. *Energies*, 13(2):391.

Bourguignon, M., Rodrigues, J., and Santos-Neto, M. (2019). Extended Poisson INAR(1) processes with equidispersion, underdispersion, and overdispersion. *Journal of Applied Statistics*, 46(1):101–118.

Bourguignon, M., Vasconcellos, K. L. P., Reisen, V. A., and Ispány, M. (2016). A Poisson INAR(1) process with a seasonal structure. *Journal of Statistical Computation and Simulation*, 86(2):373–387.

Box, G. E. P., Jenkins, G. M., and MacGregor, J. F. (1974). Some recent advances in forecasting and control. *Journal of the Royal Statistical Society*, 23(2):158–179.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509–1526.

Brownlee, J. (2018). *Deep learning for time series forecasting*. Machine Learning Mastery.

Brownlee, J. (2019). *Statistical methods for machine learning: Discover how to transform data into knowledge with Python*. Machine Learning Mastery.

Čeperić, E., Žiković, S., and Čeperić, V. (2017). Short-term forecasting of natural gas prices using machine learning and feature selection algorithms. *Energy*, 140:893–900.

CFI Team (2022). Heavy industry. `https://corporatefinanceinstitute.com/resources/economics/heavy-industry/` [Accessed on 28 May 2024].

Cheng, Y., Yi, J., Yang, X., Lai, K. K., and Seco, L. (2022). A CEEMD-ARIMA-SVM model with structural breaks to forecast the crude oil prices linked with extreme events. *Soft Computing*, 26(17):8537–8551.

Cheng, Y., Zhu, Q., Peng, Y., Huang, X. F., and He, L. Y. (2021). Multiple strategies for a novel hybrid forecasting algorithm of ozone based on data-driven models. *Journal of Cleaner Production*, 326:129451.

Chodakowska, E., Nazarko, J., and Nazarko, L. (2021). ARIMA models in electrical load forecasting and their robustness to noise. *Energies*, 14(23):7952.

Dai, S., Niu, D., and Li, Y. (2018). Daily peak load forecasting based on complete ensemble empirical mode decomposition with adaptive noise and support vector machine optimized by modified grey wolf optimization algorithm. *Energies*, 11(1):163.

David, F. N. and Johnson, N. L. (1948). The probability integral transformation when parameters are estimated from the sample. *Biometrika*, 35(1/2):182–190.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Dong, J., Dai, W., Tang, L., and Yu, L. (2019). Why do EMD-based methods improve prediction? a multiscale complexity perspective. *Journal of Forecasting*, 38(7):714–731.

Dong, X., Deng, S., and Wang, D. (2022). A short-term power load forecasting method based on k-means and SVM. *Journal of Ambient Intelligence and Humanized Computing*, 13(11):5253–5267.

Du, P., Wang, J., Yang, W., and Niu, T. (2018). Multi-step ahead forecasting in electrical power system using a hybrid forecasting system. *Renewable Energy*, 122:533–550.

Energy Information Administration (2016). Chapter 7: Industrial sector energy consumption. Report. `https://www.eia.gov/outlooks/ieo/pdf/industrial.pdf` [Accessed on 8 February 2024].

Energy Information Administration (2023). Energy use in industry. `https://www.eia.gov/energyexplained/use-of-energy/industry.php` [Accessed on 7 February 2024].

Energy Information Administration (2024). Heating oil explained: Use of heating oil. `https://www.eia.gov/energyexplained/heating-oil/use-of-heating-oil.php` [Accessed on 29 March 2024].

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 50(4):987–1007.

Engle, R. F., Hendry, D. F., and Trumble, D. (1985). Small-sample properties of ARCH estimators and tests. *Canadian Journal of Economics*, 18(1):66–93.

Fałdziński, M., Fiszeder, P., and Orzeszko, W. (2021). Forecasting volatility of energy commodities: Comparison of GARCH models with support vector regression. *Energies*, 14(1):6.

Farzanegan, M. R. and Markwardt, G. (2009). The effects of oil price shocks on the Iranian economy. *Energy Economics*, 31(1):134–151.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202.

Gandhi, H. K. (2022). Application of time-series method in company revenue. The 2022 IEEE 2nd Conference on Information Technology and Data Science.

Gandhi, H. K. (2024a). Applying hybrid forecasting model SARIMA-SVR for daily energy consumption data. In *2024 IEEE 3rd Conference on Information Technology and Data Science (CITDS)*, pages 1–6. Institute of Electrical and Electronics Engineers (IEEE).

Gandhi, H. K. (2024b). Mid-term forecasting of crude oil prices using hybrid CEEMDAN and CNN-LSTM deep learning model. *Polityka Energetyczna*, 27(4):19–38.

Gandhi, H. K. and Ispány, M. (2024). Multi-step natural gas price forecasting using ensemble empirical mode decomposition and long short-term memory hybrid model. *International Journal of Energy Economics and Policy*, 14(4):590–598.

Gandhi, H. K. and Ispány, M. (2025). Analyzing uncontrollable factors that cause defective products by Poisson and negative binomial INAR(1) for fitting model. *Proceedings on Engineering*, 7(1). [Accepted].

Ghalayini, L. (2017). Modeling and forecasting spot oil price. *Eurasian Business Review*, 7(3):355–373.

Ghosh, S. (2008). Univariate time-series forecasting of monthly peak demand of electricity in northern India. *International Journal of Indian Culture and Business Management*, 1(4):466–474.

Gray Solutions (2024). How three manufacturing giants are saving big on energy costs. `https://www.gray.com/insights/how-three-manufacturing-giants-are-saving-big-on-energy-costs/` [Accessed on 7 February 2024].

Haar, A. (1911). Zur theorie der orthogonalen funktionensysteme (on the theory of orthogonal function systems). *Mathematische Annalen*, 71(1):38–53.

Halper, E. (2022). A summer of blackouts? Wheezing power grid leaves states at risk. `https://www.washingtonpost.com/business/2022/06/02/blackout-states-summer-heat/` [Accessed on 2 January 2024].

Heatable (2018). What are the benefits of heating oil? `https://www.heatable.com/heating-oil-guide/heating-oil-101/benefits-of-oil` [Accessed on 21 February 2024].

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Hodrick, R. J. and Prescott, E. C. (1997). Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit, and Banking*, 29(1):1–16.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Yen, N., Tung, C. C., and Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A*, 454(1971):903–995.

Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 148(3):574–591.

International Energy Agency (2023). World energy balances highlights dataset. `https://www.iea.org/data-and-statistics/data-product/world-energy-balances-highlights` [Accessed on 8 February 2024].

Ismail, Z., Yahya, A., and Mahpol, K. A. (2009). Forecasting peak load electricity demand using statistics and rule based approach. *American Journal of Applied Sciences*, 6(8):1618–1625.

Ittianath, A. (2024). What is manufacturing industry - concept building. `https://www.vedantu.com/geography/what-are-manufacturing-industries` [Accessed on 28 May 2024].

Ivakhnenko, A. G. and Lapa, V. G. (1967). Cybernetics and forecasting techniques. *American Elsevier Publishing*, 168.

Jahanshahi, A., Jahanianfard, D., Mostafaie, A., and Kamali, M. (2019). An autoregressive integrated moving average (ARIMA) model for prediction of energy consumption by household sector in Euro area. *AIMS Energy*, 7(2):151–164.

Jaipuria, S. and Mahapatra, S. S. (2021). A hybrid forecasting technique to deal with heteroskedastic demand in a supply chain. *Operations and Supply Chain Management*, 14(2):123–132.

Jaiswal, R., Choudhary, K., and Kumar, R. R. (2022). STL-ELM: A decomposition-based hybrid model for price forecasting of agricultural commodities. *National Academy Science Letters*, 45(6):477–480.

Jarque, C. M. and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3):255–259.

Kilimci, Z. H., Akyuz, A. O., Uysal, M., Akyokus, S., Uysal, M. O., Atak, B. B., and Ekmis, M. A. (2019). An improved demand forecasting model

using deep learning approach and proposed decision integration strategy for supply chain. *Complexity*, 2019(1):9067367.

Komalawati, Asmarantaka, R. W., Nurmalina, R., and Hakim, D. B. (2019). Modeling price volatility and supply response of beef in Indonesia. *Tropical Animal Science Journal*, 42(2):159–166.

Krishna, P. N. and Singh, J. G. (2023). Electricity price forecasting using hybrid deep learned networks. *Journal of Forecasting*, 42(7):1750–1771.

Lazzeri, F. (2020). *Machine learning for time series forecasting with Python*. John Wiley & Sons.

Lewis, C. D. (1981). *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. Butterworth Scientific.

Li, Z. (2020). Engineering cost information management in big data era. In *Journal of Physics: Conference Series*, page 042072. IOP Publishing.

Liu, B. (2011). *Web data mining*, volume 1. Springer.

Liu, T., Luo, Z., Huang, J., and Yan, S. (2018). A comparative study of four kinds of adaptive decomposition algorithms and their applications. *Sensors*, 18(7):2120.

Liu, T., Song, S., and Duan, G. (2016). The application and management of big data in quality engineering. In *16th Asia Simulation Conference and SCS Autumn Simulation Multi-Conference*, volume 645, pages 624–631. Springer.

Lu, Q., Sun, S., Duan, H., and Wang, S. (2021). Analysis and forecasting of crude oil price based on the variable selection-LSTM integrated model. *Energy Informatics*, 4:47.

McKenzie, E. (1985). Some simple models for discrete variate time series. *JAWRA Journal of the American Water Resources Association*, 21(4):645–650.

Minsky, M. and Papert, S. (1969). Perceptron: An introduction to computational geometry. *The MIT Press, Cambridge, Expanded Edition*, 19(88):1–20.

Mirasgedis, S., Sarafidis, Y., Georgopoulou, E., Lalas, D. P., Moschovits, M., Karagiannis, F., and Papakonstantinou, D. (2006). Models for mid-term electricity demand forecasting incorporating weather influences. *Energy*, 31(2-3):208–227.

Montgomery, D. C. (2017). *Design and analysis of experiments*. John Wiley & Sons.

Montgomery, D. C., Jennings, C. L., and Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons.

Moran, P. A. and Whittle, P. (1951). Hypothesis testing in time series analysis. *Journal of the Royal Statistical Society*, 114(4).

Naikan, V. N. A., Majumdar, J., and Vijaykumar, B. K. (2012). Research trends in quality engineering and management. *International Journal of Performability Engineering*, 8(6):587–600.

Nguyen, H. T. and Nabney, I. T. (2010). Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models. *Energy*, 35(9):3674–3685.

Olakorede, N. M. and Olanrewaju, S. O. (2023). Integer-valued first order autoregressive (INAR(1)) model with negative binomial (NB) innovation for the forecasting of time series count data. *International Journal of Statistics and Probability*, 12(6).

Ondina, P. A. and Fuster, J. I. A. (2022). Rising energy prices and their impact on the manufacturing industry: Which sectors are being hit the hardest? `https://www.caixabankresearch.com/en/sector-analysis/industry` [Accessed on 12 February 2024].

Ortiz-García, E. G., Salcedo-Sanz, S., Pérez-Bellido, Á. M., Portilla-Figueras, J. A., and Prieto, L. (2010). Prediction of hourly O3 concentrations using support vector regression algorithms. *Atmospheric Environment*, 44(35):4481–4488.

Ouahilal, M., Mohajir, M. E., Chahhou, M., and Mohajir, B. E. E. (2017). A novel hybrid model based on Hodrick-Prescott filter and support vector regression algorithm for optimizing stock market price prediction. *Journal of Big Data*, 4(1):1–22.

Palkhe, S. V. (2020). Six sigma DMAIC methodology. *International Journal for Research in Applied Science and Engineering Technology*, 8(8):999–1002.

Parihar, M. and Bhaskar, M. K. (2018). Review of power system blackout. *International Journal of Research and Innovation in Applied Science*, 3(6):8–12.

Pavlyshenko, B. M. (2019). Machine-learning models for sales time series forecasting. *Data*, 4(1):1–15.

Piotrowski, P., Kopyt, M., Baczyński, D., Robak, S., and Gulczyński, T. (2021). Hybrid and ensemble methods of two days ahead forecasts of electric energy production in a small wind turbine. *Energies*, 14(5):1225.

Puig, P. and Valero, J. (2007). Characterization of count data distributions involving additivity and binomial subsampling. *Bernoulli*, 13(2):544–555.

Qin, Q., Huang, Z., Zhou, Z., Chen, Y., and Zhao, W. (2022). Hodrick-Prescott filter-based hybrid ARIMA-SLFNs model with residual decomposition scheme for carbon price forecasting. *Applied Soft Computing*, 119:108560.

Ramyar, S. and Kianfar, F. (2019). Forecasting crude oil prices: A comparison between artificial neural networks and vector autoregressive models. *Computational Economics*, 53(2):743–761.

Rennie, N., Cleophas, C., Sykulski, A. M., and Dost, F. (2021). Identifying and responding to outlier demand in revenue management. *European Journal of Operational Research*, 293(3):1015–1030.

Rochester, N., Holland, J. H., Haibt, L. H., and Duda, W. L. (1956). Tests on a cell assembly theory of the action of the brain, using a large digital computer. *IRE Transactions on Information Theory*, 2(3):80–93.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408.

Rosienkiewicz, M. (2021). Artificial intelligence-based hybrid forecasting models for manufacturing systems. *Eksploatacja i Niezawodność*, 23(2):263–277.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.

SAFIO Solutions (2020). Top benefits of forecasting. `https://www.safiosolutions.com/top-benefits-of-forecasting/` [Accessed on 19 March 2024].

Sehgal, N. and Pandey, K. K. (2015). Artificial intelligence methods for oil price forecasting: A review and evaluation. *Energy Systems*, 6(4):479–506.

Seo, M. and Kim, G. (2020). Hybrid forecasting models based on the neural networks for the volatility of bitcoin. *Applied Sciences*, 10(14):4768.

Shan, R., Dai, H., Zhao, J., and Liu, W. (2015). Forecasting study of Shanghai's and Shenzhen's stock markets using a hybrid forecast method. *Communications in Statistics: Simulation and Computation*, 44(4):1066–1077.

Shapiro, S. and Wilk, M. (1965). An analysis of variance test for normality. *Biometrika*, 52(3):591–611.

Sharma, H. K., Kumari, K., and Kar, S. (2021). Forecasting sugarcane yield of India based on rough set combination approach. *Decision Making: Applications in Management and Engineering*, 4(2):163–177.

Sharma, M., Sahni, S. P., and Sharma, S. (2019). Reduction of defects in the lapping process of the silicon wafer manufacturing: The six sigma application. *Engineering Management in Production and Services*, 11(2):87–105.

Shi, J., Guo, J., and Zheng, S. (2012). Evaluation of hybrid forecasting approaches for wind speed and power generation time series. *Renewable and Sustainable Energy Reviews*, 16(5):3471–3480.

Shumway, R. and Stoffer, D. (2019). *Time series: A data analysis approach using R*. CRC Press.

Siddiqui, R., Azmat, M., Ahmed, S., and Kummer, S. (2022). A hybrid demand forecasting model for greater forecasting accuracy: The case of the pharmaceutical industry. In *Supply Chain Forum*, pages 124–134. Taylor & Francis.

Sigauke, C. and Chikobvu, D. (2011). Prediction of daily peak electricity demand in South Africa using volatility forecasting models. *Energy Economics*, 33(5):882–888.

Singh, S. N. and Mohapatra, A. (2021). Data driven day-ahead electrical load forecasting through repeated wavelet transform assisted SVM model. *Applied Soft Computing*, 111:107730.

Soares, L. J. and Medeiros, M. C. (2008). Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data. *International Journal of Forecasting*, 24(4):630–644.

Song, J., He, J., Zhu, M., Tan, D., Zhang, Y., Ye, S., Shen, D., and Zou, P. (2014). Simulated annealing based hybrid forecast for improving daily municipal solid waste generation prediction. *Scientific World Journal*, 2014(1):834357.

Srinath, R. and Gayathri, R. (2022). Epilepsy disorder detection and diagnosis using empirical mode decomposition and deep learning architecture. *Concurrency and Computation: Practice and Experience*, 34(11):e6903.

Statista Search Department (2023). Natural gas consumption worldwide from 1998 to 2022 (in billion cubic meters) dataset. `https://www.statista.com/statistics/282717/global-natural-gas-consumption/` [Accessed on 30 October 2023].

Stein, Z. (2024). What is Heating Oil? `https://www.carboncollective.co/sustainable-investing/heating-oil` [Accessed on 8 April 2024].

Streimikiene, D., Rizwan Raheem, A. R., Vveinhardt, J., Ghauri, S. P., and Zahid, S. (2018). Forecasting tax revenues using time series techniques–a case of Pakistan. *Economic research-Ekonomska istraživanja*, 31(1):722–754.

Su, M., Zhang, Z., Zhu, Y., Zha, D., and Wen, W. (2019). Data driven natural gas spot price prediction models using machine learning methods. *Energies*, 12(9):1680.

Sudheer, C., Maheswaran, R., Panigrahi, B. K., and Mathur, S. (2014). A hybrid SVM-PSO model for forecasting monthly streamflow. *Neural Computing and Applications*, 24(6):1381–1389.

Sun, W., Wang, X., and Tan, B. (2022). Multi-step wind speed forecasting based on a hybrid decomposition technique and an improved back-propagation neural network. *Environmental Science and Pollution Research*, 29(33):49684–49699.

Svetunkov, I. and Boylan, J. E. (2020). State-space ARIMA for supply-chain forecasting. *International Journal of Production Research*, 58(3):818–827.

Tara Energy (2018). Oil as an energy source: Understanding how It works. `https://taraenergy.com/blog/oil-as-an-energy-source/` [Accessed on 12 February 2024].

Trading Economics (2023). Natural gas summary dataset. https://tradingeconomics.com/commodity/natural-gas/ [Accessed on 22 December 2023].

UniSource Energy Services (2020). 8 facts about natural gas. https://www.uesaz.com/news/8-facts-about-natural-gas/ [Accessed on 20 February 2024].

United Nations Environment Programme (2023). Is natural gas really the bridge fuel the world needs? https://www.unep.org/news-and-stories/story/natural-gas-really-bridge-fuel-world-needs [Accessed on 10 February 2024].

Valipour, M. (2015). Long-term runoff study using SARIMA and ARIMA models in the United States. *Meteorological Applications*, 22(3):592–598.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5):988–999.

Voronin, S. and Partanen, J. (2014). Forecasting electricity price and demand using a hybrid approach based on wavelet transform, ARIMA and neural networks. *International Journal of Energy Research*, 38(5):626–637.

Walker, G. T. (1931). On periodicity in series of related terms. *Proceedings of the Royal Society of London*, 131(818):518–532.

Wamwea, C., Mwelu, S., and Odin, M. (2023). Modelling COVID-19 cumulative number of cases in Kenya using a negative binomial INAR(1) model. *Open Journal of Modelling and Simulation*, 11(1):14–36.

Wang, A. (2018). Introduction to liquor of paper pulping. https://paperpulpingmachine.com/introduction-to-liquor-of-paper-pulping [Accessed on 8 February 2024].

Wang, F., Yu, Y., Zhang, Z., Li, J., Zhen, Z., and Li, K. (2018). Wavelet decomposition and convolutional LSTM networks based improved deep learning model for solar irradiance forecasting. *Applied Sciences*, 8(8):1286.

Wang, J., Hu, J., Ma, K., and Zhang, Y. (2015). A self-adaptive hybrid approach for wind speed forecasting. *Renewable Energy*, 78:374–385.

Wang, J., Lei, C., and Guo, M. (2020). Daily natural gas price forecasting by a weighted hybrid data-driven model. *Journal of Petroleum Science and Engineering*, 192:107240.

Wang, X. and Zhang, C. (2014). The impacts of global oil price shocks on China fundamental industries. *Energy Policy*, 68:394–402.

Weiß, C. H. (2018). *An introduction to discrete-valued time series*. Wiley.

Whittaker, E. T. (1922). On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41:63–75.

Wu, Z. and Huang, N. E. (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(01):1–41.

Yahoo Finance (2023). Heating oil dataset. `https://finance.yahoo.com/quote/HO%3DF/history` [Accessed on 22 December 2023].

Yang, J., Yamamoto, T., and Ando, R. (2021). The impact of mandating a driving lesson for elderly drivers in Japan using count data models: Case study of Toyota City. *Accident Analysis & Prevention*, 153:106015.

Yazdanfar, D. (2015). GARCH model and predictive performance of volatility forecasting: Evidence from oil market. *World Review of Entrepreneurship, Management and Sustainable Development*, 11(4):345–357.

Yeh, J. R., Shieh, J. S., and Huang, N. E. (2010). Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. *Advances in Adaptive Data Analysis*, 2(02):135–156.

Yu, L., Dai, W., Tang, L., and Wu, J. (2016a). A hybrid grid-GA-based LSSVR learning paradigm for crude oil price forecasting. *Neural Computing and Applications*, 27(8):2193–2215.

Yu, L., Liang, S., Chen, R., and Lai, K. K. (2022). Predicting monthly biofuel production using a hybrid ensemble forecasting methodology. *International Journal of Forecasting*, 38(1):3–20.

Yu, L., Zhao, Y., and Tang, L. (2017). Ensemble forecasting for complex time series using sparse representation and neural networks. *Journal of Forecasting*, 36(2):122–138.

Yu, Y. L., Li, W., Sheng, D. R., and Chen, J. H. (2016b). A hybrid short-term load forecasting method based on improved ensemble empirical mode decomposition and back propagation neural network. *Journal of Zhejiang University: Science A*, 17(2):101–114.

Yule, G. U. (1927). On a method of investigating periodicities in disturbed series. *Philosophical Transactions of the Royal Society of London*, 226:167–298.

Yun, P., Huang, X., Wu, Y., and Yang, X. (2023). Forecasting carbon dioxide emission price using a novel mode decomposition machine learning hybrid model of CEEMDAN-LSTM. *Energy Science and Engineering*, 11(1):79–96.

Zeng, L., Li, Z., Yang, J., and Xu, X. (2022). CEEMDAN-IPSO-LSTM: A novel model for short-term passenger flow prediction in urban rail transit systems. *International Journal of Environmental Research and Public Health*, 19(24):16433.

Zhang, K., Gençay, R., and Yazgan, M. E. (2017a). Application of wavelet decomposition in time-series forecasting. *Economics Letters*, 158:41–46.

Zhang, X., Wang, J., and Zhang, K. (2017b). Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by cuckoo search algorithm. *Electric Power Systems Research*, 146:270–285.

Zhang, Y. J. and Zhang, J. L. (2018). Volatility forecasting of crude oil market: A new hybrid method. *Journal of Forecasting*, 37(8):781–789.

Zheng, Y., Xu, Z., Liao, W., Lin, B., and Chen, J. (2023). Multi-step forecasting for household power consumption. *IEEJ Transactions on Electrical and Electronic Engineering*, 18(8):1255–1263.

Zhou, Q., Wang, C., and Zhang, G. (2019). Hybrid forecasting system based on an optimal model selection strategy for different wind speed forecasting problems. *Applied Energy*, 250:1559–1580.

Appendix A

Residual dependency check using LB test

L	SARIMA		SARIMA-ARCH		SARIMA-GARCH		SVM Only		SARIMA-SVR	
	Stat	p-value	Stat	p-value	Stat	p-value	Stat	p-value	Stat	p-value
1	6.3669	0.0116	5.0448	0.0247	4.8840	0.0271	0.4571	0.4990	0.5997	0.4387
2	11.0391	0.0040	9.0137	0.0110	8.8552	0.0119	0.8061	0.6683	3.3892	0.1837
3	12.3537	0.0063	9.9255	0.0192	10.0681	0.0180	2.6729	0.4448	3.4404	0.3286
4	12.3608	0.0149	9.9258	0.0417	10.1431	0.0381	2.8025	0.5914	5.6303	0.2285
5	12.3613	0.0302	9.9269	0.0773	10.3696	0.0654	3.2555	0.6607	6.8087	0.2353
6	12.7857	0.0466	10.4055	0.1086	10.5544	0.1032	3.3471	0.7642	10.4515	0.1550
7	13.5681	0.0594	11.3058	0.1258	11.1689	0.1314	6.0598	0.5328	10.4290	0.1511
8	13.7120	0.0896	11.4757	0.1762	11.2278	0.1891	7.9080	0.4425	10.8244	0.1627
9	13.7864	0.1301	11.5645	0.2390	11.4893	0.2437	8.7559	0.4601	9.8967	0.1938
10	14.9597	0.1335	13.0067	0.2233	13.5268	0.1957	11.2224	0.3405	14.9470	0.1340
11	15.1524	0.1756	13.2467	0.2775	13.9264	0.2371	14.4710	0.2080	14.9645	0.1841
12	15.4585	0.2173	13.6527	0.3234	14.4897	0.2705	15.2788	0.2265	17.6026	0.1283
13	15.4752	0.2786	13.6668	0.3977	14.5231	0.3381	15.6340	0.2695	17.6866	0.1698
14	16.3558	0.2921	14.7915	0.3926	15.8237	0.3243	17.3466	0.2382	18.7828	0.1734
15	16.7493	0.3341	15.1775	0.4387	16.0701	0.3774	17.3575	0.2979	19.4680	0.1933
16	16.9623	0.3880	15.3490	0.4992	16.2230	0.4375	18.0488	0.3211	19.5697	0.2402
17	18.1061	0.3822	16.4562	0.4918	17.28	0.4356	18.8957	0.3346	20.6534	0.2422
18	22.2167	0.2225	20.8740	0.2858	22.1732	0.2244	21.3245	0.2634	22.2113	0.2227
19	24.6197	0.1734	23.3534	0.2221	24.9837	0.1611	26.0151	0.1298	22.7366	0.2491
20	24.8387	0.2077	23.5005	0.2649	25.2409	0.1924	26.1253	0.1617	22.7435	0.3016

Appendix B

N-step energy consumption forecasting results

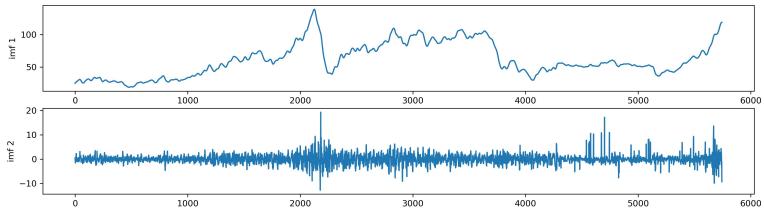
No	Horizon forecast	MSE	MAE	RMSE	MAPE	R2
1	SARIMA-SVR					
	1-step	0.0048	0.0436	0.0694	2.3214	0.7571
	2-step	0.0069	0.0629	0.0836	3.379	0.6598
	3-step	0.0057	0.0565	0.0756	2.9762	0.7222
	5-step	0.0066	0.0781	0.0814	3.2795	0.7271
	10-step	0.0081	0.0771	0.0895	3.5609	0.6839
	15-step	0.0075	0.0871	0.0866	3.4827	0.6925
2	SARIMA					
	1-step	0.0051	0.0545	0.0713	2.8546	0.7588
	2-step	0.0095	0.0734	0.0973	3.9624	0.5515
	3-step	0.0072	0.0666	0.0849	3.5003	0.6583
	5-step	0.0089	0.0857	0.0941	4.0078	0.5797
	10-step	0.0097	0.0819	0.0985	4.4224	0.5398
	15-step	0.0103	0.0957	0.1014	4.5103	0.3832
3	SARIMA-ARCH					
	1-step	0.0059	0.0611	0.0772	3.1911	0.7173
	2-step	0.0082	0.0719	0.0904	3.7362	0.5651
	3-step	0.0062	0.0618	0.0786	3.2237	0.677
	5-step	0.0064	0.0764	0.0799	3.0743	0.7339
	10-step	0.0063	0.0708	0.0791	3.2007	0.7028
	15-step	0.0073	0.082	0.0857	3.2796	0.7282

4	SARIMA-GARCH					
	1-step	0.0053	0.0542	0.0730	2.8613	0.7473
	2-step	0.0079	0.0683	0.0874	3.6359	0.6013
	3-step	0.0059	0.0587	0.0761	2.7237	0.7073
	5-step	0.006	0.0759	0.0749	2.9936	0.7402
	10-step	0.006	0.0683	0.0765	3.0837	0.7181
	15-step	0.0069	0.0764	0.0821	3.2062	0.7465
5	SVR					
	1-step	0.0082	0.0718	0.0906	3.8209	0.6109
	2-step	0.0233	0.1269	0.1529	6.8765	-0.1086
	3-step	0.0468	0.1728	0.2165	9.5067	-1.2234
	5-step	0.0395	0.1421	0.1989	8.0287	-0.8767
	10-step	0.043	0.1653	0.2074	9.1206	-1.0397
	15-step	0.0265	0.1352	0.1628	7.4761	-0.2575

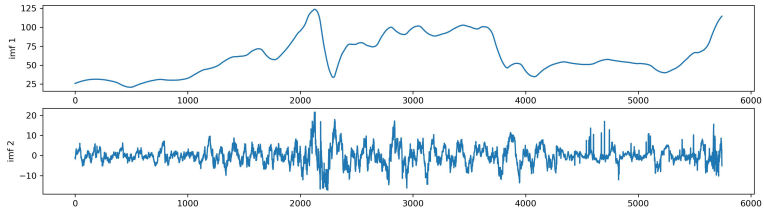
Appendix C

Decomposition results for three datasets

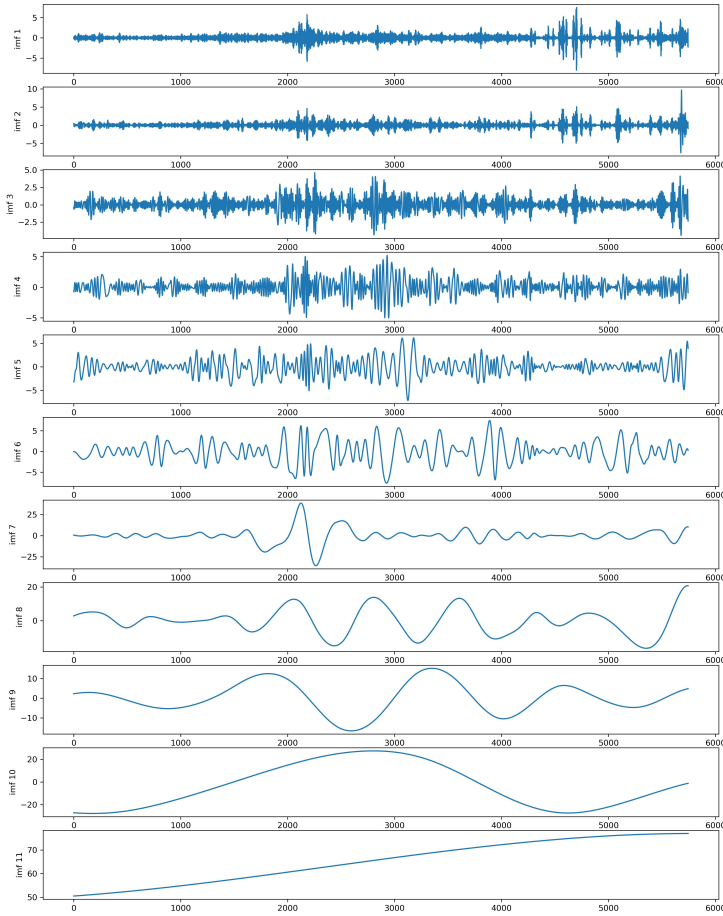
1.1 Hodrick-Prescott decomposition for WTI crude oil price



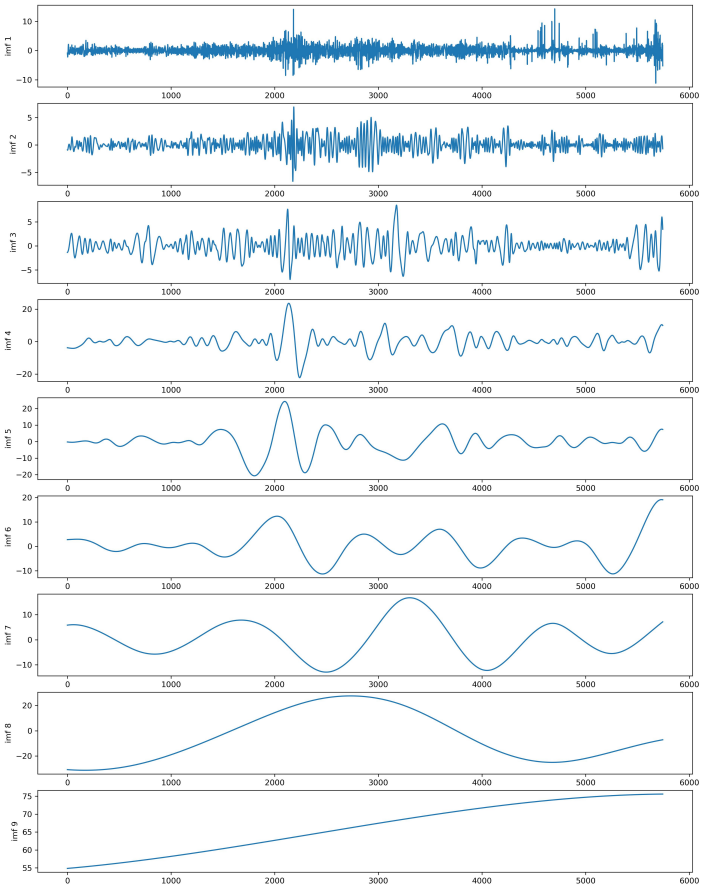
1.2 Wavelet decomposition for WTI crude oil price



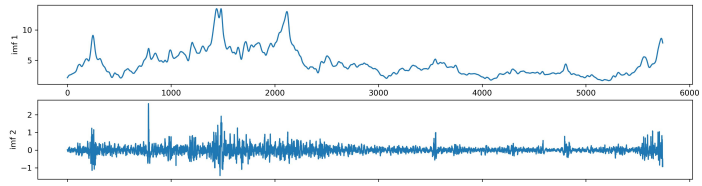
1.3 EMD for WTI crude oil price



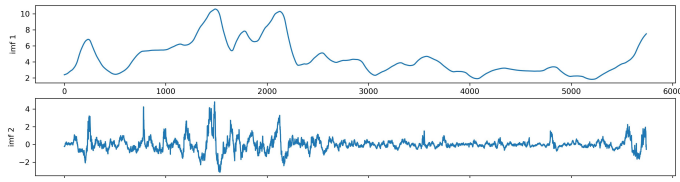
1.4 CEEMD for WTI crude oil price



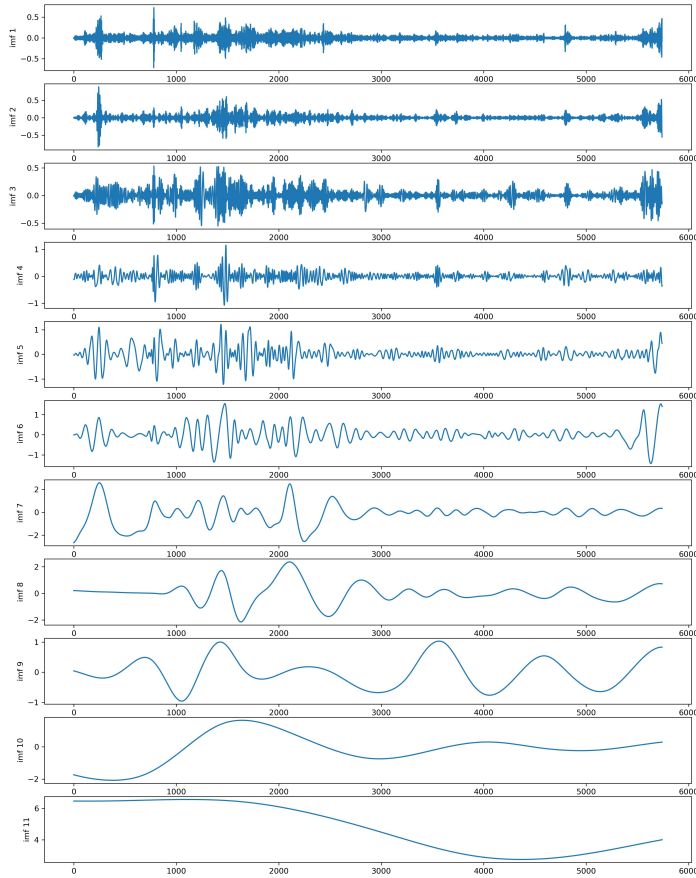
2.1 Hodrick-Prescott decomposition for natural gas price



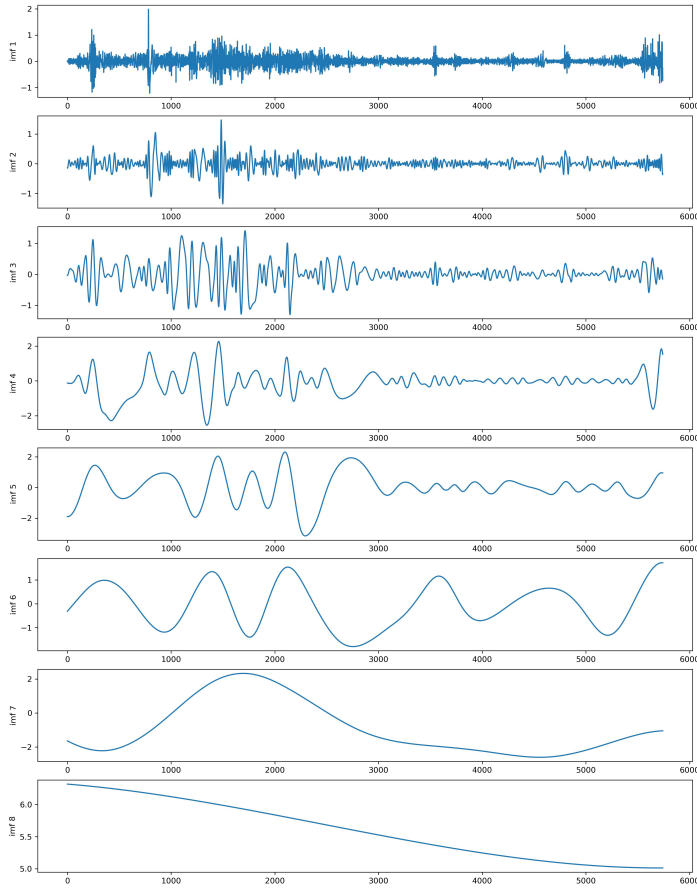
2.2 Wavelet decomposition for natural gas price



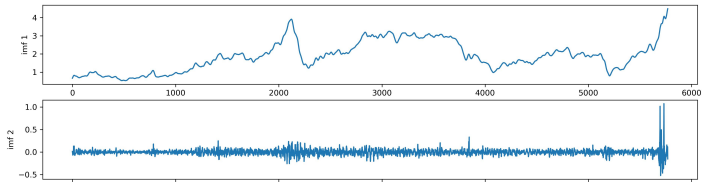
2.3 EMD for natural gas price



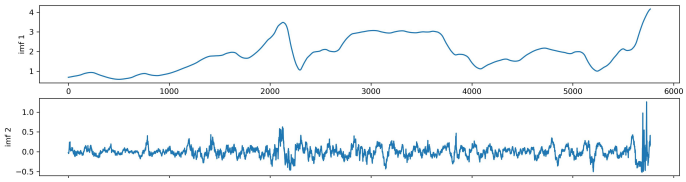
2.4 CEEMD for natural gas price



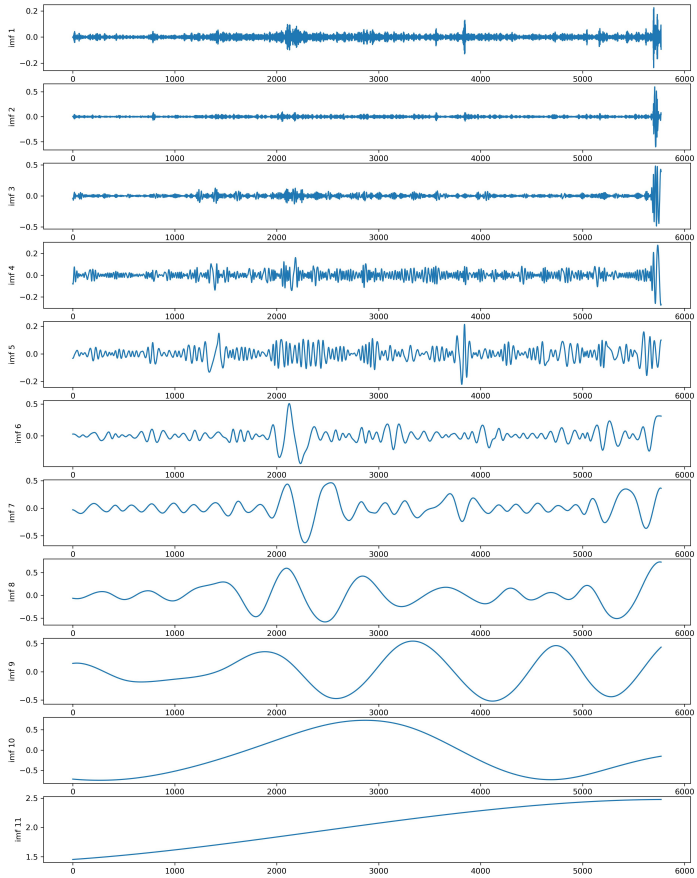
3.1 Hodrick-Prescott decomposition for heating oil price



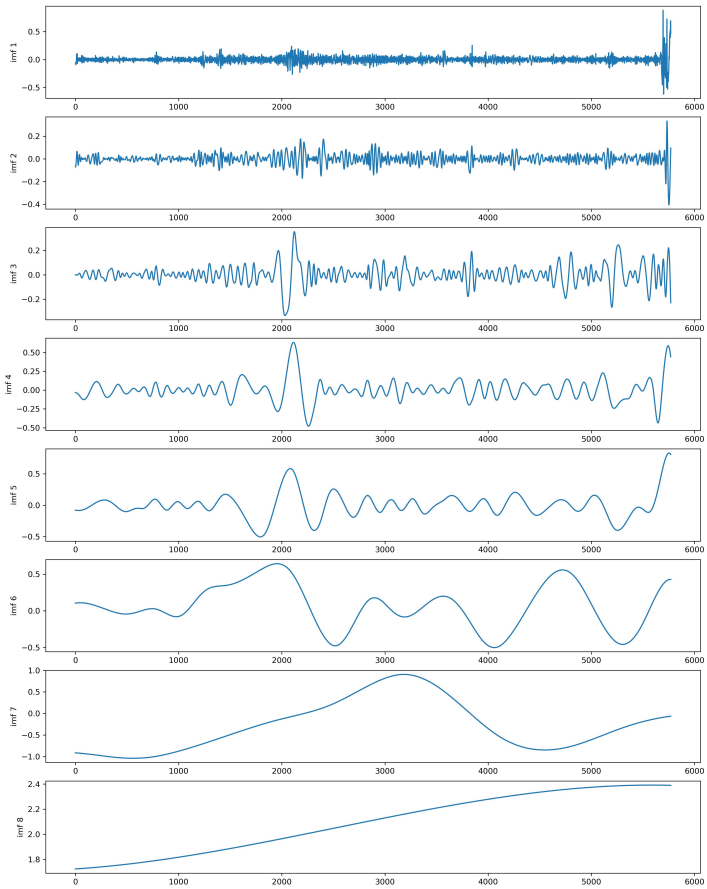
3.2 Wavelet decomposition for heating oil price



3.3 EMD for heating oil price



3.4 CEEMD for heating oil price



Appendix D

Percentage decreasing MAE between decompositions

1 Brent crude oil price

Step	% Difference MAE between CEEMD and			
	EMD	Wavelet	HP	No Decomp
30-step	7.93	16.19	55.35	56.47
35-step	22.24	7.43	51.05	52.23
40-step	7.65	0.74	41.61	47.03
45-step	10.59	1.85	45.72	49.41
50-step	20.97	12.56	52.93	54.54
55-step	2.25	1.50	45.32	48.39
60-step	9.54	0.48	33.33	45.47
65-step	22.63	10.45	48.22	50.94
70-step	20.89	12.06	41.05	46.88
75-step	6.04	13.90	45.69	47.73
80-step	14.99	25.35	45.53	46.20
85-step	16.56	26.25	48.62	50.73
90-step	12.87	14.45	37.32	44.31

2 WTI crude oil price

Step	% Difference MAE between CEEMD and			
	EMD	Wavelet	HP	No Decomp
30-step	6.96	18.06	47.45	52.84
35-step	12.25	14.59	45.09	50.7
40-step	18.97	18.27	54.00	58.18
45-step	13.34	19.18	48.96	58.11
50-step	24.09	24.92	54.22	60.17
55-step	16.99	21.31	54.68	56.71
60-step	22.84	28.42	51.98	56.58
65-step	27.20	39.24	59.63	61.98
70-step	37.82	29.04	56.77	61.91
75-step	15.86	31.07	61.33	53.85
80-step	32.95	33.23	60.65	56.87
85-step	17.09	24.38	56.13	59.02
90-step	3.50	11.90	50.15	56.88

3 Natural gas price

Step	% Difference MAE between CEEMD and			
	EMD	Wavelet	HP	No Decomp
30-step	-1.39	27.36	52.86	51.48
35-step	10.35	25.92	56.15	49.87
40-step	8.46	27.27	51.45	51.03
45-step	-0.95	14.67	51.76	49.98
50-step	4.70	17.80	46.86	51.90
55-step	9.13	29.18	53.95	54.04
60-step	9.08	27.80	54.80	54.54
65-step	6.32	27.40	55.22	58.19
70-step	3.71	24.23	54.36	58.73
75-step	3.23	25.39	56.52	60.95
80-step	1.55	28.51	53.17	58.06
85-step	6.37	23.29	50.78	57.95
90-step	-0.79	20.16	57.15	59.58

4 Heating oil price

Step	% Difference MAE between CEEMD and			
	EMD	Wavelet	HP	No Decomp
30-step	37.43	23.74	29.85	35.67
35-step	30.33	21.19	36.20	37.12
40-step	31.03	20.82	34.48	37.18
45-step	25.17	17.21	33.72	42.40
50-step	25.49	19.13	38.59	39.87
55-step	30.75	24.38	36.73	45.82
60-step	30.57	25.26	34.46	49.08
65-step	35.44	31.41	38.32	50.40
70-step	36.33	33.56	41.23	53.21
75-step	37.03	34.62	43.38	53.50
80-step	32.71	29.67	35.41	43.33
85-step	38.16	34.40	42.85	49.11
90-step	29.87	23.80	43.67	47.64

Appendix E

RMSE from 30 simulations of the best model in every step

1 Brent crude oil price

steps												
30	35	40	45	50	55	60	65	70	75	80	85	90
3.384	3.891	4.449	4.968	4.521	5.096	5.504	4.655	6.203	6.24	6.107	6.265	7.367
3.539	4.136	4.704	4.739	4.698	5.353	5.731	4.86	6.43	6.484	5.898	6.096	7.067
3.497	4.07	4.63	4.956	4.65	5.283	5.625	4.764	6.324	6.37	6.065	6.232	7.307
3.493	4.063	4.627	4.804	4.645	5.276	5.625	4.764	6.324	6.37	5.918	6.113	7.096
3.387	3.897	4.454	4.783	4.525	5.102	5.603	4.744	6.302	6.346	6.12	6.275	7.384
3.542	4.14	4.708	4.922	4.7	5.357	5.705	4.837	6.405	6.456	5.997	6.176	7.209
3.438	3.976	4.537	4.704	4.582	5.185	5.714	4.845	6.414	6.466	6.014	6.19	7.234
3.428	3.961	4.521	4.74	4.572	5.169	5.784	4.908	6.484	6.541	6.032	6.204	7.259
3.495	4.067	4.631	4.802	4.648	5.28	5.613	4.753	6.312	6.356	6.104	6.262	7.362
3.411	3.934	4.493	4.972	4.552	5.141	5.538	4.685	6.237	6.276	5.898	6.097	7.068
3.442	3.983	4.544	4.89	4.587	5.193	5.705	4.836	6.404	6.456	5.896	6.095	7.065
3.45	3.995	4.556	4.925	4.596	5.204	5.559	4.704	6.258	6.298	5.954	6.142	7.148
3.54	4.137	4.705	4.703	4.699	5.354	5.781	4.904	6.481	6.537	6.029	6.202	7.255
3.488	4.055	4.619	4.727	4.639	5.268	5.774	4.898	6.474	6.53	6.07	6.235	7.314
3.473	4.032	4.595	4.967	4.622	5.243	5.701	4.832	6.401	6.451	6.061	6.228	7.301

steps												
30	35	40	45	50	55	60	65	70	75	80	85	90
3.482	4.046	4.609	4.792	4.633	5.258	5.772	4.896	6.472	6.528	6.052	6.221	7.288
3.391	3.902	4.46	4.851	4.529	5.107	5.782	4.905	6.481	6.538	6.005	6.183	7.221
3.397	3.911	4.469	4.866	4.535	5.117	5.536	4.684	6.235	6.274	6.056	6.224	7.294
3.52	4.105	4.671	4.947	4.675	5.32	5.677	4.812	6.377	6.426	6.099	6.259	7.355
3.437	3.975	4.536	4.796	4.581	5.184	5.499	4.651	6.198	6.235	6.096	6.256	7.351
3.416	3.942	4.501	4.776	4.557	5.149	5.619	4.759	6.318	6.363	6.04	6.211	7.271
3.407	3.928	4.487	4.905	4.548	5.135	5.535	4.683	6.234	6.273	5.893	6.093	7.061
3.436	3.974	4.534	4.914	4.58	5.183	5.781	4.905	6.481	6.537	5.907	6.104	7.082
3.482	4.046	4.61	4.781	4.633	5.258	5.524	4.673	6.223	6.261	5.998	6.177	7.211
3.404	3.924	4.482	4.855	4.544	5.130	5.708	4.838	6.407	6.458	6.025	6.199	7.25
3.466	4.021	4.584	4.695	4.615	5.232	5.514	4.664	6.213	6.25	5.897	6.096	7.066
3.418	3.944	4.504	4.719	4.559	5.152	5.568	4.713	6.268	6.309	6.005	6.183	7.221
3.455	4.002	4.564	4.706	4.601	5.213	5.723	4.852	6.422	6.475	6.071	6.236	7.315
3.433	3.969	4.53	4.869	4.577	5.178	5.558	4.704	6.257	6.298	5.93	6.122	7.114
3.452	3.914	4.473	4.873	4.538	5.12	5.623	4.89	6.323	6.368	5.947	6.136	7.138

2 WTI crude oil price

steps												
30	35	40	45	50	55	60	65	70	75	80	85	90
2.474	2.560	2.679	2.621	2.620	2.803	2.990	3.020	3.126	3.056	3.117	3.735	4.100
2.445	2.533	2.637	2.580	2.594	2.772	2.946	2.996	3.094	3.012	3.090	3.697	4.066
2.404	2.496	2.580	2.523	2.559	2.728	2.885	2.964	3.050	2.952	3.052	3.645	4.019
2.459	2.546	2.658	2.601	2.607	2.787	2.968	3.008	3.110	3.034	3.104	3.716	4.083
2.515	2.597	2.736	2.678	2.655	2.847	3.050	3.052	3.171	3.117	3.155	3.788	4.147
2.407	2.498	2.584	2.527	2.562	2.731	2.890	2.966	3.053	2.956	3.055	3.649	4.022
2.408	2.499	2.586	2.529	2.562	2.732	2.891	2.967	3.054	2.957	3.056	3.650	4.023
2.434	2.523	2.622	2.564	2.585	2.760	2.929	2.987	3.082	2.996	3.079	3.683	4.053
2.465	2.552	2.666	2.609	2.612	2.794	2.976	3.013	3.117	3.043	3.109	3.724	4.089
2.479	2.564	2.686	2.628	2.624	2.808	2.997	3.024	3.132	3.064	3.122	3.742	4.105
2.459	2.546	2.658	2.600	2.607	2.787	2.967	3.008	3.110	3.034	3.103	3.716	4.083

steps												
30	35	40	45	50	55	60	65	70	75	80	85	90
2.451	2.539	2.646	2.589	2.600	2.778	2.955	3.001	3.101	3.022	3.096	3.705	4.073
2.432	2.521	2.619	2.562	2.583	2.758	2.927	2.986	3.080	2.993	3.078	3.681	4.051
2.487	2.571	2.696	2.639	2.631	2.817	3.008	3.030	3.140	3.075	3.129	3.751	4.114
2.410	2.501	2.589	2.532	2.565	2.735	2.895	2.969	3.057	2.961	3.058	3.653	4.026
2.482	2.567	2.690	2.633	2.627	2.812	3.002	3.026	3.136	3.069	3.125	3.746	4.109
2.401	2.493	2.576	2.519	2.557	2.725	2.881	2.962	3.047	2.947	3.049	3.641	4.015
2.476	2.562	2.681	2.624	2.621	2.805	2.992	3.021	3.128	3.059	3.119	3.738	4.102
2.419	2.509	2.601	2.544	2.572	2.744	2.907	2.976	3.066	2.974	3.066	3.664	4.036
2.499	2.583	2.713	2.656	2.641	2.830	3.027	3.039	3.154	3.093	3.141	3.767	4.128
2.456	2.543	2.653	2.595	2.604	2.783	2.962	3.005	3.106	3.028	3.100	3.711	4.078
2.485	2.570	2.694	2.637	2.629	2.815	3.006	3.028	3.138	3.073	3.128	3.750	4.112
2.434	2.523	2.623	2.565	2.585	2.760	2.930	2.988	3.083	2.997	3.080	3.684	4.054
2.476	2.562	2.682	2.625	2.622	2.806	2.993	3.022	3.129	3.060	3.120	3.738	4.102
2.510	2.593	2.729	2.672	2.651	2.842	3.043	3.048	3.166	3.110	3.151	3.782	4.141
2.440	2.529	2.631	2.574	2.590	2.767	2.939	2.993	3.089	3.005	3.086	3.691	4.060
2.500	2.583	2.714	2.657	2.642	2.830	3.028	3.040	3.154	3.094	3.141	3.768	4.129
2.516	2.598	2.737	2.680	2.656	2.848	3.052	3.053	3.172	3.119	3.156	3.789	4.148
2.430	2.520	2.617	2.560	2.582	2.756	2.924	2.985	3.078	2.991	3.076	3.679	4.049
2.448	2.536	2.642	2.585	2.597	2.776	2.951	2.999	3.098	3.018	3.093	3.702	4.070

3 Natural gas price

steps												
30	35	40	45	50	55	60	65	70	75	80	85	90
0.267	0.289	0.310	0.344	0.374	0.374	0.378	0.382	0.399	0.400	0.406	0.411	0.418
0.260	0.281	0.302	0.339	0.367	0.368	0.373	0.378	0.394	0.393	0.400	0.405	0.412
0.266	0.288	0.309	0.344	0.374	0.373	0.377	0.382	0.399	0.399	0.405	0.411	0.418
0.252	0.270	0.292	0.331	0.357	0.361	0.366	0.372	0.387	0.384	0.393	0.397	0.405
0.252	0.270	0.292	0.332	0.357	0.361	0.367	0.372	0.387	0.385	0.393	0.397	0.405
0.256	0.276	0.297	0.335	0.362	0.365	0.370	0.375	0.391	0.389	0.397	0.401	0.409
0.259	0.279	0.301	0.338	0.365	0.367	0.372	0.377	0.393	0.392	0.399	0.404	0.411
0.256	0.275	0.297	0.335	0.362	0.364	0.369	0.375	0.390	0.389	0.396	0.401	0.408

steps												
30	35	40	45	50	55	60	65	70	75	80	85	90
0.259	0.280	0.301	0.338	0.366	0.367	0.372	0.377	0.393	0.392	0.400	0.404	0.412
0.260	0.280	0.301	0.338	0.366	0.367	0.372	0.377	0.393	0.392	0.400	0.404	0.412
0.267	0.290	0.310	0.345	0.375	0.374	0.378	0.383	0.400	0.400	0.406	0.412	0.419
0.271	0.294	0.315	0.348	0.379	0.377	0.381	0.385	0.402	0.404	0.409	0.415	0.422
0.251	0.270	0.292	0.331	0.357	0.360	0.366	0.371	0.387	0.384	0.393	0.396	0.404
0.260	0.280	0.302	0.338	0.366	0.368	0.373	0.377	0.394	0.393	0.400	0.405	0.412
0.264	0.286	0.307	0.342	0.371	0.371	0.376	0.381	0.397	0.397	0.404	0.409	0.416
0.259	0.280	0.301	0.338	0.366	0.367	0.372	0.377	0.393	0.392	0.400	0.404	0.412
0.263	0.284	0.305	0.341	0.370	0.370	0.375	0.380	0.396	0.396	0.402	0.408	0.415
0.265	0.287	0.308	0.343	0.373	0.373	0.377	0.382	0.398	0.399	0.405	0.410	0.417
0.271	0.294	0.315	0.348	0.379	0.377	0.381	0.385	0.402	0.404	0.409	0.415	0.422
0.261	0.282	0.304	0.340	0.368	0.369	0.374	0.379	0.395	0.394	0.401	0.406	0.414
0.272	0.295	0.316	0.348	0.380	0.378	0.382	0.386	0.403	0.405	0.410	0.416	0.423
0.260	0.281	0.302	0.338	0.367	0.368	0.373	0.378	0.394	0.393	0.400	0.405	0.412
0.258	0.278	0.300	0.337	0.365	0.366	0.371	0.376	0.392	0.391	0.399	0.403	0.411
0.266	0.288	0.309	0.343	0.373	0.373	0.377	0.382	0.398	0.399	0.405	0.410	0.417
0.264	0.285	0.307	0.342	0.371	0.371	0.376	0.380	0.397	0.397	0.403	0.409	0.416
0.265	0.286	0.307	0.342	0.372	0.372	0.376	0.381	0.397	0.398	0.404	0.409	0.416
0.251	0.270	0.292	0.331	0.357	0.360	0.366	0.371	0.387	0.384	0.393	0.396	0.404
0.270	0.294	0.314	0.347	0.379	0.377	0.381	0.385	0.402	0.403	0.409	0.415	0.422
0.261	0.282	0.303	0.339	0.368	0.369	0.373	0.378	0.394	0.394	0.401	0.406	0.413
0.256	0.275	0.297	0.335	0.362	0.364	0.369	0.374	0.390	0.389	0.396	0.401	0.408

4 Heating oil price

steps												
30	35	40	45	50	55	60	65	70	75	80	85	90
0.159	0.160	0.171	0.185	0.186	0.188	0.189	0.194	0.196	0.200	0.216	0.225	0.230
0.161	0.161	0.172	0.186	0.187	0.188	0.190	0.195	0.197	0.201	0.217	0.227	0.231
0.159	0.160	0.170	0.184	0.186	0.187	0.189	0.193	0.196	0.199	0.215	0.225	0.229
0.156	0.156	0.167	0.181	0.183	0.185	0.187	0.190	0.192	0.196	0.212	0.221	0.226
0.156	0.155	0.167	0.180	0.183	0.184	0.187	0.189	0.192	0.196	0.211	0.221	0.226
0.156	0.156	0.167	0.181	0.183	0.185	0.187	0.190	0.192	0.196	0.212	0.222	0.226

steps												
30	35	40	45	50	55	60	65	70	75	80	85	90
0.163	0.165	0.175	0.189	0.190	0.191	0.192	0.198	0.200	0.204	0.220	0.230	0.234
0.158	0.159	0.170	0.184	0.185	0.187	0.189	0.193	0.195	0.199	0.214	0.224	0.229
0.160	0.161	0.171	0.185	0.187	0.188	0.190	0.194	0.197	0.200	0.216	0.226	0.230
0.155	0.154	0.166	0.179	0.182	0.184	0.186	0.188	0.191	0.195	0.210	0.220	0.225
0.162	0.163	0.173	0.187	0.188	0.189	0.191	0.196	0.199	0.202	0.218	0.228	0.232
0.165	0.167	0.177	0.191	0.192	0.192	0.193	0.200	0.203	0.206	0.222	0.232	0.236
0.160	0.161	0.171	0.185	0.187	0.188	0.190	0.194	0.197	0.200	0.216	0.226	0.230
0.162	0.164	0.174	0.188	0.189	0.190	0.191	0.197	0.199	0.203	0.219	0.229	0.233
0.155	0.154	0.166	0.179	0.182	0.184	0.186	0.188	0.191	0.195	0.211	0.220	0.225
0.157	0.157	0.168	0.182	0.184	0.186	0.188	0.191	0.194	0.197	0.213	0.223	0.227
0.162	0.163	0.174	0.188	0.189	0.190	0.191	0.197	0.199	0.203	0.218	0.228	0.233
0.162	0.163	0.174	0.188	0.189	0.190	0.191	0.197	0.199	0.203	0.218	0.228	0.232
0.159	0.159	0.170	0.184	0.185	0.187	0.189	0.193	0.195	0.199	0.215	0.224	0.229
0.165	0.167	0.177	0.191	0.191	0.192	0.193	0.200	0.202	0.206	0.221	0.231	0.235
0.162	0.163	0.173	0.187	0.188	0.189	0.191	0.196	0.199	0.202	0.218	0.228	0.232
0.162	0.163	0.174	0.187	0.188	0.190	0.191	0.196	0.199	0.202	0.218	0.228	0.232
0.165	0.167	0.177	0.191	0.191	0.192	0.193	0.200	0.202	0.206	0.221	0.231	0.235
0.161	0.162	0.172	0.186	0.187	0.189	0.190	0.195	0.198	0.201	0.217	0.227	0.231
0.160	0.160	0.171	0.185	0.186	0.188	0.190	0.194	0.196	0.200	0.216	0.225	0.230
0.160	0.160	0.171	0.185	0.186	0.188	0.190	0.194	0.196	0.200	0.216	0.226	0.230
0.163	0.164	0.175	0.188	0.189	0.190	0.192	0.197	0.200	0.203	0.219	0.229	0.233
0.160	0.161	0.172	0.186	0.187	0.188	0.190	0.195	0.197	0.201	0.216	0.226	0.230
0.157	0.156	0.168	0.181	0.183	0.185	0.187	0.190	0.193	0.197	0.212	0.222	0.227
0.159	0.159	0.170	0.184	0.185	0.187	0.189	0.193	0.195	0.199	0.215	0.224	0.229