THESIS

ESTIMATION OF COAL PRICES USING HYBRID ARIMA-GARCH MODELS

Submitted by:

LA ODE RIDWAN RINALDY TONDA
D111 19 1049

MINING ENGINEERING BACHELOR PROGRAM STUDY
FACULTY OF ENGINEERING
HASANUDDIN UNIVERSITY
GOWA
2024

LEGALIZATION

ESTIMATION OF COAL PRICE USING HYBRID ARIMA-GARCH MODELS

Compiled and Submitted by

La Ode Ridwan Rinaldy Tonda
D111191049

Has been defended in front of the Examination Committee established for the Completion of the Mining Engineering Undergraduate Study Program
Faculty of Engineering Hasanuddin University
On date September 17th, 2024
And declared eligible

Approved by,

Supervisor,	Co-Supervisor,
Dr. Eng. Rini Novrianti S. Tui., ST., MT., M.BA.	Rizki Amalia, S.T., M.T.
NIP. 198311142014042001	NIP. 199205042019016001

Head of Study Program,

Dr. Aryanti Virtanti Anas, S.T., M.T.
NIP. 197010052008012026

LEMBAR PENGESAHAN

ESTIMATION OF COAL PRICE USING HYBRID ARIMA-GARCH MODELS

Disusun dan diajukan oleh

La Ode Ridwan Rinaldy Tonda
D111191049

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Teknik Pertambangan Fakultas Teknik Universitas Hasanuddin Pada Tanggal 17 September 2024 dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,	Pembimbing Pendamping,
Dr. Eng. Rini Novrianti S. Tui., ST., MT., M.BA.	Rizki Amalia, S.T., M.T.
NIP. 198311142014042001	NIP. 199205042019016001

Ketua Program Studi,

Dr. Aryanti Virtanti Anas, S.T., M.T.
NIP. 197010052008012026

STATEMENT OF AUTHENTICITY

The Undersigned below:

Name	: La Ode Ridwan Rinaldy Tonda
Student ID	: D111191049
Study Program	: Mining Engineering
Degree	: Bachelor (S1)

Declare that my bachelor thesis titled,

Estimation of Coal Prices Volatility Using Hybrid ARIMA-GARCH Model

Is my own writing and not a claim of others. The Thesis which I wrote is truly my own work.

All Information written in the thesis are written by another writer have been credited, by citing source and publishing year. Therefore, all of the content of this thesis hhas fully become responsibility of the writer. If there shall be anybody who found similarity in title or results in this thesis, the writer is ready to be clarified and take responsibilities of every risks.

If in the future, it is proven or can be proven that a part or whole of this thesis is the work of others, I am willing to accept any penalty for such act.

Gowa, 17 October 2024

Declared by,

La Ode Ridwan Rinaldy Tonda

ABSTRACT

LA ODE RIDWAN RINALDY TONDA. *Estimation of Coal Prices Using Hybrid ARIMA-GARCH Models* (supervised by Rini Novrianti Sutardjo Tui dan Rizki Amalia).

The estimation of coal prices is an important aspect in Indonesia's energy industry, considering that significant price fluctuations can have an impact on business decisions and economic policies. This research aims to estimate the best hybrid ARIMA-GARCH model and predict coal prices using the hybrid ARIMA-GARCH model for September 2024 to January 2025. The ARIMA method is used to catch trend and seasonal patterns in coal price data, while the GARCH method is used to properly model price volatility that often appears in financial data. This method combination is expected to produce more accurate predictions compared to a single method. The Hybrid ARIMA-GARCH model is a combination model of the ARIMA and GARCH models, which can be used to overcome the residual problems of ARIMA models that indicate heteroscedasticity in residual variance (volatility). The steps in this analysis and discussion are descriptive statistics, stationarity testing, forming the best ARIMA model, forming the best GARCH model, combining hybrid ARIMA-GARCH models, determining the best hybrid ARIMA-GARCH model, measuring the accuracy of hybrid ARIMA-GARCH forecasting, and forecasting. This research was performed using Eviews in analyzing reference coal price data. The outcome of this research is that the best model for coal price is hybrid ARIMA(3,1,1)-GARCH(1,1) with MAPE value = 9.76%. According with the formula of the ARIMA(3,1,1)-GARCH(1,1) model is $\mathrm{D(HBA)} = 4.474\, log\, z_{t-1} + 4.973\, log\, z_{t-2} + 0.001 + 0.146\alpha_{t-1} + 0{,}0{,}596\alpha_{t-2} - 0.005 + 0{,}057\varepsilon_{t-1}^{2} + 0{,}221\sigma_{t-1}^{2}$. Based on the best model, the forecasting outcomes for September 2024 to January 2025 are \$115,7; \$120,83; \$118,99; \$129,2; \$117,31, respectively, which indicates that coal prices in June to October 2024 have decreased in price.

Key Words: Coal, Hybrid, ARIMA, GARCH, E-views.

ABSTRAK

LA ODE RIDWAN RINALDY TONDA. *Estimasi Harga Batubara Mengunakan Metode Hybrid ARIMA-GARCH* (dibimbing oleh Rini Novrianti Sutardjo Tui dan Rizki Amalia).

Estimasi harga batubara merupakan aspek penting dalam industri energi, mengingat fluktuasi harga yang signifikan dapat berdampak pada keputusan bisnis dan kebijakan ekonomi. Penelitian ini bertujuan mengestimasi model terbaik hybrid ARIMA-GARCH dan Memprediksi harga batubara menggunakan model hybrid ARIMA-GARCH untuk periode September tahun 2024 sampai Januari 2025. Metode ARIMA digunakan untuk menangkap pola tren dan musiman dalam data harga batubara, sementara metode GARCH digunakan untuk memodelkan volatilitas harga yang sering muncul dalam data keuangan. Kombinasi kedua metode ini diharapkan dapat menghasilkan prediksi yang lebih akurat dibandingkan dengan penggunaan metode tunggal. Model Hybrid ARIMA-GARCH merupakan model penggabungan dari model ARIMA dan GARCH, yang dapat digunakan untuk mengatasi masalah residual model ARIMA yang terindikasi adanya heteroskedastik dalam variansi residual (volatilitas). Tahapan dalam analisis dan pembahasan yaitu statistika deskriptif, pengujian stasioneritas, pembentukan model terbaik ARIMA, pembentukan model terbaik GARCH, penggabungan model hybrid ARIMA-GARCH, menentukan model terbaik hybrid ARIMA-GARCH, melakukan pengukuran akurasi peramalan hybrid ARIMA-GARCH, dan peramalan. Penelitian ini dilakukan dengan menggunakan software Eviews dalam menganalisis data harga batubara acuan. Hasil dari penelitian ini diperoleh model terbaik untuk harga batubara acuan adalah hybrid ARIMA(3,1,1)-GARCH(1,1) dengan nilai MAPE = 9,76%. Sehingga persamaan model ARIMA(3,1,1)-GARCH(1,1) adalah $D(HBA) = 4.474\, log\, z_{t-1} + 4.973\, log\, z_{t-2} + 0.001 + \ 0.146\alpha_{t-1} + \ 0{,}0{,}596\alpha_{t-2} - 0.005 + \ 0{,}057\varepsilon_{t-1}^2 + \ 0{,}221\sigma_{t-1}^2$. *Berdasarkan model terbaik tersebut diperoleh hasil peramalan untuk periode September 2024 sampai Januari 2025 berturut–turut adalah $115,7; $120,83; $118,99; $129,2; $117,31, yang menunjukkan bahwa harga batubara pada bulan September 2024 sampai dengan Januari 2025 mengalami fluktuasi harga.*

***Kata kunci**: Hybrid, ARIMA, GARCH, E-views*

LIST OF CONTENT

LIST OF FIGURE

LIST OF TABLE

TABLE OF ACRONYMS AND SYMBOL DESCRIPTION

Symbol/Acronym	Description
ARIMA	Autoregressive Integrated Moving Area
ARMA	Autoregressive Moving Average
ACF	Autocorrelation Function
AR	Autoregressive
EWMA	Exponential Weighted Moving Average
GARCH	Generalized Autoregressive Conditional Heteroscedastic
GDP	Gross Domestic Product
MA	Moving Average
MEMR	Ministry of Energy Mineral Resource
PACF	Partial Autocorrelation Fuction
PPP	Purchasing Power Parity
p-value	Probability Value
R^2	Determination Coeficient
μ	Constant
D	Differencing Parameters
n	Total Observation
ε	Random Variable with Zero Mean
S	Number of Periode in Season
ϕ	Autoregressive Parameters
θ	Moving Average Coeficient Parameters
σ	Variance

LIST OF APENDIX

ACKNOWLEDGEMENT

First of all I would like to express my gratitude to the presence of God, who has bestowed mercy and grace to me so that I can complete this thesis. This thesis is prepared as a mandatory for completing the bachelor program. Blessings and greetings delivered to Our Majesty Rasulullah Shalallahu Alaihi Wassalam, the leader and the revolutionary who has led mankind to an age of jihad in seeking knowledge and fighting for the truth.

The dynamic in the activities and preparation of this research is found no significant obstalces. Therefore, I would like to express respect and gratitude to Dr. Aryanti Virtanti Anas, S.T., M.T. as Head of the Department of Mining Engineering, Mrs. Dr. Eng. Rini Novrianti Sutardjo Tui, S.T., M.BA., M.T. as supervisor and Mrs. Ir. Rizki Amalia S.T., M.T as co-supervisor. who has taught, guided, and briefed so that the thesis composing took place smoothly.

A big gratitude to my dearest parents, Mr. H. Ria La Ode Misbahu Tonda, S.Pd.SD., M.Pd. and Mrs. Hj. Hamrina, S.Pd., M.Pd. as well as the entire extended family who have provided motivational support and material support to me so that the writing of this thesis can be completed.

I would also express my gratitude for IGNEOUZ 2019 friends who always accompany and support in since the very beginning of my study program, thanks to SINEMA members, and all of PVT members for providing countless support, motivation, and advices.

I apologize if there were any miswords in the preparation of this thesis, because I relized perfection belongs only to the creator. Hopefully in the future this thesis can be useful for overyone.

CHAPTER I
INTRODUCTION

1.1 Research Background

Indonesia's economy has shown significant growth during recent decades. At the moment, Indonesia is the largest economy in Southeast Asia and the 7th largest economy worldwide in terms of Gross Domestic Product (GDP) on purchasing power parity (PPP) basis. The limited resources of petroleum energy require every country especially Indonesia, to save petroleum energy usage. Coal as one of the alternative energy sources that began to be viewed by the industrial world, which made this energy quite strategic. The usage of coal in fulfillment of national energy demands, including for steam power plants (PLTU) and as an alternative energy source which can be used as a substitute for fuel oil (BBM). In 2016, the total domestic coal consumption was 90.78 million tons, 69 million tons or 76% of which were used by PLTU (Haryadi and Suciyanti, 2018).

Indonesia's large coal reserves paired with high coal prices in international markets have forced coal companies to increase their production. As seen from the increase in exports by 31.3 million tons from 272.7 million tons to 304 million tons in 2018. Meanwhile, domestic coal demand also increased by 4.4 million tons from 57.2 million tons to 61.6 million tons in 2018 (BPS, 2019). The amount of exports and domestic demand creates problems such as the unbalanced distribution of coal between the fulfillment of national energy needs and global demand. In 2021, the Ministry of Energy and Mineral Resources implemented a coal export ban policy due to the gap between exports and fulfillment of DMO (Domestic Market Obligation) for PLTU needs (Haryanto, 2022). Domestic Market Obligation (DMO) is a policy made by the government that aims to ensure adequate coal supply for domestic needs. Regarding the DMO, each company is required to sell at least 25% of the planned production (KEPMEN ESDM, 2019).

At the moment, global coal usage is getting out of control, even major countries such as China, India, United States, and Australia, which are the largest coal users, have not signed an agreement at the UN COP26 Climate Summit in Glasgow to discontinue coal usage because the amount of coal used is considered to have exceeded the normal limit of world coal use. This has resulted in the world experiencing a global crisis that affects many economic sectors. Indonesia was also affected, especially in the mining sector. This certainly caused a decrease in coal demand which resulted in an oversupply of coal itself in the market. Such oversupply has an impact on unstable coal prices (Haryadi and Suciyanti, 2018).

The projected coal price as the basis for decisions when exporting coal is based on historical coal price data. In case this prediction changes drastically to a higher one, the price of coal will increase quickly. Therefore, proper price planning or forecasting will affect the long-term success in exporting coal. The time series analysis is a quantitative forecasting analysis method that considers time, where data is collected periodically based on time sequence to determine patterns from past data collected regularly. Time series forecasting techniques are divided into two parts. First, forecasting models based on statistical mathematical models such as moving average, exponential smoothing, regression, and ARIMA (Box Jenkins). Second, artificial intelligence-based forecasting models such as artificial neural networks, genetic algorithms, classification, and hybrids (Wijayanti, 2012).

Autoregressive Integrated Moving Average (ARIMA) method is the method most often used as a financial data forecasting method, because the estimation results of this method are the best model for some cases. The usage of the ARIMA method has great short-term forecasting accuracy, as it has a high level of accuracy. However, the ARIMA method itself has shortcomings in terms of identifying data that has heteroscedastic properties. As said by Hyndman (2021) that ARIMA takes into account past values (autoregressive) as well as past errors (moving average), so that it can capture patterns in the data dynamically. ARIMA may not be able to capture complex or non-linear patterns in the data, making it less effective if the data pattern is too complex or has a non-linear relationship and also ARIMA does not handle volatility or variance changes in the data well, so it is not suitable for data with high volatility. GARCH (Generalised Autoregressive Conditional Heteroskedasticity) models are usually better suited under these conditions. Therefore, to estimate heteroscedastic data, the GARCH (Generalized Autoregressive Conditional Heteroskedasticity) method is used. In this method, there is also another method that can be used to estimate data that has heteroscedasticity, called EWMA (Exponential Weighted Moving Average). Sukma (2012) shows that the GARCH method is able to produce a smaller error value than the EWMA (Exponential Weighted Moving Average) method. Both methods can be a solution for forecasting coal prices and can also see price anomalies, in order for companies to take the right steps in determining coal prices. Therefore, it is necessary to conduct research on combining ARIMA and GARCH, which is then called hybrid ARIMA-GARCH to forecast coal prices.

1.2 Research Problem

One of the main challenges in estimating coal prices is the price fluctuations that often cannot be predicted easily, considering many factors that influence, either of demand and supply side. To address this issue, this research focuses on the application of hybrid ARIMA-GARCH model in coal price estimation. The ARIMA model is used to catch linear patterns in coal price data, while the GARCH model is applied to handle volatility that often appears unexpectedly. By combining both models, it is hoped that the estimation results can be more accurate to the market changes. This research will see how much coal prices change during September 2024 until January 2025 by using hybrid ARIMA-GARCH.

1.3 Research Objective

The objectives of this research are including:

1. Estimating the best model hybrid ARIMA-GARCH.
2. Forecasting coal prices using hybrid ARIMA-GARCH models for September 2024 until January 2025.

1.4 Research Benefit

The benefits of this research are including:

1. The outcomes of this research are useful for the government in decision-making as well as regulators to quantify an cost-effective strategies for energy and coal sector management.

2. The outcomes of this research contribute to improving insight about coal price dynamics and increasing the available instrument for the management of coal commodities market risks.

1.5 Reseach Scope

This research will focus on coal prices which have fluctuated in recent years. This research will analyze historical data on Indonesian coal prices start from 2009 January till 2024 August. Quantitative methods, such as Autoregressive Intergrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) econometric modeling, are combined to resolve the ARIMA model residual problem which is indicated by heteroscedasticity in the residual variance.

CHAPTER II
LITERATURE STUDY

2.1 Overview of Coal Markets

Economic recovery and growth were gaining traction throughout 2022, supported by strong export, investment and household spending growth. However, downside risks, such as weak global demand, capital outflow, currency pressures and tight global financial conditions could potentially hinder growth momentum over the next four years. However, there has been a slowdown in the country's economic growth in 2023. As of now, the country's GDP growth is projected to be stable at an average of 5% (IMF,2023). Nevertheless, the figure is encouraging considering that the world's economies are expected to experience a major slowbalisation in 2023 as the battle against inflation continues. This robust and steady economic growth may drive Indonesia to be the fifth-largest economy in the world by 2030 and the fourth-largest one by 2050 on purchasing power parity basis (PwC, 2023).

Indonesia's economic growth can also be measured through GNI per capita. It is usually related to inflation, productivity, infrastructure growth, as well as social factors: such as the country's population health, education, and skill. World Bank categorizes Indonesia as a lower-middle income country (based on GNI per capita using the Atlas method (current USD). Despite the stable economic growth, the growth of GNI per capita in Indonesia has stagnated around $3,500 over the last six years. It experienced a short downtrend after 2013 which coincided with the end of the commodity boom period. However, a closer look at the GNI per capita with constant price of 2011 shows a steady increase instead. This showcases that the average income of the residents in the country still experienced improvement along with economic growth. The stagnated GNI per capita could then be attributed to devaluation of IDR toward USD since the commodity boom period ended.

Historically, Indonesia's primary energy mix shows a different story. Coal is on the rise in the primary energy mix in the last decade due to the acceleration of the power plant development program. Renewable mix is also increasing although at a much slower pace. The renewables increase is mainly contributed by the biofuel usage to replace fuel diesel in industry and transportation sectors and by geothermal for electric power.

In addition to its role in energy sector, coal also contributes to national development as a revenue stream for the State Budget. According to government regulation no. 9/2012, there are three ways on how coal sector can contribute to state revenue: land rent, royalty/tax, and sales of mining product. For the last four years, coal revenue collected is averaging around IDR 31 trillion (2.17 billion USD) or averaging close to 80% of total non-oil & gas revenue. However, coal revenue contribution to the state budget is relatively low, around 1.5 to 2 % of total revenue which can be seen in figure 1 (Mariatul Aini, 2018).

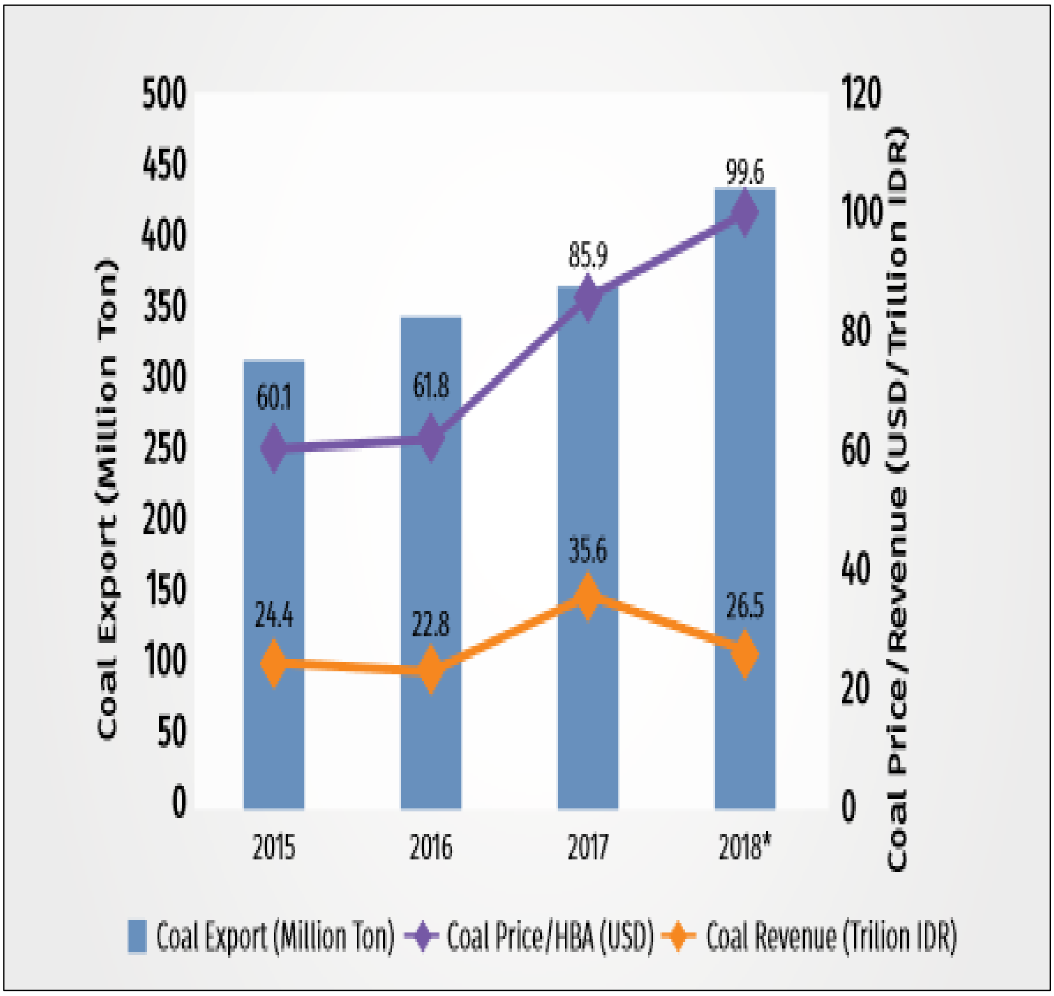

Figure 1 Revenue from the coal sector (Mariatul Aini, 2018; Ministry of Finance, 2019).

The government's reasoning over the exploitation of coal is to increase trade revenue and help in counterbalancing deficit coming from oil and gas trade.

Indonesia's import has risen by 22.2 % from 2017's figure, mainly dominated by the increasing of raw material import for industry and fuel (Ministry of Trade, 2019). Nevertheless, Indonesia experienced the worst net trade record in 2018, reaching minus 8.57 billion USD. The record is worse compared to the 2013 and 2014 trade deficit value of 4.08 and 1.89 billion USD, respectively (Fajriah, 2019). It is more than likely that the trend will continue in 2019 and thus the government will still look to coal export as one of the options for trade deficit balancing (given that the international price of coal stays high at >90 USD/ton) while building a strategy on reducing imports of consumer goods.

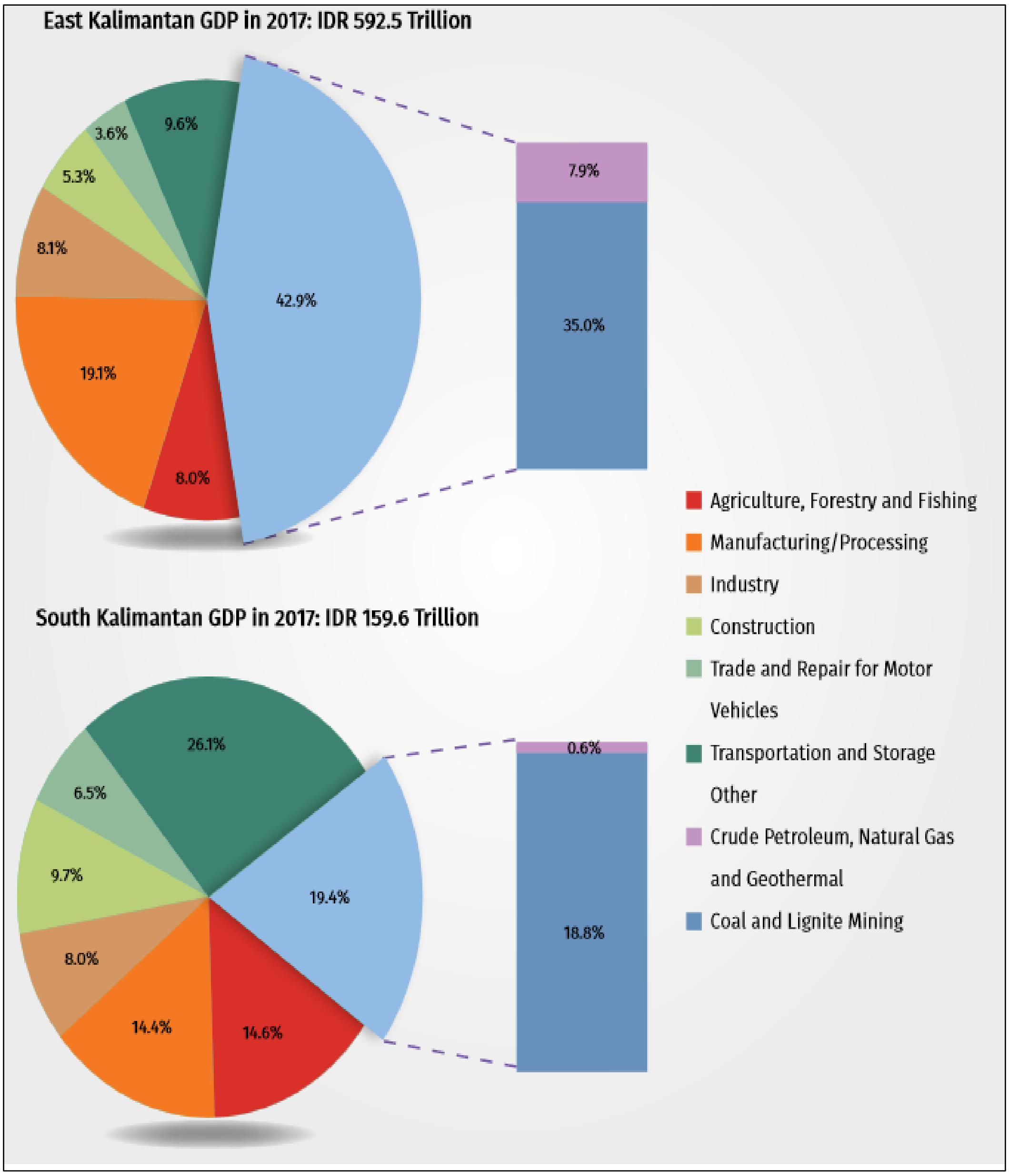

Figure 2 Comparison of GDP of South Kalimantan and East Kalimantan (BPS, 2018).

Indonesia’s coal resources and production are mainly distributed over only four provinces out of 34: East Kalimantan, South Sumatera, South Kalimantan, and Central Kalimantan. Kutai, Tarakan, and Barito coal basins located in East Kalimantan have medium-quality coal (calorific value between 5100-6100 kcal/kg) while the Central and South Sumatera Basins have low-quality coal reserves (calorific value <5100 kcal/kg). Coal has a substantial contribution to the local economy of the four provinces. In East Kalimantan, coal sector contributed up to 35% of the provincial GDP in 2017. By adding oil and gas to the figure, the number almost reach half of the provincial GDP. This indicates that East Kalimantan economy relies heavily on fossil fuel. A similar condition can be found in South Kalimantan province. Although South Kalimantan has lower GDP value compared to East Kalimantan, South Kalimantan’s coal sector contribution is rather high, ranging between 19-26% of the provincial GDP in the last five years. Considering the high share of GDP from coal sector and also the discrepancy between coal and other sectors’ development in both provinces, coal transition may have more impacts on their economics, social, and political environment (Adiatma et al., 2018).

East Kalimantan’s economy is four times larger than South Kalimantan’s. The sources of South Kalimantan’s economy are diverse (e.g., coal mining, industry, trade, and transportation) and are comparable in size. On the other hand, the East Kalimantan’s economy depends mostly on coal sector with more than a third of its GDP contribution coming from coal. The next largest GDP contributor in this province is the manufacturing/processing industry, agriculture, and construction with a considerable difference of value compared to coal sector’s contribution. Overcoming this gap would be more challenging and would be a crucial strategy for East Kalimantan to shift away from its coal-dominated industry.

2.2 Factors Influencing Coal Prices

In the third quarter of 2021, due to tight supply and unabated demand, the price of thermal coal rose sharply. After the National Development and Reform Commission implemented intervention measures in late October, the price of thermal coal declined and operated at a low level. Coal accounts for the highest proportion of primary energy use in China. The ratio of coal consumption to total energy consumption in 2021 was 56%. From 1994 to 2021, coal consumption accounted for 75% at the highest. Coal can be divided into two categories according to its use: thermal coal and coke. The iron and steel industries and the non-ferrous metal-smelting industry mainly use coke, since coke is effective and efficient in metal smelting. The main demand side of thermal coal is power plants. The fluctuation of the thermal coal price will affect the economic benefits of downstream industries such as power plants and steel, etc. Therefore, the stability of thermal coal prices is crucial to the stability of the economic situation. The national planning and intervention measures for the energy market will have an impact on the price of thermal coal.

The value of coal is subject to many constraints, including coal mine costs, international coal transportation prices, macroeconomic policies of countries around the world, global coal prices, coal stocks, the rise in the price of alternative resources, etc., because there are commodities have prices, and the most fundamental reason affecting the world commodity prices is the issue of supply and demand, with the introduction of various environmental protection measures in recent years, the price of coal as a non-clean resource will also be subject to (Xingchi et al, 2023).

2.2.1 Supply and Demand

Coal mine production capacity is the most important reason for coal supply and demand, it is the core of supply side to determine coal supply and demand. At present, due to the excessive production of coal in China, it is easy to cause overcapacity, and also easy to lead to the downward movement of coal prices.

The net import of coal is another element to reflect the supply level of coal, which is by directly reflecting the supply of coal, thus affecting the price of coal products.

Coal demand is a fundamental factor on the supply and demand side that governs the direction of coal prices, and determines coal price trends on a larger scale. High coal consumption will cause a tight market supply, which will lead to higher coal prices; conversely, the market is depressed, and coal prices will fall.

2.2.2 Cost Factors

Coal production costs directly affect the supply side of coal. Coal production costs are mainly divided into mining costs and transportation costs, production costs rise, coal companies can hardly afford, will inevitably share the risk to consumers by increasing the price of coal, and this leads to an increase in price.

2.2.3 Industry Concentration

Industrial concentration refers to the degree of dominance of a few enterprises in a certain industry in the market in terms of production volume, sales volume, total assets, etc. Coal industry concentration, is another important factor affecting the price of coal, and play an important role in the role of coal supply and demand factors. With too many coal producers, in a highly fragmented market structure, any coal company is a passive recipient of market prices, and the high or low price of coal is determined entirely by supply and demand. On the contrary, under a completely monopolistic market structure, the only coal enterprise will be the dominant price taker, while under a fully competitive market structure, coal prices face a certain degree of uncertainty (Wang, 2021).

2.2.4 Other Energy Prices Influence

Coal, oil and natural gas have become the main components of China's economy and the world's energy structure, and to some extent, the three may be substitutes for each other. Especially with the increasing global climate change issues, optimize the energy consumption structure, reduce human dependence on coal resources, has become an important initiative in energy conservation. Therefore, as the traditional coal alternative products such as oil, natural gas and other new energy products price changes, will also have an important impact on global coal prices.

2.2.5 The Impact of Government Policies

In order to promote economic development and energy conservation and emission reduction, improve the urban atmosphere, China has introduced various aspects of environmental protection preferential policies, but with the expansion of China's regulatory scope, coal prices have been greatly affected, especially at present, China's environmental protection capacity has reached or even close to the upper limit, so the price of coal by China's energy conservation and emission reduction preferential policies and carbon emission reduction measures are also more influential.

Supply and demand is the main factor affecting coal prices, coal prices are below cost period, coal mines into losses, naturally will reduce investment, supply reduced to a certain extent, there will be an oversupply, prices rise; coal prices are above cost period, investors profitable, naturally increase investment, supply increased to a certain extent, there will be more than demand, prices fall (Zhu, 2017).

2.3 Time Series Data

Time series data is a series of observations ordered by time with the same distance. This type of data is often encountered in everyday life because the data is collected at intervals, namely daily, weekly or monthly. Based on the collected data, it can be seen that there is a pattern in it. The time series data pattern is divided into three, namely trend, cyclical and seasonal patterns. Seasonal patterns are patterns that experience the same repetition many times at certain intervals. Based on domain division, time series data is divided into two domains, namely time domain and frequency domain. The time area examines the significance of autocorrelation, data stationarity, parameter estimation of time series regression models and forecasting. Meanwhile, the frequency area examines hidden frequencies in difficult seasonal data obtained in the time area. The aim is to find out special things or certain conditions in the data (Al'afi et al., 2020).

Time series data is data that consists of an object but covers several time periods, for example daily, monthly, weekly and yearly. It can be seen from examples of time series data on reference prices for several commodities, stock prices, production data, and so on. If you observe that each data is related to time and occurs sequentially, it will be very helpful so that it is easy to recognize the type of data. Time series data is also very useful for decision making to predict future events. While its because the pattern of changes in time series data from previous periods will repeat itself in the present. Time series data also usually depends on lag or difference. For example, in several cases, such as the supply and demand for coal commodities in previous years, this will influence coal prices in the following year, thus lag data on coal reference prices will be needed (Kusuma and Solihin, 2020).

Each value of an observation can always be related to the time of observation. Only at the time of the analysis, the connection variable time with observations is not a problem, because time series data is a collection of data based on time and one aspect of time series data is the involvement of a quantity called autocorrelation, which has the same concept as correlation for bivariate data. Another aspect of time series data is the stationarity of the data which is classified into strong stationary (first order stationary) and weak stationary (second order stationary), and this stationarity is a necessary condition in the analysis of time series data, because it will reduce standard errors (Kusumah and Solihin , 2020).

Time series data in the financial sector, especially return data, has a tendency to have a certain character, where this term is known as stylized fact. Sewell (2011) believes that stylized fact is a term commonly used in economics which refers to empirical evidence that there is the same consistency in a particular field so that it is accepted as truth. Some stylized facts found in time series data include unit roots, heteroscedasticity, volatility clustering and probability distributions that are fat tails relative to the normal distribution. In time series there are four types of data patterns, namely (Kusumah and Solihin, 2020):

1. Horizontal, When the data in an observation varies around a level or with a constant average. For example, monthly sales of a product do not increase or decrease consistently at any given time.
2. Seasonality is a data pattern when observations are influenced by seasonality which is characterized by a pattern of change that repeats itself automatically from 10 years to the next. An example is the data pattern of purchasing new books in the new school year.

3. By date Cyclical, is a form of data pattern characterized by wavy fluctuations in data that occur around the trend line. An example is data on economic and business activities.
4. Trend is a form of data pattern when observations increase or decrease over an extended period of time. An example is an example of population data.

Time series data is a more efficient and cheaper option for producing accurate forecasts. There are several things that need to be considered when processing time series data, namely (Wardhono et al, 2019):

1. Based on the assumption of stationarity.
2. If the stationarity assumption is not met, autocorrelation will arise.
3. Regression with R value^2 higher than 0.9 indicates an insignificant relationship.
4. There is a random walk phenomenon. For example, tomorrow's stock price is the same as today's stock price plus random errors.
5. Regression with time series data often used for forecasting.
6. Testing for stationarity is carried out before the causality test.

Data stationarity is the main issue related to spurious regression problems in time series analysis. The unit roots test is one tool for testing data stationarity. If any variable consists of a unit root, is non-stationary and combines with variable If anything else is not stationary, the two series will form stationarity in the cointegration relationship. Testing stationarity on a variable aims to see if there is a linear combination of terintegrasi forming stationarity or balance relationships (Wardhono et al., 2019).

Stationarity is related to the consistency of time series data movements. If stationary data has a constant mean and variance throughout time followed by the value of the variance between two periods only depends on the distance. Stationary data will move stably and converge around the average value with small deviations without positive or negative trend movements. The cointegration test is a long-term relationship between a variable which is not stationary and produces a linear combination so as to create a stationary condition or in the long term reach an equilibrium condition. Error Correction Model (ECM) dynamic model for correct regression equation variable which is not stationary so that it returns to a condition of equilibrium in the long term provided that there is a cointegration relationship between variables. As explained, the stationarity test is used to test stationarity, which can be seen as follows (Wardhono et al., 2019):

1. Graphical analysis
    a. Data movement deviation
    b. Trend truth

       If the data is not stationary, it indicates that the deviation tends to be further away from the average and has a certain trend.
2. Autocorrelation Function (ACF) and Correlogram Indication are not stationary
    a. AC (Autocorrelation) and Partial Autocorrelation (PAC) correlogram graphs pass the limit values.
    b. The AC and PAC statistical values are above 0.5.
3. Unit Root Test

   The unit root test is one component in testing data stationarity. If any variable consists of a unit root, is non-stationary and combines with variable If anything else is not stationary then the two series will form stationarity in the cointegration relationship. Testing stationarity on a variable The aim is to see whether the data contains a linear combination terintegrasi establish stationarity or equilibrium relationships.
    a. Dickey-Fuller Test

       The Dickey Fuller test is used to test whether a time series is stationary or not. The concept of this test is closely related to random walks. One important thing to know about this test is that the random walk in question is a situation where the series is not stationary. Consider the following AR(1) process.

$$y_t = \Phi y_{t-1} + \mathscr{E}t \quad (1)$$

       If $\Phi = 1$, then the series becomes a random walk. Then the AR(1) process above is expressed in the Dickey-Fuller equation by subtracting both sides of y_{t-1}.

$$yt - y_{t-1} = \Phi y_{t-1} + \mathscr{E}t - y_{t-1}$$

$$\Delta y_t = (\Phi - 1)y_{t-1} + \mathscr{E}_t$$

$$\Delta y_t = \Phi y_{t-1} - y_{t-1} + \mathscr{E}_t \qquad (2)$$

Then it can be seen that the hypothesis of this equation is as follows:

$H_0 : \Phi = 0$ (there is a unit root)

$H_1 : \Phi < 0$ (there is no unit root)

The Dickey-Fuller test has a one-way hypothesis. The unit root test results can be seen by comparing the t-statistic results with the McKinnon critical value, where:

yt = independent variable

Phi = independent variable coefficient

$\mathscr{E}$ = residual value

H = hypothesis

b. Augmented Dickey-Fuller Test

Perbandingan Augmented Dickey-Fuller test dengan Dickey-Fuller test is by adding the number of lags to the first difference of variable dependent to overcome the autocorrelation of variable which was removed. Augmented Dickey-Fuller model test for the null hypothesis that has train stochastic and alternative with deterministic trends.

c. Phillip-Perron Test

An important assumption in the Dickey-Fuller test is that the error term values are independent and identically distributed. The Augmented Dickey-Fuller test adapts the Dickey_fuller test to overcome the possibility of serial correlation in the error term by adding a lag to the difference in the explanatory variables.

Philip-Perron test is a nonparametric statistical mode to overcome the occurrence of serial correlation in the error term without adding lag to the difference in the explanatory variables. Philip's approach-Perron test adds a correction factor to the Dickey-Fuller test.

4. Unit Root Test and Structural Break

A stationary series that is along a deterministic trend and experiences permanent shifts over a period sometimes gives rise to some failure in slope changes when using the Augmented Dickey-Fuller test. Unit root tests that do not include breaks will have weak power.

If the break in a series is known, adjustments in the Augmented Dickey-Fuller test relatively guarantee the presence of a deterministic component in the data generating process. One approach to testing the possibility of a break in a series is a sequential approach which is calculated based on usage samples in full.

2.4 Autoregressive Integrated Moving Area (ARIMA)

Autoregressive Integrated Moving Average (ARIMA) model is a model that completely ignores independent variables in making forecasts. ARIMA uses the past and present values of the dependent variable to produce accurate short-term forecasts. ARIMA is suitable if observations from a time series are statistically related to each other (dependent). The aim of this model is to determine a good statistical relationship between the predicted variables and the historical values of these variables so that forecasting can be done using this model (Hendranata, 2003; Pradana et al., 2020).

Advantages and Disadvantages of ARIMA (Autoregressive Integrated Moving Average) Method (Hyndman, 2021):

1. Advantages of ARIMA Models
   1) ARIMA is a relatively simple model to apply, so it can be used for time series data without requiring in-depth knowledge of probability theory.
   2) ARIMA is very effective for forecasting data that is stationary or can be made stationary through the differencin process.
   3) ARIMA takes into account past values (autoregressive) as well as past errors (moving average), so it can capture patterns in the data dynamically.
   4) ARIMA models can be applied to data with different time scales (e.g. daily, monthly, or yearly) by adjusting the model parameters.
   5) ARIMA can be extended to SARIMA to accommodate seasonal patterns, making it suitable for data that has both trends and seasonality.
2. Disadvantages of ARIMA Models
   1) One of the main limitations of ARIMA is the assumption that the data must be stationary. If the data is not stationary, differencing or other

transformations need to be performed, which may reduce the precision of the results.

2) ARIMA may not be able to capture complex or non-linear patterns in the data, making it less effective if the data pattern is too complex or has a non-linear fulfilment.
3) ARIMA is mainly used for univariate (one variable), so it is less suitable for models with multiple variables. ARIMA models do not take into account the influence of external variables directly.
4) The selection of ARIMA (p, d, q) parameters must be done carefully. Errors in parameter selection can result in poor prediction results.
5) ARIMA does not handle volatility or variance changes in the data well, so it is not suitable for data with high volatility. GARCH (Generalised Autoregressive Conditional Heteroskedasticity) models are usually better suited under these conditions.

The ARIMA model consists of three basic steps, namely the identification stage, the assessment stage and testing and inspection diagnostics. Furthermore, the ARIMA model can be used to forecast if the model obtained is adequate. The stages that must be carried out in the ARIMA Model are (Pradana et al., 2020):

2.4.1 Stationarity and Nonstationarity

The thing to note is that most periodic series are non-stationary. The AR and MA aspects of the ARIMA model only concern stationary periodic series. Stationarity means there is no growth or decline in the data. The data should be roughly horizontal along the time axis. In other words, data fluctuations are around a constant average value, independent of time and the variance of these fluctuations remains essentially constant at all times.

The non-stationary time series must be converted into stationary data by doing *differencing*. What is meant by differencing is calculating changes or differences in observation values. The difference value obtained is checked again whether it is stationary or not. If it is not stationary then do it *differently* Again. If the variance is not stationary, then a logarithmic transformation is carried out.

2.4.2 ARIMA Model Classification

The Box-Jenkins model (ARIMA) is divided into 3 groups, namely: model autoregressive (AR), moving average (MA), and ARIMA mixed models (autoregressive moving average) which has the characteristics of the first two models.

1. Autoregressive Model (AR), a model that explains the movement of a variable through the variable itself period previous time. General form of the model autoregressive with order p (AR(p)) or ARIMA model (p,0,0) expressed as follows:

   $$Z_t = \mu' + \varphi_1 Z_{t-1} + \varphi_2 Z_{t-2} + \ldots + \varphi_p Z_{t-p} + a_t \quad (3)$$

   Information:

   μ' = a constant

   φ_p = autoregressive parameter to -p

   a_t = error value at time t

2. Moving Average Model (MA), a moving average that looks at the movement of the variable through residuals in the past. General form of the model moving average order q (MA(q)) or ARIMA (0,0,q) is expressed as follows:

   $$Z_t = \mu' + a_t - \theta_1 a_{t-1} – \theta_2 a_{t-2} – \theta_q a_{t-k} \quad (4)$$

   Information:

   μ' = a constant

   θ_1-θ_q = moving average parameters

   a_{t-k} = error value at time t – k

3. Mixed Model

   a) ARMA Process

   The general model for a mixture of AR(1) and MA(1) processes, for example ARIMA (1,0,1) is expressed as follows:

   $$Z_t = \mu' + \varphi_1 Z_{t-1} + \varphi_2 Z_{t-2} + \ldots + \varphi_p Z_{t-p} - \theta_1 a_{t-1} – \theta_2 a_{t-2} – \theta_q a_{t-q} \quad (5)$$

   or

   $$\underbrace{(1-\varphi_1 B-\varphi_2 B^2 - \ldots -\varphi_p B^p)}_{AR(1)} Zt = \mu' + \underbrace{(1-\theta_1 B-\theta_2 B^2 - \ldots -\theta_q B^q)}_{MA(1)} a_t \quad (6)$$

   b) ARIMA Process

When the *non*-stationarity model is added to the ARMA process mixture, so the general ARIMA model (p,d,q) is satisfied. The equation for the simple case of ARIMA (1,1,1) is:

$\Phi_P (B)(1-B)^d Zt = \theta_q (B) a_t$ (7)

Information:

p = AR Orde

d = Non Seasonal Differencing Orde

q = MA Orde

2.4.3 Seasonality and ARIMA Models

Seasonality is defined as a pattern that repeats itself over a fixed time interval. For stationary data, seasonal factors can be determined by identifying autocorrelation coefficients at two or three time-lag which is significantly different from zero. Autocorrelation that is significantly different from zero indicates the presence of a pattern in the data. To recognize the presence of seasonal factors, one must look for high autocorrelation. To handle seasonality, a short common notation is:

ARIMA (p,d,q) (P,D,Q)S

Information:

(p,d,q) = non-seasonal part of the model

(P,D,Q) = seasonal part of the model

S = number of periods in season

2.4.4 Identification

The identification process of a seasonal model depends on statistical tools in the form of autocorrelation and partial autocorrelation, as well as knowledge of the system being studied.

2.4.5 Parameter Estimation

Ada dua cara yang mendasar untuk mendapatkan parameter-parameter tersebut:

1. By trial and error, testing several different values and select one value (or set of values, if there is more than one parameter to be estimated) that minimizes the sum of the squares of the residual values.
2. Iterative refinement, choosing an initial estimate and then letting a computer program refine the estimate iteratively.

2.4.6 Model Parameter Testing

1. Testing each model parameter partially (t-test)
2. Overall model testing (Overall F test)

   The model is said to be good if the error value is random, which means that is no longer has a specific pattern. In other words, the obtained model can capture the existing data patterns. To see the randomness of the value *error* Testing was carried out on the autocorrelation coefficient value of *error* using one of the following two statistics:

   a) Q Box and Pierce Testing:

   $Q = n' \sum_{k=1}^{m} r\hat{}2k$ (8)

   b) Ljung-Box Test:

   $Q = n'(n'+2) \sum_{k=1}^{m} r^{2k} /(n' - k)$ (9)

   Chi Squared Spread (x^2) with random degrees

   (db) = (k-p-q-P-Q) (10)

   Information:

   n' = n-(d+SD)

   d = differentiation ordo it is not a seasonal factor

   D = differentiation ordo a seasonal factor

   S = number of periods in season

   m = *lag* maximum time

   rk = autocorrelation for time lags 1, 2, 3, 4,..., k

   Test criteria:

   If $Q \le x^2$ (α ,db) , means: value *error* nature *random* (acceptable model).

   If $Q > x^2$ (α ,db) , means: value *error* non-native *random* (unacceptable model).

2.4.7 Selection of The Best Models

To determine the best model that can be used standard error estimate as follow:

$$S = \left[\frac{SSE}{n-n_p}\right]^{1/2} = \left[\frac{\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2}{n-n_p}\right]^{1/2} \tag{11}$$

Information:

Yt = actual value at time t

Ytˆ = estimated value at time t

The best model is the one that has the value standard error estimate (S) is the smallest. Apart from value standard error estimate, the average percentage forecasting error (MAPE) value can also be used as consideration in determining the best model, namely:

$$\text{MAPE} = \left(\frac{1}{n}\sum_{t=1}^{n}\left|\frac{Yt-\hat{Y}t}{Yt}\right|\right) \text{ x } 100\% \tag{12}$$

Information:

MAPE = average value of forecasting error percentage

Yt = actual value at time t

Ytˆ = estimated value at time t

T = number of forecasting in estimated periods.

2.4.8 Forecasting with the ARIMA Model

The notation used in ARIMA is an easy and general notation, but to use it in forecasting requires elaborating the equation and making it a more general regression equation.

2.5 Generalized Autoregressive Conditional Heteroskedasticity (GARCH)

The Autoregressive Moving Average (ARCH) was first introduced by Engle (1982) it was developed to answer the problem of volatility in financial data. This model was developed into Generalized Autoregressive Moving Average (GARCH) by Bollerslev in 1986. Sukma (2012) showed that the GARCH method was able to produce smaller error values compared to the EWMA (Exponential Weighted Moving Average) model. The advantages of the GARCH model compared to other time series models are:

1. This model does not view heteroscedasticity as a problem, but instead uses it to a form model.
2. This model not only produces forecasts of variable Y, but also forecasts of variance. Changes in variance are very important example for understanding coal price movements.

According to Bollerslev (1986), the ACF and PACF patterns, apart from being used to identify time series behavior from ARIMA in the form of conditional mean, can also be used to assist the square process in identifying GARCH behavior in conditional variance equations (Heteroskedastic). Some applications that use linear ARCH(p) models require large p values. However, this creates problems in determining the number of parameters α_0, α_1, …, α_p, which is describe the time evolution of the economic time series. The GARCH process is defined by the following equation:

$$\sigma_1^2 = \alpha_0 + \alpha_1 \mathscr{E}_{t-1}^2 + \ldots + \alpha_p \mathscr{E}_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_p \sigma_{t-p}^2 + \beta_q \mathscr{E}_{t-q}^2 \quad (13)$$

$\alpha_0 > 0$, $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q \geq 0$

Information:

$\alpha 1, \ldots, \alpha p, \beta 1, \ldots, \beta q$ = Parameter control

$\mathscr{E}1$ = Random variable with zero mean

$\sigma t2$ = Variance

While conventional time series and econometric models operate under an assumption of constant variance, the ARCH (Autoregressive Conditional Heteroskedastic) process introduced in Engle (1982) allows the conditional variance to change over time as a function of past errors leaving the unconditional variance constant.

This type of model behavior has already proven useful in modelling several different economic phenomena. In Engle (1982), Engle (1983) and Engle and Kraft (1983), models for the inflation rate are constructed recognizing that the uncertainty of inflation tends to change over time. In Coulson and Robins (1985) the estimated inflation volatility is related to some key macroeconomic variables. Models for the term structure using an estimate of the conditional variance as a proxy for the risk premium are given in Engle, Lilien and Robins (1985). The same idea is applied to the foreign exchange market in Domowitz and Hakkio (1985). In Weiss (1984) ARMA models with ARCH errors are found to be successful in modelling thirteen different U.S. macroeconomic time series. Common to most of the above applications however, is the introduction of a rather arbitrary linear declining lag structure in the condi tional variance equation to take account of the long memory typically found in empirical work, since estimating a totally free lag distribution often will lead to violation of the non-negativity constraints.

2.5.1 The GARCH(p,q) process

The ARCH process introduced by Engle (1982) explicitly recognizes the difference between the unconditional and the conditional variance allowing the latter to change over time as a function of past errors. The statistical properties of this new parametric class of models has been studied further in Weiss (1982) and in a recent paper by Milhoj (1984). In empirical applications of the ARCH model a relatively long lag in the conditional variance equation is often called for, and to avoid problems with negative variance parameter estimates a fixed lag structure is typically imposed. In this light it seems of immediate practical interest to extend the ARCH class of model to allow for both a longer memory and a more flexible lag structure.

Let ε_t denote a real-valued discrete-time stochastic process, and Ψ_t, the information set (σ-field) of all information through time t. The GARCH(p,q) process (Generalized Autoregressive Conditional Heteroskedasticity) is given by

$$\varepsilon_t \; \Psi_{t-1} \sim N(\theta, h_t), \tag{14}$$

$$\begin{aligned} h_t &= a_0 + \sum_{i-1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i-1}^{p} \beta_i h_{t-i} \\ &= a_0 + A(L)\varepsilon_t^2 + B(L)h_t \end{aligned} \tag{15}$$

Where

$p \geq 0, \;\; q > 0$

$\alpha_0 > 0, \; \alpha_i > 0, \; i = 1,\ldots,q,$

$\beta_t \geq 0, \; i = 1,\ldots,q.$

For p = 0 the process reduces to the ARCH(q) process, and for p = q = 0 ε_t is simply white noise. In the ARCH(q) process the conditional variance is specified as a linear function of past sample variances only, whereas the GARCH(p,q) process allows lagged conditional variances to enter as well. This corresponds to some sort of adaptive learning mechanism.

The GARCH(p,q) regression model is obtained by letting the ε_t's be innovations in a linear regression,

$$\varepsilon_t = y_t - x_t' b \tag{16}$$

where y_t is the dependent variable, x_t a vector of explanatory variables, and b is a vector of unknown parameters. This model is studied in some detail in section 5. If all the roots of 1 - B (z) = 0 lie outside the unit circle, (2) can be rewritten as a distributed lag of past ε_t's,

$$\begin{aligned} h_t &= a_0(1 - B(1))^{-1} + A(L)(1 - B(L))^{-1}\, \varepsilon_t^2 \\ &= a_0(1 - \textstyle\sum_{i=1}^{p} \beta \mathrm{i})^{-1} + \sum_{i=1}^{\infty} \delta i\ \varepsilon_t^2 \end{aligned} \tag{17}$$

which together with (1) may be seen as an infinite-dimensional ARCH(oo) process. The 8i's are found from the power series expansion of $D(L) = A(L)(1 - B(L))^{-1}$,

$$\begin{aligned} \delta i &= \alpha_i + \textstyle\sum_{i=1}^{\infty} \beta j \delta i, && i = 1,\ldots,q, \\ &= \textstyle\sum_{i=1}^{\infty} \beta j \delta i, && i = q + 1,\ldots, \end{aligned} \tag{18}$$

where n = min(p, i - 1}. It follows, that if B(1) < 1, 6i will be decreasing for i greater than m = max{p, q}. Thus if D(1) < 1, the GARCH(p,q) process can be approximated to any degree of accuracy by a stationary ARCH(Q) for a sufficiently large value of Q. But as in the ARMA analogue, the GARCH process might possibly be justified through a Wald's decomposition type of arguments as a more parsimonious description.

From the theory on finite-dimensional ARCH(q) processes it is to be expected that D(1) < 1, or equivalently A(1) + B(1) < 1, suffices for wide-sense stationarity; cf. Milhoj (1984). This is indeed the case.

Theorem 1. The GARCH(p,q) process as defined in (1) and (2) is wide-sense stationary with $E(\varepsilon_t) = 0$, $\mathrm{var}(\varepsilon_t) = \alpha_0 (1 - A(1) - B(1))^{-1}$ and $\mathrm{cov}(\varepsilon_t, \varepsilon_s) = 0$ for $t \neq s$ if and only if A(1) + B(1) < 1.

2.5.2 The GARCH(1,1) process

The simplest but often very useful GARCH process is of course the GARCH(1,1) process given by (1) and

$$h_t = \alpha_0 a_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}, \qquad \alpha_0 > 0,\ \alpha_1 \geq 0,\ \beta_1 \geq 0. \qquad (19)$$

From Theorem 1, a 1 +/31 < 1 suffices for wide-sense stationarity, and in general we have:

Theorem 2, For the GARCH(1, 1) process given by (1) and (6) a necessary and sufficient condition for existence of the 2mth moment is

$$\mu\,(\alpha_1, \beta_1, m) = \sum_{j=0}^{\infty} \binom{j}{m}\ \alpha_j\ \alpha_1^j\ \beta_1^{m-j} < 1, \qquad (19)$$

where

$$\alpha_0 = 1, \quad \alpha_j = \prod_{i=1}^{j} (2j - 1), \quad j = 1 \dots .$$

The 2mth moment can be expressed by the recursive formula

$$E(\varepsilon_t^{2m}) = \alpha_m \left[\sum_{n=0}^{m-1} \alpha_n^{-1}\, E(\varepsilon_t^{2n}) \alpha_0^{m-n} \binom{m}{m-n} \mu(\alpha_1, \beta_1, n) \right] \times [1 - \mu(\alpha_1, \beta_1, m)]^{-1}$$

The conditions for existence of the first twelve moments are illustrated in fig. 3. It follows by symmetry that if the 2 ruth moment exists, E(ε_t^{2m-1}) = 0. For β_1 = 0, () reduces to the well-known condition for the ARCH(l) process, $\alpha_m\ \alpha_1^m$ > 1. Thus if $\alpha_1 > (\alpha_m)^{-1/m}$ in the ARCH(l) process, the 2mth moment does not exist, whereas even if $\sum_{i=1}^{\infty} \delta i = \alpha_1 (1 - \beta_1)^{-1} > (\alpha_m)^{-1/m}$ in the GARCH(1,1) process, the 2mth moment might very well exist because of the longer memory in this process.

In the GARCH(1,1) process the mean lag in the conditional variance equation is given by

$$\gamma = \sum_{i=1}^{\infty} i\delta_i \Big/ \sum_{i=1}^{\infty} \delta_i = (1 - \beta_1)^{-1}, \qquad (20)$$

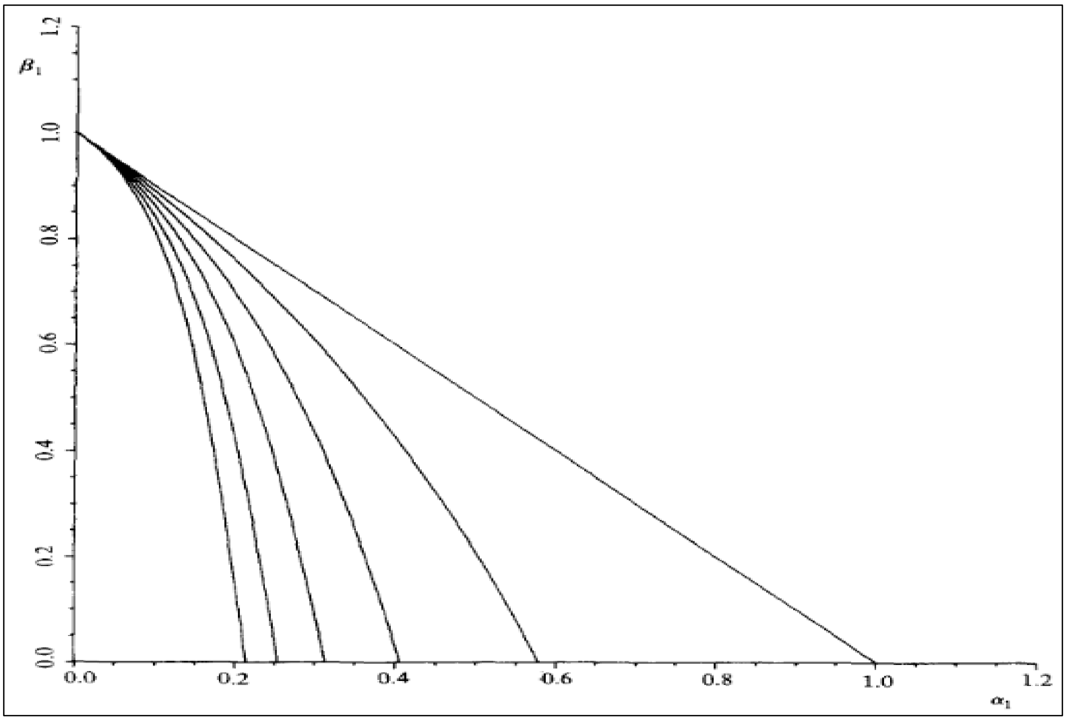

Figure 3 Moment Conditions for GARCH(1,1)(Engle, 1982).

and the median lag is found to be

$\gamma = -log2/log\beta_1$,

Where $\sum_{i=1}^{\infty} i\delta_i / \sum_{i=1}^{\infty} \delta_i = \frac{1}{2}$ and the δ_i's.

If $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$, the fourth-order moment existed by Theorem 2

$\mathrm{E}(\varepsilon_t^2) = \alpha_1(1 - \alpha_1 - \beta_1)^{-1}$,

and

$\mathrm{E}(\varepsilon_t^4) = 3\alpha_0^2(1 + \alpha_1 + \beta_1)[(1 - \alpha_1 - \beta_1)(1 - \beta_t^2 - 2\alpha_1\beta_1 - 3\alpha_1^2)]^{-1}$

The coefficient of kurtosis is therefore

$\mathrm{K} = (\mathrm{E}(\varepsilon_t^4) - 3\mathrm{E}(\varepsilon_t^2)^2)\mathrm{E}(\varepsilon_t^2)^{-2}$

$\mathrm{K} = 6\alpha_1^2(1 - \beta_t^2 - 2\alpha_1\beta_1 - 3\alpha_1^2)^{-1}$,

which is greater than zero by assumption. Hence the GARCH(1,1) process is leptokurtic (heavily tailed), a property the process shares with the ARCH(q) proces

2.5.3 Autocorrelation (AC) and Prtial Autocorrelation (PACF) Structure

The use of autocorrelation and partial autocorrelation functions to identify and check time series behaviour of the ARMA form in the conditional mean is well established. In this section, the autocorrelation and partial autocorrelation functions for the squared process are shown to be useful in identifying and checking time series behaviour in the conditional variance equation of the GARCH form. The idea of using the squared process to check for model adequacy is not new, where it is found that some of the series modelled in Box and Jenkins (1976) exhibit autocorrelated squared residuabs even though the residuals themselves do not seem to be correlated over time.

Consider the general GARCH(p,q) process as specified in (1) and (2), and let us assume the process has finite fourth-order moment. Let the covariance function for ε_t^2 be denoted.

$$\gamma_n = \gamma_{-n} = cov(\varepsilon_t^2, \varepsilon_{t-n}^2) \tag{21}$$

The general conditions for the existence of finite fourth-order moment are unknown. However, given a specific order of the model the conditions may be derived following the same line of arguments as lead to Theorem 2 for the GARCH(1,1) process. For instance the necessary and sufficient condition for the GARCH(1,2) process is found to be

$$\alpha_2 + 3\alpha_1^2 + 3\alpha_2^2 + \beta_1^2 + 2\alpha_1\beta_1 - 3\alpha_2^3 + 3\alpha_1^2\alpha_2 + 6\alpha_1\alpha_2\beta_1 + \alpha_1\beta_1^2 < 1$$

and for the GARCH(2,1) the condition is

$$\beta_2 + 3\alpha_1^2 + \beta_1^2 + \beta_2^2 + 2\alpha_1\beta_1 - \beta_2^3 - \alpha_1^2\beta_2 + 2\alpha_1\beta_1\beta_2 + \beta_1^2\beta_2 < 1$$

Thus, the first p autocorrelations for ε_t^2 depend 'directly' on the parameters $\alpha_1, \ldots, \alpha_q$, $\beta_1, \ldots, \beta_p$, but given $\rho_p, \ldots, \rho_{p+1-m}$ the above difference equation uniquely determines the autocorrelations at higher lags. This is similar to the result for the autocorrelations for an ARMA(m,p) process. Note also, that (24) depends on the parameters $\alpha_1, \ldots, \alpha_q$, $\beta_1, \ldots, \beta_p$, only through $\varphi_1, \ldots, \varphi_m$.

Let ϕ_{kk} denote the *k*th partial autocorrelation for ε_t^2 found by solving the set of *k* equations in the k unknown $\phi_{k1}, \ldots, \phi_{kk}$.

$$\rho_n = \sum_{i=1}^{k} \phi_{ki}\rho_{n-i}, \qquad n = 1, \ldots, k. \tag{22}$$

By (24) ϕ_{kk} cuts off after lag q for an ARCH(q) process

$$\phi_{kk} \neq 0, \qquad k \leq q$$
$$\phi_{kk} = 0, \qquad k > q.$$

This is identical to the behaviour of the partial autocorrelation function for an AR(q) process. Also from (24) and well known results in the time series literature, the partial autocorrelation function for ε_t^2, for a GARCH(p,q) process is in general non-zero but dies out. In practice, of course, the ρ_n's and ϕ_{kk}'S will be unknown. However, the sample analogue, sayρ_n, yields a consistent estimate for ρ_n, and ϕ_{kk} is consistently estimated by the *k* th coefficient, say ϕ_{kk}, in a kth-order autore- gression for ε_t^2. These estimates together with their asymptotic variance under the null of no GARCH 1/T can be used in the preliminary identification stage, and are also useful for diagnostic checking.

2.5.4 Estimation of The GARCH Regression Model

In this section we consider maximum likelihood estimation of the GARCH regression model (1), (2), (3). Because the results are quite similar to those for the ARCH regression model, our discussion will be very schematic**.**

Let $z_t' = \left(1,\ \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-q}^2, h_{i-1}, \ldots, h_{i-p}\right)$, $z_t' = \left(\alpha_0, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p\right)$ and $\theta \in \Theta$, where $\theta = (b', w')$ and Θ is a compact subspace of a Euclidean space such that ε_t^2 possesses finite second moments.

The log likelihood function for a sample of T observations is apart from some constant

$$L_T(\theta) = T^{-1}\sum_{t=1}^{T} \mathrm{l}_t(\theta), \tag{23}$$

$$\mathrm{l}_t(\theta) = -\frac{1}{2}\log h_t - \frac{1}{2}\varepsilon_t^2 h_t^{-1}.$$

2.5.5 Testing for GARCH

Because of the complication involved in estimating a GARCH process, it seems of interest to have a formal test for the presence of GARCH instead of just relying on the more informal tools. Consider the GARCH(p,q) regression model (27). As in Engle and Kraft (1983) let us partition the conditional variance equation

$$h_t = z'_t\omega = z'_{1t}\omega_1 + z'_{2t}\omega_2, \tag{24}$$

The Lagrange multiplier test statistic for H_0: $\omega_2 = 0$ is then given by

$$\varepsilon_{LM} = \frac{1}{2} \int_{0}^{'} Z_0 (Z'_0 Z_0)^{-1} Z'_0 \tag{25}$$

The alternative as represented by z_{2t} needs some consideration. Straight forward calculations show that under the null of white noise, $Z'_0 Z_0$ is singular if both p > 0 and q > 0, and therefore a general test for GARCH(p,q) is not feasible. In fact if the null is an ARCH(q) process, $Z'_0 Z_0$ is singular for GARCH (r_l, q+r_2) alternatives, where $r_1 > 0$ and $r_2 > 0$. It is also interesting to note that for an ARCH(q) null, the LM test for GARCH(r,q) and ARCH(q + r) alternatives coincide. This is similar to the results in Godfrey (1978), where it is shown that the LM tests for AR(p) and MA(q) errors in a linear regression model coincide and that the test procedures break down when a full ARMA(p, q) model is considered. These test results are, of course, not peculiar to the LM test, but concern the Likelihood Ratio and the Wald tests as well.