NPL forecasting under a Fourier residual modified model: An empirical analysis of an unsecured consumer credit provider in South Africa

Presented to

UNIVERSITY OF CAPE TOWN

In partial fulfilment of the requirements for the degree of

Master of Commerce (Financial Management)

By

Pranisha Luckan

Supervisor: Chun-Sung Huang

2016

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Commerce at the University of the Cape Town. It has not been submitted before for any degree or examination to any other University.

Pranisha Luckan

Disclaimer

The market data used in this study is private and confidential. The credit provider has verified the accuracy of the data and reserves the right to access this data. A copy of this dissertation may only be published in the library at the University of Cape Town after 31 March 2017.

____

Pranisha Luckan

____

The credit provider

Abstract

Forecasting nonperforming loans (NPLs) is a primary objective for credit providers. NPL forecasts assist in financial budgeting and provisioning for bad debts. The difficulty in accurately identifying the determinants of domestic NPLs has led to a review of time series forecasting techniques. This dissertation explores whether a forecasting model combining a traditional time series approach with a Fourier series residual modification technique performs well in projecting NPLs. It also seeks to establish if selecting an adequate time series model before modifying its residual terms is of benefit. Using the data of an unsecured consumer credit provider in South Africa, the in-sample and out-of-sample performance for a seasonal time series model and residual modified model were evaluated. The results demonstrate that a time series model performs well but the out-of-sample forecasting errors may be reduced by including the lowest Fourier frequencies to modify the residual terms.

Acknowledgements

I am thankful to my supervisor, Chun-Sung for his input and guidance.

I would like to express gratitude to my manager and colleagues.

I would like to acknowledge my family for their love, support and encouragement.

I certify that it is my own work and all references used are accurately reported in the text.

List of abbreviations

ABIL: African Bank Investments Limited

ACF: Autocorrelation function

ANN: Artificial neural network

AR: Autoregressive

ARDL: Autoregressive Distributed Lag

ARIMA: Autoregressive Integrated Moving Average

ARMA: Autoregressive Moving Average

BoE: Bank of England

CAR: Capital adequacy ratio

CAGR: Compound annual growth rate

CESEE: Central, Eastern and South-eastern Europe

CPI: Consumer Price Index

DFT: Discrete Fourier transform

DW: Durbin-Watson

DWT: Discrete wavelet transform

EU: European Union

ECB: European Central Bank

FDI: Foreign Direct Investment

FFT: Fast Fourier Transform

FOMC: Federal Open Market Committee

FRB: Federal Reserve Bank

GAAP: Generally Accepted Accounting Principles

GCC: Gulf Cooperation Council

GDP: Gross Domestic Product

GMM: Generalised Method of Moments

IAS: International Accounting Standards

IFRS: International Financial Reporting Standards

IMF: International Monetary Fund

MA: Moving average

MAE: Mean absolute error

MAPE: Mean absolute percentage error

MoM: Month-on-month

NCA: National Credit Act
NCR: National Credit Regulator
NN: Neural network
NNs: Neural networks
NPL: Nonperforming loan
NPLs: Nonperforming loans
NPLR: Nonperforming loan ratio
NPV: Net present value
OLS: Ordinary least squares
PACF: Partial autocorrelation function
PP: Phillips-Perron
PSO: Particle Swarm Optimization
Q1: Quarter one
Q2: Quarter two
Q3: Quarter three
Q4: Quarter four
QoQ: Quarter-on-quarter
QQ: Quantile-Quantile
ROA: Return on assets
ROE: Return on equity
ROI: Return on investment
SA: South Africa
SARB: South African Reserve Bank
SARIMA: Seasonal Autoregressive Integrated Moving Average
SM: Special mention
UK: United Kingdom
USA: United States of America
VAR: Vector Autoregressive
VECM: Vector Error Correction Model
YoY: Year-on-year

List of figures

Page

List of tables

Page

Table of contents

Chapter 1

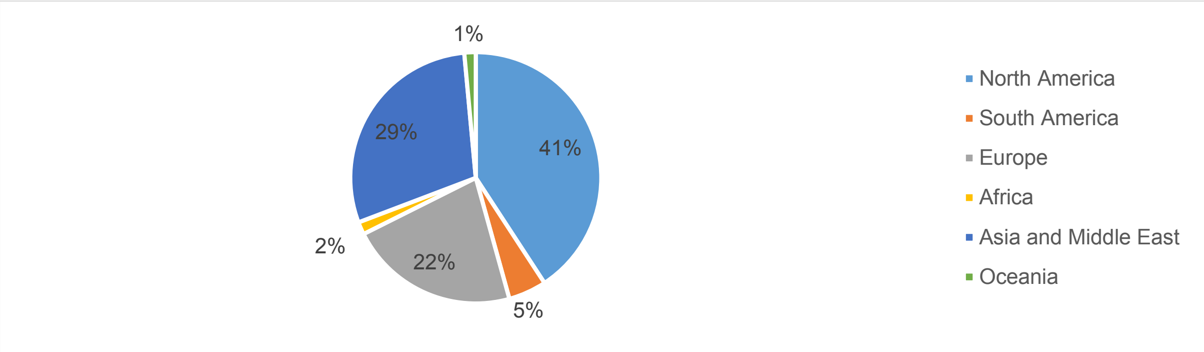
Introduction

Chapter one covers the definition of an NPL and provides a discussion of the consumer credit industry. A background to the reason for this investigation is presented by reviewing the Greek crisis and African Bank failure. Trends in the global and domestic NPLR conclude this chapter.

1.1. The consumer credit industry

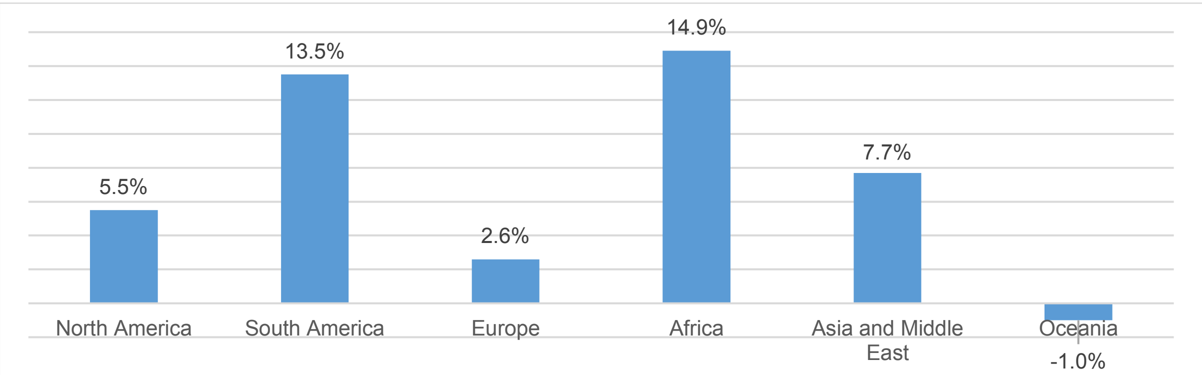
By the end of 2012, the outstanding value of consumer credit loans worldwide was estimated at 6.383 billion Euros. North America and Europe are dominant players, capturing almost two thirds of the international consumer credit market, as shown in Figure 1.1. Emerging economies have gradually increased their share of the market, registering healthy YoY growth in 2012, evident from Figure 1.2.

Figure 1.1: Proportion of outstanding consumer loans worldwide at year end 2012



Source: Crédit Agricole Consumer Finance (2013)

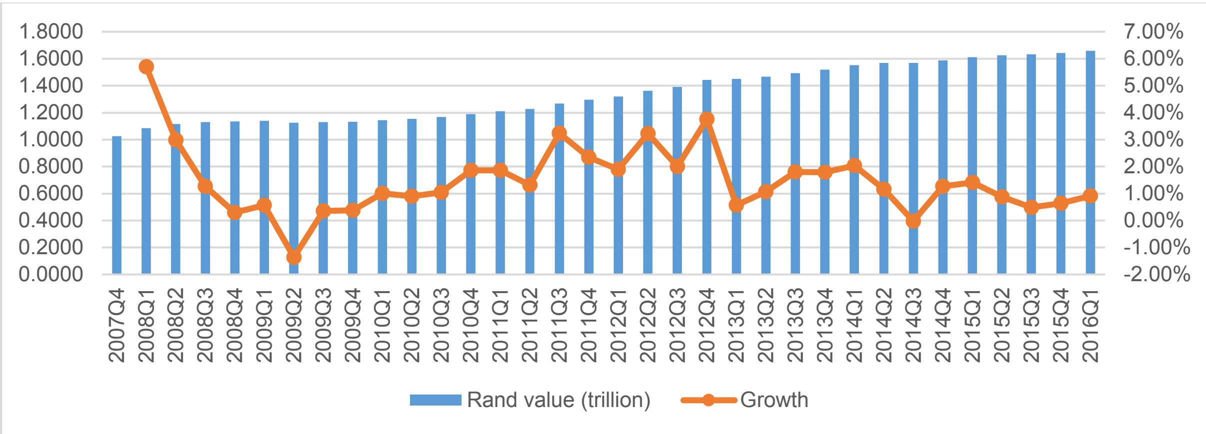
Figure 1.2: YoY growth in outstanding consumer loans at year end 2012



Source: Crédit Agricole Consumer Finance (2013)

By Q1 of 2016, the size of the consumer credit market in SA was approximately R1.658 trillion. Between Q3 of 2011 and Q4 of 2012, growth in the Rand value of outstanding consumer credit facilities was greater than 1.9%. QoQ growth has trended upwards in the most recent three quarters, but was still below 1%, as observed in Figure 1.3.

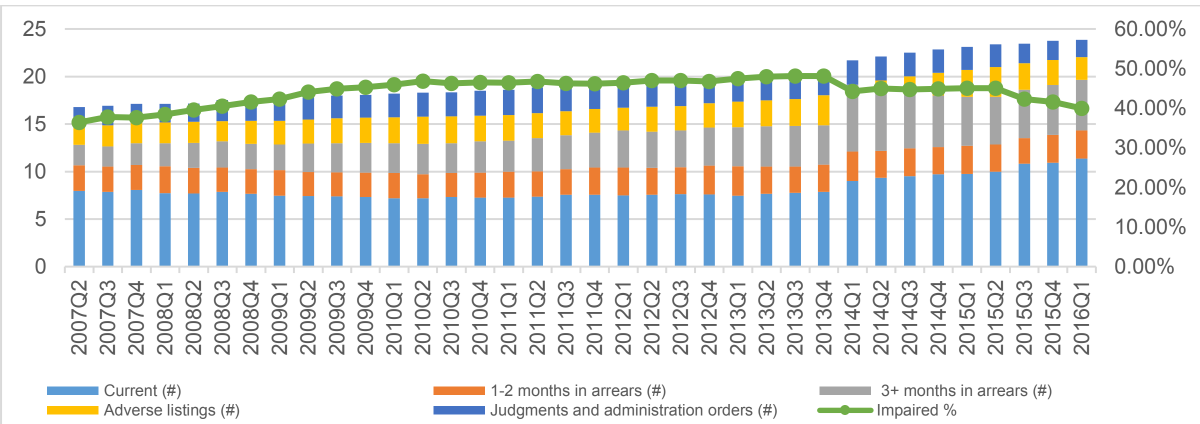
Figure 1.3: Gross debtor's book for the consumer credit market in SA



Source: NCR (2016)

Based on quarterly data issued by the NCR, the health of the consumer credit market in SA has improved since the inception of credit amnesty in 2014. The number of credit active customers in SA was 23.88 million by Q1 of 2016, up from 16.78 million in Q2 of 2007. The domestic data reveals that four out of ten active credit consumers are impaired. The percentage of customers three or more months in arrears has increased from 12.8% in Q2 of 2007 to 22.3% by Q1 of 2016. This upward trend in nonperforming customers is observed by the growing grey block in Figure 1.4.

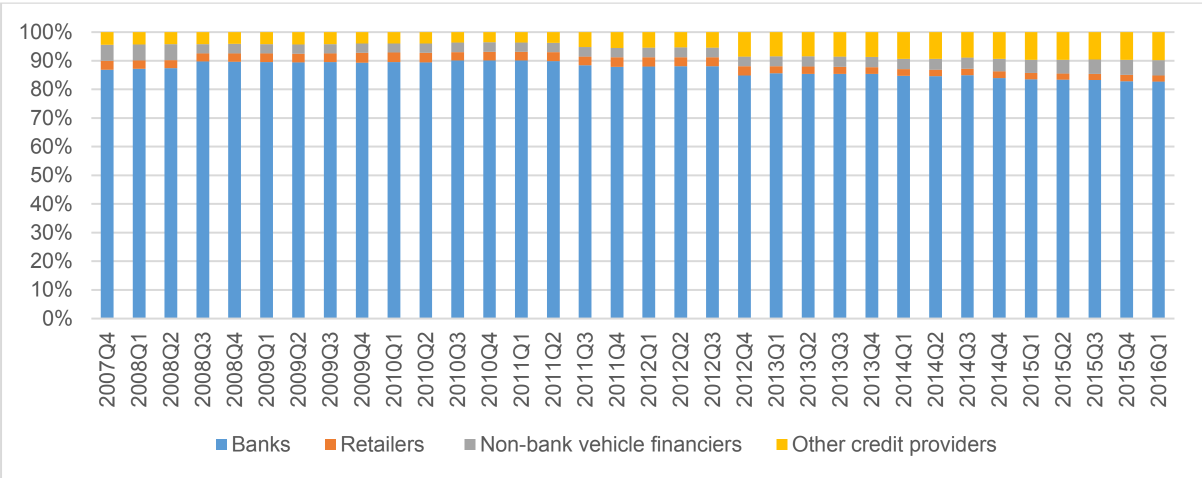
Figure 1.4: Credit standing of active consumers in millions



Source: NCR (2016)

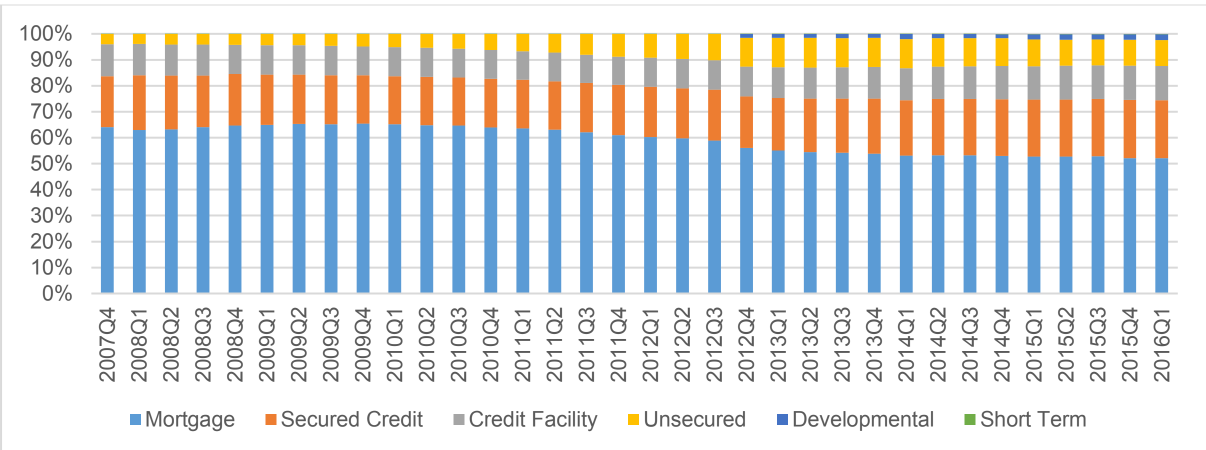
From Figure 1.5, it is evident that there has been a gradual loss in market share for banks as less traditional credit providers gain footprint. When the gross debtor's book for consumer credit is graphed by the facility type, shown in Figure 1.6, the growth in unsecured credit is notable. In Q1 of 2008, unsecured credit contributed 3.87% to the total consumer credit market in SA. This increased to 9.96% by Q1 of 2016. The BoE (2015) has also reported positive YoY growth in unsecured consumer credit since 2013, mainly concentrated in overdrafts and personal loans. The increased consumer credit levels was partly due to the greater availability of such credit from lenders.

Figure 1.5: Gross debtor's book by industry type weighted by balance



Source: NCR (2016)

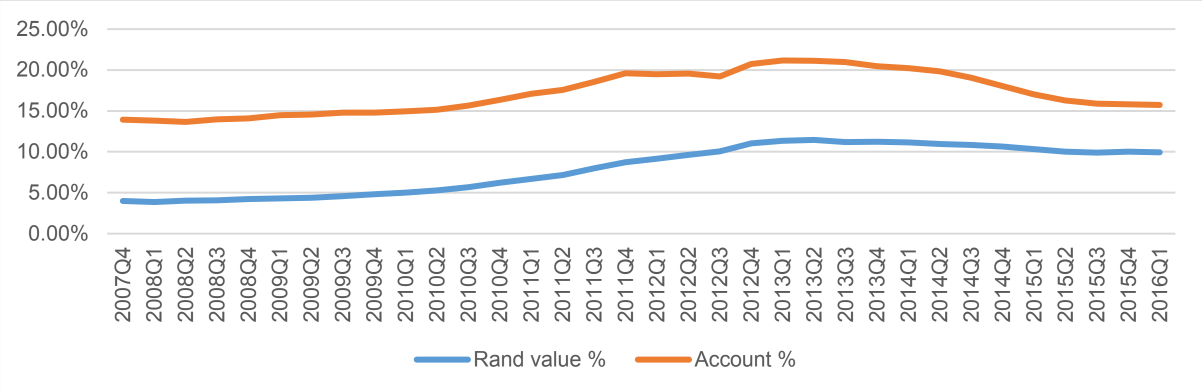
Figure 1.6: Gross debtor's book by credit type weighted by balance



Source: NCR (2016)

Unsecured credit as a percentage of the total Rand value and total number of accounts for the South African consumer credit market is displayed in Figure 1.7. After a steady increasing contribution to the total debtor's book, the Rand value of unsecured credit has flattened at about 10% of the total. The number of unsecured accounts relative to the total credit market represented just over 15% in Q1 of 2016.

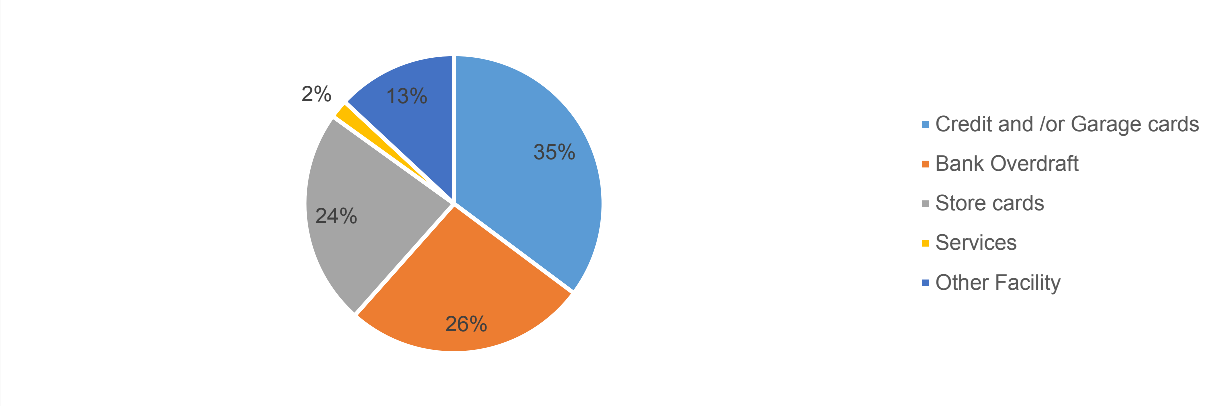
Figure 1.7: Unsecured consumer credit as a percentage of the total market



Source: NCR (2016)

When the type of credit facilities were reviewed, it is evident that there has been increased preference for overdrafts. In Q1 of 2008, overdrafts represented 15.54% of the total Rand value of credit facilities granted in SA. By Q1 of 2016, this had nearly doubled to 26.35%, as observed in Figure 1.8. Overdrafts and credit cards assist with MoM cash flow management and may be subject to seasonal patterns. Transunion (2015) noted QoQ seasonal trends in distressed borrowing, with a spike usually in Q1. The Federal Reserve Bank of Minneapolis (2015) described general purpose credit cards as the most popular type of consumer debt, although the balances owed are typically small. On average, these balances generally follow a seasonal trend, increasing toward the end of the year during the festive period, and being paid down toward the start of the following year. According to Williams (2014), balances on revolving consumer debt in the USA begin to rise in August for school shopping and decrease in September and October, before increasing sharply in November and into December, for holiday travel, gift purchases and festive parties. Credit cards serve to bridge the cash flow gap for periods when heavier than normal consumer spending occurs. Balances are usually paid down during the first three months of the following year. Federal Reserve figures show peaks in delinquency rates in the first quarter, due to consumer overspend during the end of the prior year.

Figure 1.8: Credit facilities granted in Q1 of 2016 for the consumer market



Source: NCR (2016)

1.2. NPL definition

"A loan is nonperforming when payments of interest and/or principal are past due by 90 days or more, or interest payments equal to 90 days or more have been capitalized, refinanced, or delayed by agreement, or payments are less than 90 days overdue, but there are other good reasons—such as a debtor filing for bankruptcy—to doubt that payments will be made in full" (Bloem & Freeman, 2005, page 4).

Li & Zou (2014) describe default or NPLs as the following:

- The non-payment of interest for 90 days after the due date.
- The non-payment of the loaned amount 90 days after its maturity date.
- Restructure of the borrower's debt agreements.
- Declaration of bankruptcy, appointment of administrators or filing liquidation.

An NPL is generally defined as a loan that is overdue greater than 90 days (Bloem & Freeman, 2005). The Basel II accord recommends that banks' capital requirements comply with modern risk management practices and are comparable and comprehensive (Fofack, 2005). Under Basel II, NPLs are past due for greater than 90 days (Islam, Shil & Mannan, 2008).

The NPLR is calculated as: NPLR= $\frac{\text{NPLs}}{\text{Total loans advanced}}$

In its numerator, the ratio has the level of NPLs and its denominator represents the gross value of loans advanced, as reported on the credit provider's balance sheet. The denominator does not only reflect the amount of loans outstanding but rather the total amount extended to borrowers (Li & Zou, 2014).

NPLs refer to loans which do not generate income for relatively long time (Fofack, 2005). According to Islam *et al.* (2008), the NPL definition generally stems from a lending institution's viewpoint. A loan becomes nonperforming when it fails to be recovered in a stipulated time as governed by law. This is consistent with Bloem & Freeman (2005), where a loan is classified as nonperforming or impaired when it is likely that contractual payments will not be made. Accounting and banking practices often term a loan as impaired instead of nonperforming. IAS state that the carrying amount of assets should be decreased by an amount equivalent to the loss resulting from impairments. The Basel Committee on Banking Supervision also labels impairment as the probable lack of payment of amounts due on a loan. With reference to IAS, Bloem & Gorter (2001) state that when the carrying amount of an asset is greater than its estimated recoverable amount, the asset is treated as impaired.

Bloem & Gorter (2001) mention that distinguishing between good and bad loans may involve the use of quantitative criteria, such as the days overdue on payment. It may also regard qualitative criteria, like the information related to a customer's financial standing, as well as the discretion of management about future payments. In absolute terms, good or bad may not exist, as there is instead a sliding scale for measuring credit quality, from default free loans to impaired loans. The following categories are used to classify loans:

- Standard: Principal and interest payments are up-to-date. Under current conditions, repayment difficulties are not expected and no default is projected.
- Watch: If certain conditions are not corrected, it could raise concerns about receiving full repayment.
- Substandard: Inadequate protection, such as a decrease in the collateral value, may create doubt for full repayment. The interest and/or principal payments of the loan may be greater than 90 days in arrears. There are underlying weaknesses that could lead to impairment and loss.

- Doubtful: Collection on the overdue amount is not likely under the current landscape, and the interest and/or principal payments are overdue for greater than 180 days.
- Loss: The facility is virtually uncollectible and interest or principal is in arrears greater than a year.

Sometimes, NPLs may correspond to substandard, doubtful and loss loans. In other cases, only doubtful and loss loans may be classified as NPLs and in some instances, only loss loans are recognised as NPLs (Shingjergji, 2013).

Differences in accounting practices and financial reporting standards cause slight variation in the recognition and treatment of NPLs for lenders. According to Moody's Investor Services (2011), the treatment of NPLs depends on the credit provider's reporting framework. Under IFRS, problem loans are best classified by impaired loans. A loan is deemed impaired if there is objective evidence of a loss event that will influence the reliably estimated future cash flows generated from the loan. Under GAAP, problem loans consist of accruing loans plus non-accruing loans that are overdue for more than ninety days. Bloem & Freeman (2005) state that NPLs ought to be valued at market value as this represents the actual agreed upon price between transacting parties. However, since loans are usually not actively traded, a market equivalent value is applied. Fair value is a close approximation to market value as it estimates the value arising between the counterparties engaging in a market transaction. This fair value is determined by using comparable transactions or by computing the discounted NPV of cash flows generated from similar non-traded instruments. For some countries, this information is difficult to extract because fair valuation may not be practiced. In the absence of such data, the market equivalent value approach uses a nominal loan value minus the expected losses.

Late payment is usually categorized as nonperforming rather than default. At some point, an NPL will be written off as a default loss, which is funded from a financial institution's capital reserves, often at a hundred percent of the notional value outstanding. Post write-off, a portion of this amount may be recovered (Li & Zou, 2014). An NPL is classified as such until there is a write off or interest and/ or principal payments are made on the original loan or any subsequent facility that replaces the original. When an NPL is replaced with new loans, it cannot simply be reclassified as performing. Nonperforming essentially means that some losses on the loan are likely to be incurred as orderly repayment is compromised. It does not mean that all losses will occur and on loans backed by collateral pledges, losses may not necessarily be expected (Bloem & Freeman, 2005).

The definition of arrears is different to nonperforming. Arrears comprises of interest and principal amounts that are unpaid and overdue for payment. Arrears data is generally scrutinized by regulators. In comparison to the 90 day past due definition for NPLs, arrears related impairments are identified earlier. Arrears amounts only include those that are due for payment. If there are payments owed on a loan exceeding 90 days, only the interest and amortization payments are included in the arrears amount, whereas the entire loan amount is recognized as an NPL (Bloem & Freeman, 2005).

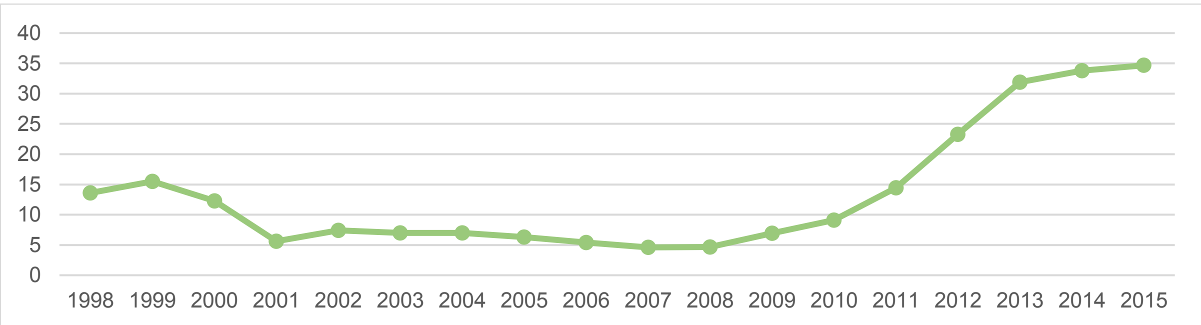
1.3. Background to reason for investigation

The management of credit risk is integral for fiscal stability and economic activity. The NPLR serves as an indicator of financial soundness (Boudriga, Tatak & Jellouli, 2009), as a proxy for credit risk management (Li & Zou, 2014) and as an economy-wide default probability of a banking sector's loan exposure (Akinlo & Emmanuel, 2014). Serwa (2013) and Makri, Tsagkanos & Bellas (2014) describe the NPLR as a commonly used and standardized measure of financial performance for a credit institution. It can be used to evaluate asset quality (Mendoza & Terrones, 2012), establish bank equity values (Aman & Miyazaki, 2006), to build early warning systems for financial crises (Cheang, 2009) and to predict bank failure (Messai & Jouini, 2013).

By May 2007, material presented at the FOMC meeting had already revealed a strain to the subprime mortgage book. There was a substantial uptick in the sixty day plus delinquency for adjustable rate mortgages originated in 2006. This was likely to translate into higher NPL rates (more than 90 days delinquent). The credit crunch ensued in 2008, which suggests that the NPLR was a leading indicator of the crisis (FRB, 2007).

Greece was the centre of the sovereign debt crisis, with the largest budget deficits and public debt in the Eurozone. The country's public finances were strained by the onset of the global credit crisis and falsified statistical data placed further upward pressure on its borrowing costs. The EU, ECB and IMF intervened and provided financial assistance to Greece, while the government pledged to economic reforms (Nelson, Belkin & Mix, 2011). However, by June 2015, Greece failed to meet its $1.7 billion payment to the IMF, becoming the first developed nation to formally default (Harrison & Liakos, 2015). In an article by Giakoumis (2014), NPLs were highlighted as one of the most important risks faced by Greek banks. Described as the Achilles heel of banking system, the Greek NPLR remains the highest in the world. With better management of NPLs, Greece could improve its banking sector profitability and economic status. According to the World Bank (2015), the NPLR for Greece was 34.3% in 2014. Only San Marino and Cyprus reported higher NPL rates during that year. The steady upward trend in the Greek NPLR is confirmed in Figure 1.9. Koutras (2015) described the banking system of Greece as a greater challenge than its debt crisis. Although total government debt is excessive, the cash flow maturity profile of this debt favours Greece. While the government debt level is a long term problem, the banking problem is current and urgent. The balance sheets of Greek banks are tainted with NPLs. By the end of 2014, Greek banks reported that 39% of all loans were nonperforming. The high level of NPLs in Greece can be attributed to factors such as poor bank policies in the form of imprudent lending and illegal practices, the severity of the 2008-2009 recession, which eroded 25% of the country's GDP, consumer and business over indebtedness, political party influence around the culture of non-payment and cumbersome legal processes that allowed defaulters to escape or delay payment. By resolving the NPL problem, Greece can define its path for prosperity.

Figure 1.9: NPLs as a percentage of total gross loans for Greece



Source: World Bank, 2015

In the context of SA, the demise of ABIL may be studied for further insight into the role of NPLs as a leading indicator of distress. On 10 August 2014, the Governor of the SARB, Gill Marcus, announced that African Bank will be placed under curatorship. The concerns around the viability of the unsecured credit provider stemmed from its impairment and provision policies, its excessive credit growth and its non-deposit taking business model (SARB, 2014).

The inherent flaw of African Bank was that it failed to hold enough provision against its bad debts (Sanchez, 2014). An equity report by Legae Securities (2010), showed the increase in NPLs for the ABIL Group, observed in Table 1.1. NPLs as a percentage of gross loans increased from 25.4% in 2005 to 35.3% in 2009. The CAGR of total NPLs (54.1%) exceeded the CAGR in gross advances (41.9%) and total impairment provisions (50%). The Group's coverage ratio fell from 70.1% in the 2008 to 61.2% in 2009, which was the lowest in the five year period under review. This relatively lower provisioning level reduced earnings quality for the credit provider. The NPLR for Ellerines in 2009 was 41%, 7% higher than African Bank's.

Table 1.1: Key credit metrics for the ABIL Group

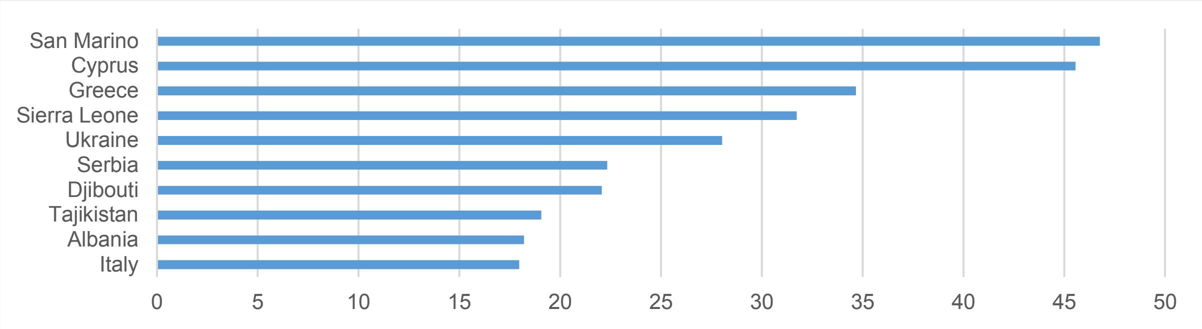
	2005	2006	2007	2008	2009	CAGR
Gross advances	6 454	7 727	10 890	20 908	26 181	41.9%
Growth rate		19.7%	40.9%	92.0%	25.2%	
Total impairment provisions	1 117	1 435	1 892	4 376	5 661	50.0%
Growth rate		28.5%	31.8%	131.3%	29.4%	
Total NPLs	1 642	2 213	3 004	6 239	9 253	54.1%
Growth rate		34.8%	35.7%	107.7%	48.3%	
Impairment/ Gross loans	17.3%	18.6%	17.4%	20.9%	21.6%	
NPLs/ Gross loans	25.4%	28.6%	27.6%	29.8%	35.3%	
Provisions/ NPLs	68.0%	64.8%	63.0%	70.1%	61.2%	

Source: Legae Securities, 2013

1.4. Global and South African NPL rates

The ten countries with the highest NPL rates in 2015 have been shown in Figure 1.10. The highest value was reported by San Marino with a 46.76% NPLR, while Macao reported the lowest NPLR at 0.1205% in 2015.

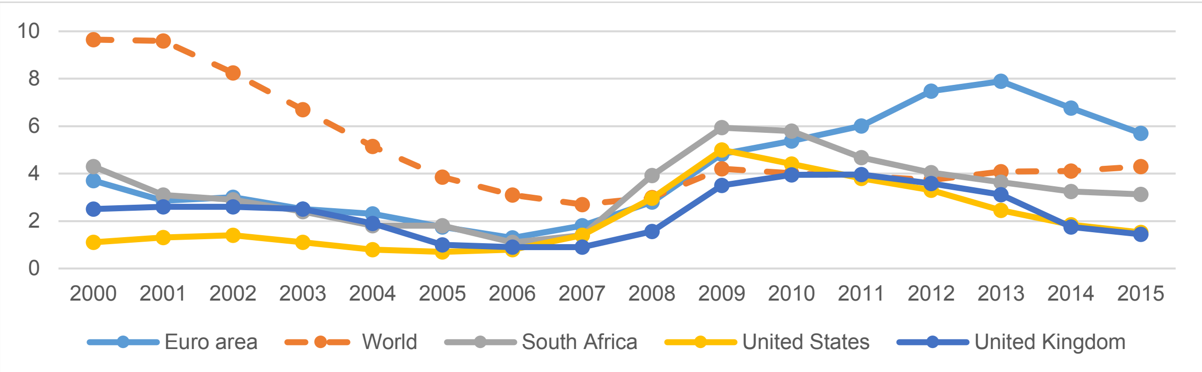
Figure 1.10: Top ten NPL rates for country data from around the world in 2015 (%)



Source: World Bank (2015)

NPLs as a percent of all bank loans for Euro area, UK, USA and SA have been plotted in Figure 1.11. Following the Asian currency crisis in 1997-1998, the NPLR for the globe has steadily declined. During the recession, NPL rates rose but have since recovered. Due to the sovereign debt crisis and Greek default, the NPLR for the Euro area remains elevated. SA has shown an improvement in its NPLR post 2009, reporting a figure of 3.1% in 2015, below the world figure of 4.3%. According to PWC (2015), both macroeconomic and bank specific factors drove drown NPLs in the first half of 2015.

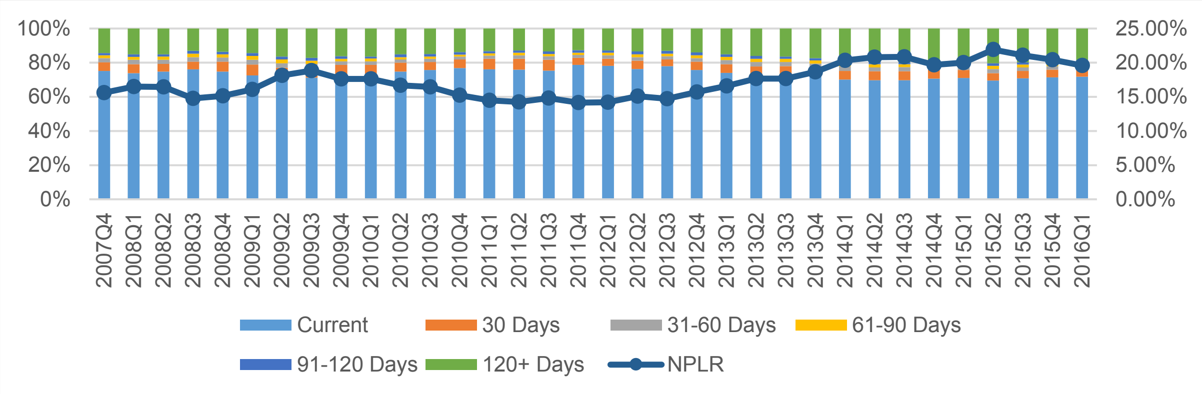
Figure 1.11: Trends in NPLRs for selected country data from around the world (%)



Source: World Bank (2015)

In SA, the NPLR for the unsecured credit consumer market has been trending upward since 2012, consistent with the positive growth experienced in the global industry. Data from the NCR in Q1 of 2016 revealed an NPL to total credit advanced ratio of roughly 20%. The age analysis displayed in Figure 1.12 shows the distribution of unsecured consumer credit balances by the number of days in arrears. NPLs include the 91-120 day and the 120+ day arrears buckets. The primary axis shows the percentage per arrears bucket, while the NPLR is plotted on the secondary axis.

Figure 1.12: Age analysis for the unsecured consumer credit industry in SA weighted in balance



Source: NCR (2016)

This dissertation explores the use of a Fourier residual modification technique in forecasting NPLs. It also seeks to determine if it is beneficial to select an adequate model to fit a data series before modifying its residual terms. The unsecured consumer credit industry has been characterised by positive growth and sound risk management practice is imperative for sustainable growth. There are competitive and regulatory implications in accurately projecting NPLs. A credit provider incurs an opportunity cost if too much capital is held against impairments, while an increased risk of bankruptcy arises with under-provisioning.

The structure of this thesis is organized as follows. A review of literature relating to the reasons for and significance of NPLs is presented in chapter two. Modelling the determinants of NPLs, as well as applications of Fourier residual modification models is included in the second chapter. Contained in chapter three is the methodology and modelling framework adopted in this dissertation and a review of some statistical tools. A descriptive analysis of the data applied in this study is conducted in chapter four, followed by tests for seasonality and stationarity, and the selection of adequate time series models. Chapter five evaluates and compares the performance of residual modified models against a seasonal unmodified time series model. The dissertation is concluded in chapter six and areas for further research are suggested. The list of references and appendices follow chapter six.

Chapter 2

Literature review

This chapter contains three subsections. The first section highlights the causes and relevance of NPLs. Literature related to modelling the determinants of NPLs has been discussed thereafter. Application of the Fourier residual modification approach to forecasting is presented in the final subsection.

2.1. The causes and significance of NPLs

A few reasons for NPLs were listed by Islam *et al.* (2008). The first is that reduced attention to borrowers may increase non-performance on loans. Secondly, lenders will tend to move along the risk curve because the low risk market becomes saturated. This increases the level of risk assumed by the credit provider as well as the degree of unknown risk. Thirdly, risk increases as the loan size increases, implying that NPLs are higher when larger credit amounts are granted. Another reason is that lenders may lack plans to manage risk and borrowers probe at a credit institution's weakness. Poor collections ability drive up NPLs, especially if a credit program is weakly designed, or if there is lack of serious consequence for defaulting customers. The absence of credible risk models leads to higher default as the good credit counterparts cannot be accurately separated from the bad. Viswanadham & Nahid (2015) state that sound credit processes ensure good customer selection and proper risk identification. There must be proactive monitoring of the loans written on book and a transparent recovery strategy for bad debts. A proper policy framework should be formulated to address ethical standards, organizational capacity, prudent lending policies and a credit culture for effectively managing loans. Clementina & Isu (2014) also highlight possible causes of NPLs. Consistent with Viswanadham & Nahid (2015), these include the lack of a sound credit framework and policy, poor credit management, weak undertaking of credit analysis, compromising loan quality for higher profitability, fraudulent practices, erroneous documentation, political turmoil, economic recessions, high competition among lenders, regulatory and policy inconsistencies, and social or political influence on the financial sector. Saba, Kouser, Azeem (2012) also state that flexible credit rationing policies may lead to high NPLs.

Risk-taking is affected by several factors including regulatory practices, moral hazard, ownership structure and agency problems (Espinoza & Prasad, 2010). This is supported by Viswanadham & Nahid (2015), who discuss moral hazard, adverse selection, the principal agent problem and patronizing effect. The principal agent problem suggests that management may fail to align to shareholder interests by taking actions and decisions that do not maximize shareholder wealth. The patronizing effect stems from lenders being unwilling to collect. Unwillingness results from poor internal policies, structures and incentives. Borrowers may be motivated not to repay a loan if there is lack of serious action against non-payment. Underlying adverse selection is that borrowers fail to provide all the required and correct information. Borrowers may possess more accurate private or internal information than lenders. As a result, the lender may have difficulty in assessing and controlling the behaviour of the counterparty due to uncertainty around the risk of default. By setting contractual terms aligned to the average expected quality of credit applicants, the lender tries to be protected against default risk, giving rise to adverse selection. Moral hazard is brought about by borrowers who have private information and take hidden action that increases the risk of default. The incentives of the two parties may change once the transaction is entered into, giving rise to moral hazard. Fofack (2005) argues that moral hazard is especially high when bank capitalisation is low. This leads to the undertaking of imprudent credit strategies that favour high risk projects.

Ekanayake & Azeez (2015) mention that financial system shocks can arise from macroeconomic conditions or bank specific factors. This is consistent with Fofack (2005) who states that dramatic increases in NPLs are largely driven by external shocks and macroeconomic volatility that expose the vulnerability of undiversified countries. Bloem & Gorter (2001) discuss incidents that may influence the amount of NPLs. Abrupt changes in interest rates, foreign exchange rates, prices in key export goods and services and the cost of fuel are a few such events. A decline in the value of underlying loan collaterals may lead to more loans being categorized as doubtful. If the financial system becomes inundated with large amounts of bad loan portfolios, the ability of lenders to provide credit will be hindered leading to a liquidity crunch. Inaba, Kozu & Sekine (2005) suggest that NPLs are caused mainly by a sharp drop in asset prices, especially land. Ahmad & Nor (2015) indicate that political stability and corruption levels impact the amount of impaired finance in a banking sector.

Minimizing NPLs is compulsory for improving economic growth (Messai & Jouini, 2013). This is supported by Ekanayake & Azeez (2015) who state that NPLs hinder economic growth and efficiency. Credit risk depends on asset quality, which is reflected through the level of NPLs. Li & Zou (2014) consider the NPLR as a significant economic indicator. A lower NPLR suggests lower credit risk. The NPL measure represents a probability of loss for the lender. A provision amount is held against this expected loss. The accounting amount of this provision is deducted from profit. Hence, higher NPLs increase the required level of provision and decrease a credit provider's profitability. Mwengei (2013) further solidifies this argument. Higher NPLRs affect the provision for doubtful debts and consequently, the write-off strategies for banks. This reduces profitability, increases the cost of capital and erodes the ROI. Excessive NPLs also leads to a widening of the asset liability mismatch.

Consistent with Li & Zou (2014) and Mwengi (2013), Mesnard, Margerit, Power & Magnus (2016) associate high levels of NPLs with erosion in profitability. High levels of NPLs decrease GDP growth and increase unemployment rates. Economic activity is especially mired when bank financing is heavily relied upon. The impact of high NPLs is channelled into the real economy by reducing a credit provider's ability to lend. Higher NPLs require higher provision amounts, leading to lower operating income for an institution. NPLs attract higher risk weights than performing loans, resulting in higher capital requirements. Due to higher NPLs, investors are less willing to lend their funds leading to increased funding costs and lower profit generating capacity for a credit institution. This is further supported in Japan's annual economy and public finance report. NPLs are problematic as it causes erosion in bank profitability, which leads to a disintermediation of the financial lending system. It creates stagnation of resources, which hinders economic growth and efficiency. The failure to effectively manage NPLs creates a decrease in risk-taking capacity for lending institutions and brings about more cautious behaviour from market participants due to lower trust in the banking system (Government of Japan, 2001). This sentiment is shared by Škarica (2014) who states that NPLs induce uncertainty, as well as a credit provider's willingness and ability to sustain lending, which affects aggregate demand and investment. It traps capital in unproductive resources, which suppresses growth.

The level of NPLs was found to be a leading indicator of recessions. According to Jeon (2010), accumulated NPLs were one of the factors that led to a vulnerable corporate sector and contributed to the outbreak and depth of the Korean crisis in 1997-1998. In an analysis of the Malaysian financial system, Abdullah, Ahmad, Asari, Jusoff, Latif and Muhamad (2011) report a significant relationship between recessions and credit risk. Credit risk had already begun to increase before the 1997 Asian currency crisis. Westernhagen, Harada, Nagata, Vale, Ayuso, Saurina, Daltung, Ziegler, Kent, Reidhill & Peristiani (2004) focused on the role of NPLs in the Japanese banking crisis in the 1990s. During that period, Japanese financial institutions were heavily exposed to the real estate industry. When prices of real estate fell, the amount of NPLs in the economy significantly increased. In Japan, the banking sector had been a dominant credit supplier to the corporate sector, but after sharp decreases in asset prices, investment from the corporate sector declined and the capacity for banks to grant new loans diminished. This economic contraction led to further reduction in the credit quality of bank portfolios.

Mendoza and Terrones (2012) found that lending increased during the expansionary period of a credit upswing, with a simultaneous worsening in asset quality, suggested by higher NPLs. This result suggests that lending standards ease in a credit boom, which attracts more risky customers, leading to deterioration in the quality of a bank's assets. Messai & Jouini (2013) also studied the quality of loans in relation to the macroeconomic environment and phases of the business cycle. During an expansion, there is a small volume of bad loans, as there is sufficient income and revenue to meet debt obligations within the pre-specified time frames. However, in an economic boom, credit may be granted without considering the quality of receivables.

Cheang (2009) discussed an Early Warning System (EWS) to indicate the onset of financial crises. One of the bank specific indicators listed in the model was NPLs as a percentage of total loans, as a measure to identify the asset quality of a credit portfolio. In attempting to construct a Financial Conditions Index (FCI) for South Africa, Gumata, Klein, and Ndou (2012) show that the NPLR is a key indicator of real activity as it sheds insight into the health of the banking sector and the build-up of risks in the financial system. Messai & Jouini (2013) state that the main cause of credit crunch in developed countries was the deterioration in loan quality for bank portfolios. The increase in defaults in the USA's mortgage sector emphasises the link between credit market friction and the likelihood of financial instability. This is supported by Saba *et al.* (2012) who suggest that the financial crisis of the late 2000s was the effect of high NPLs in the USA's banking industry. The increased NPLR was a dominant reason for a decline in bank earnings. Similarly, Curak, Pepur & Poposki (2013) state that the credit crisis highlighted the importance for credit institutions and regulatory authorities to monitor NPL levels as poor loan portfolio quality increases the risk of insolvency, erodes financial performance, and creates fragility in the system.

According to Mileris (2012), there is historical evidence that associates bank failure with poor credit risk management activities. This is consistent with Messai & Jouini (2013) who mention that before bankruptcy, banks generally have a high level of NPLs. This large amount of bad loans often leads to bank failure. Excessive levels of NPLs cause economic stagnation, as each NPL exacerbates the risk of the lender experiencing financial difficulty and becoming unprofitable. Adjei-Mensah (2014) found asset quality to be a statistically significant predictor of insolvency in many studies and insolvents often had excessive amounts of NPLs before failing. From Fofack (2005), NPLs of insolvent institutions contribute a sizeable share to its assets, especially during systemic crises. NPLs rapidly accumulated prior to the 1990s banking crisis in sub-Saharan African countries.

Aman & Miyazaki (2006) found that NPLs are negatively correlated with the valuation of new equity issues by Japanese banks. Japanese commercial banks have endured an increase in NPLs since the early 1990s, reflecting deterioration in bank asset quality, which in turn affects the valuation of new issues. The amount of NPLs was interpreted as the magnitude of information asymmetry between issuing firms and external investors. The reasoning is that the book value of NPLs is determined at the discretion of bank management, whereas the quality of performing loans is more objectively quantified. For external investors, the valuation of NPLs is less certain than the value of up to date loans. Hence, negative market sentiment from information asymmetry is reduced when institutions have a lower amount of NPLs. If the NPLs are further split by bankrupted borrowers versus delayed payers, a higher proportion of borrowers categorized as bankrupted may lead to greater negative market reaction to new issuances.

Podpiera (2006) explored the relationship between banking sector performance and levels of compliance indicated by the Basel Core Principles for Effective Banking Supervision (BCP). After controlling for macroeconomic variables, financial system sophistication and structural factors, it was found that higher compliance with BCP was positively related to banking sector performance, which was measured by NPLs and net interest margin. NPLs reflect the degree to which banks are able to fulfil their basic task of retrieving money from borrowers and a high NPL level frequently suggests a problematic banking sector. The net interest margin is a measure of efficiency as it represents the cost of financial intermediation incurred by customers. A higher degree of compliance with the BCP had a significantly positive impact on asset quality and was associated with lower net interest margins, implying that banking sector performance is better when there is improved compliance with the BCP. This is consistent with Adjei-Mensah (2014) who found that loan quality was enhanced by the improved monitoring and governance brought about by public listing and institutional ownership of Ghanaian banks. In addition, an increase in bank net interest margins was associated with improved loan quality.

There are also regulatory implications for maintaining an acceptable level of NPLs. The NPLR is a measure of credit risk quality and impacts the level of provisions held by a financial institution, which subsequently affects its liquidity requirements as per the Basel III reform. The Basel III accord aims to improve the banking sector's ability to withstand shocks, promote sound risk management and governance, and to increase transparency and disclosures (www.bis.org[1]). According to Cuza & Thu (2012), more regulations have been placed on the financial sector, with focus on transparency and standardized accounting procedures. This ensures that financial institutions have sufficient capital and reserves to meet contractual obligations. However, these excessive regulations could have contributed to the financial crisis. When banks face increased risks, the requirement for capital increases and decreased lending ensues. Basel III provides a stricter framework for risk coverage and capital quality, aimed to raise the ability of banks to face systemic risks. The liquidity framework has been strengthened by introducing a liquidity coverage ratio and a net stable funding ratio.

[1] Basel III: international regulatory framework for banks. Available: http://www.bis.org/bcbs/basel3.htm

Serwa (2013) discusses two limitations of the NPLR. In contrast to Mendoza and Terrones (2012), who suggest that the NPL measure may be used to evaluate and compare the quality of portfolios, Serwa (2013) argues that it is difficult to compare the NPLR between financial systems in different countries at different phases of the business cycle. Furthermore, the NPL measure is affected by the supply of new credit, as new debt is generally originated at better quality than older debt. The NPLR may also vary for reasons unrelated to changing credit risk or economic conditions, such as the rate of originating new credit products, the buying and selling of credit tranches and the maturing of old debt.

2.2. Modelling the determinants of NPLs

Joseph, Edson, Manuere, Clifford & Michael (2012) implemented a qualitative approach to modelling the causes of NPLs for a Zimbabwean commercial bank. The authors followed a case study design and used questionnaires and interviews for data collection. Among the factors contributing to the determinants of NPLs were government policy, inadequate market information, integrity of the borrower and natural disasters. Poor credit policy and monitoring as well as inadequate risk management were also listed. Viswanadham & Nahid (2015) attempted to establish the determinants of NPLs for the National Bank of Commerce in Tanzania by using questionnaires, interviews and documentary evidence. The study examined the impact of GDP, interest rate, economic conditions, concentration of lending activity and the bank's supervision capacity on loans. With the exception of the concentration of lending activity, all other aforementioned factors were found to influence NPLs. Mwengei (2013) used secondary journal and annual report data in an investigation of the factors contributing to NPLs in the Kenyan banking sector for the 2008 to 2012 period. The conclusion of the study was that macroeconomic elements, such as the interest rate level and spread, lead to changes in NPLs. In a study of nonperforming assets in the Indian banking sector, Rajput, Arora & Kaur (2011) used secondary data collected from journals, reports and websites between 2009 and 2010. Nonperforming assets were impacted by ineffective monitoring and poor lending practices. A decline in nonperforming assets is necessary to improve bank profitability and fulfil capital adequacy requirements.

The quality of loans can be affected by a wide range of variables. Using a single approach time series model, Bofondi & Ropele (2011) examine the macroeconomic drivers of loan quality for Italian banks between Q1 of 1990 and Q2 of 2010. The ratio of the previous quarter's flow of new bad loans to performing loans was used as an indicator of loan quality of firms and households. The first category of variables analysed was related to the general state of the economy and price stability. These included the annual growth in GDP, the seasonally adjusted unemployment rate, annual CPI and the growth rate in money supply. Next to be considered was debt servicing costs, as measured by the three month Euribor rate. For households, the debt burden was expressed as the ratio of loans to disposable income and for firms; it was reflected by leverage. Financial and real wealth factors were also reviewed, such as the growth rate in the Italian stock price index, and economic growth outlook, namely, the slope of the yield curve. The analysis found that the debt burden and cost of borrowing variables explained new bad loan rates. Loan quality was significantly affected by general economic conditions, being inversely related to annual GDP growth and positively related to the unemployment rate.

Messai & Jouini (2013) sought to identify the determinants of NPLs for Italian, Greek and Spanish banks between 2004 and 2008. The sample consisted of 85 non-randomly selected large banks with high NPLs post-crisis. The NPLR was regressed against three macroeconomic characteristics and three bank-specific measures. The macroeconomic factors were annual real growth rate in GDP, the current period's real interest rate and the current unemployment rate. The three bank-specific variables included ROA as profitability measure, current loan loss reserves and the current year's growth in loans granted. An improvement in the real economy led to a decrease in NPLs, confirming that an inverse and significant relationship exists between GDP growth and NPLs. When economic growth is positive, there is higher income levels and improved capacity for borrowers to repay obligations and reduce bad debts. A positive relationship between NPLs and the rate of unemployment was evident, because customers fail to meet financial commitments without a steady cash flow stream, resulting in higher NPL levels. When interest rates increase, it is more difficult for borrowers to satisfy debt obligations, leading to a positive association between NPLs and the variable real interest rate. The ROA measure was negatively related to NPLs, as strong profitability provides less incentive for institutions to engage in risky lending. A significant, positive relationship was found between NPLs and loan loss reserves, as the expectation for high capital losses lead to higher provision amounts, which lowers earning volatility and improves solvency in the medium-term. The change in the amount of loans granted did not significantly explain the variation in the NPLR.

Makri *et al.* (2014) examine factors affecting NPLs in the Eurozone's banking sector between 2000 and 2008 using a dynamic panel regression model. The econometric model contained both macro-variables, such as the annual percentage change in GDP, public debt as a percent of GDP, government budget surplus/ deficit as a percent of GDP, unemployment rate and the average annual rate of inflation, as well as micro-variables, such as ROA, ROE, loans to deposits ratio and CAR. The regression also consisted of one period lagged macroeconomic and bank-specific factors. The CAR and ROE measure were found to be negatively correlated with the NPLR, as lower levels of profitability result in higher risk activities. Public debt as a proportion of GDP and the unemployment rate revealed significant, positive relationships with the NPLR. The annual growth rate in GDP had negative and significant correlation with NPLs, as rising income levels and reduced unemployment improves repayment ability. ROA, loans to deposit ratio, inflation and government budget surplus/ deficit as a proportion of GDP showed no significant explanatory power on the NPLR. In contrast, Messai & Jouini (2013) found a negative relationship between NPLs and ROA.

Curak *et al.* (2013) analysed the drivers of NPLs in the South Eastern European banking system using a dynamic panel function on 69 banks in 10 countries between 2003 and 2010. The current period's NPL was the dependent variable in the model and the independent variables included a set of macroeconomic and bank specific variables. Their findings showed that GDP growth explained NPLs as the economy influenced a borrower's repayment ability. Inflation displayed a positive and significant relationship with NPLs, which suggests that increased levels of monetary instability erodes real income and diminishes a debtor's capacity to repay a loan. NPLs were positively correlated with real interest rates, as higher interest rates increase the debt burden, consistent with Messai & Jouini (2013). The size of the bank was negatively associated with NPLs as larger banks were more capable at solving information asymmetry problems relative to smaller sized banks. Larger banks may also be more equipped in credit analysis and monitoring, contrary to the argument by Khemraj &

Pasha (2009). A positive relationship emerged between NPLs and the solvency ratio, while NPLs were negatively affected by ROA. In comparison, Makri *et al.* (2014) found that the NPLR was not significantly explained by the ROA measure.

In exploring the determinants of the NPLR for the ten largest banks in the CESEE region, Klein (2013) applied a dynamic panel regression with a logit transformation of the dependent variable. Higher NPLs were found to be positively related to unemployment, inflation and currency depreciation, confirming the strong link between the banking sector and business cycle. A high equity to assets ratio resulted in lower NPLs, suggesting the moral hazard effect. Lower NPLs were driven by higher profitability, measured by ROE, which suggests that better managed banks had better asset quality, on average. Excessive lending, represented by the loans to assets ratio, led to higher NPLs. In contrast to Curak *et al.* (2013) and Khemraj & Pasha (2009), Klein (2013) found that bank size did not significantly impact NPLs.

Škarica (2014) modelled the NPL determinants for Central and Eastern European countries between Q3 of 2007 and Q3 of 2012 using a fixed effects model. The study found that GDP growth rate negatively influenced the NPLR, but the unemployment rate was positively related to the NPLR, confirming that NPLs rise in recessions and fall in economic expansions. Higher inflation rates adversely impacted the asset quality of banks, resulting in higher NPLs. Due to higher rates of inflation, interest rates may increase, which leads to a higher debt burden for borrowers. The share price index was not a significant determinant of the NPLR as interactions between the financial and macroeconomic sectors are not often pronounced in countries with under-developed or small stock markets relative to their GDP. The quarterly growth in outstanding loans was also not statistically significant in explaining the NPLR. This result may have been impacted by the time period chosen for this study as credit growth was stifled by global liquidity shocks during 2007 to 2012. In contrast to Klein (2013) who found currency depreciation to be positively related to NPLs, Škarica (2014) finds that the annual percentage change in the nominal effective exchange rate is not significant as an explanatory variable, although the region is characterized by a large amount of foreign currency loans, implying that the NPLR is expected to respond strongly to volatility in the exchange rate.

Espinoza & Prasad (2010) support the view that bank-specific factors and macroeconomic environments influence the level of NPLs in the GCC region. The study used a dynamic panel estimation technique on data consisting of 80 banks between 1995 and 2008. A strong inverse relationship emerged between NPLs and real GDP. Another dominant variable contributing to the build-up of NPLs was the interest rate. World trade growth and a dummy variable for the 1997-1998 Asian crisis were not found to be significant predictors of the NPLR. However, the independent variable used as a proxy for global risk aversion and credit tightening was highly significant, implying that external financing conditions is a stronger driver of credit risk than the global trade cycle. Bank characteristics for efficiency and size were significant explanatory variables for NPLs. Expenses relative to average assets measured efficiency and banks with lower expenses had lower NPLs. The logarithm of equity was used to measure bank size and revealed an inverse correlation with NPLs, in contrast to Khemraj & Pasha (2009).

Fofack (2005) investigated the leading causes of NPLs that affected sub-Saharan countries in the 1990s banking crises. The study employed a correlation analysis to examine the relationship between macroeconomic and bank specific variables on NPLs. This was further complemented by a Granger causality analysis. Contrary to Klein (2013), Fofack (2005) found a positive association between NPLs and real exchange rate appreciation. Currency appreciation may restrict growth prospects in export driven industries, leading to reduced profit margins and receding economic output, which adversely impacts loan performance. However, the direction of this relationship did not remain constant throughout the sample period.

In establishing the determinants of NPLs for the Guyanese banking sector, Khemraj & Pasha (2009) apply a fixed effect regression function to model NPLs against macroeconomic and bank-specific characteristics. The panel dataset consisted of macroeconomic factors and bank level data for six institutions between 1994 and 2004. The analysis showed that NPLs were positively associated with the loan to asset ratio, suggesting that greater risk taking behaviour causes higher NPLs, similar to Klein (2013). Real interest rates revealed a positive relationship with NPLs but the correlation coefficient was weak. Growth in loans was negatively related to NPLs, indicating that banks granting higher levels of credit incurred lower NPLs. This is contrary to Škarica (2014) and Messai & Jouini (2013) where loans growth was not found to be significant in influencing change to the NPLR. Khemraj & Pasha (2009) also found that a positive and significant relationship arose between NPLs and the real effective exchange rate, implying that international competitiveness is a key driver of credit risk. A mixed relationship emerges between NPLs and inflation. The current period's NPL was negatively related to inflation but the previous period's NPL was positively correlated with inflation. This implies that high inflation rates experienced in the past year will increase current NPLs. However, the inflation variable was not statistically significant. Makri *et al.* (2014) also found that inflation was not a significant explanatory variable when regressed against the NPLR, while Curak *et al.* (2013) find a significant and positive association between inflation and NPLs.

Ekanayake & Azeez (2015) used a fixed effect panel regression model, similar to Khemraj & Pasha (2009), to explain the NPLR against macroeconomic and bank-specific variables in an investigation of the determinants of credit risk for 9 Sri Lankan commercial banks from 1999 to 2012. The NPLR was found to vary positively with the prime lending rate and negatively with the inflation and the GDP growth rate. High lending rates increase the debt obligation of the borrower, causing more payment defaults. During periods of high inflation, banks change their credit policy to experience lower NPLs. This is in contrast to Curak *et al.* (2013) who found a significantly positive relationship between NPLs and inflation. In terms of unique lender factors, Škarica (2014) failed to observe outstanding loans growth as a statistically significant predictor of the NPLR, while Ekanayake & Azeez (2015) suggest that high credit growth is negatively correlated with NPL levels. The bank size also influenced NPLs, where larger banks incurred a lower number of loan defaults than smaller banks. While this result is consistent with Curak *et al.* (2013), it is in contradiction to Khemraj & Pasha (2009) who suggest that larger banks are less effective at screening loan applicants, relative to smaller lenders, which may lead to higher levels of NPLs.

Prasanna (2014) used a panel dataset of 31 Indian banks with annual data from 2000 to 2012 in a bivariate regression analysis that investigated the determinants of NPLs. The natural logarithm of GDP at factor cost, the growth rate of GDP, the per capital income growth rate, foreign trade proxies and growth rate in savings were found to have a significant inverse relationship on NPLs. The most influential variables in lowering NPLs were GDP and savings growth rates. Stock market and exchange rate volatility also revealed an inverse association with NPLs but were not statistically significant. In contrast, Fofack (2005) and Khemraj & Pasha (2009) find a positive and significant relationship between NPLs and the real effective exchange rate. Prasanna (2014) further reveals that both interest rates and inflation have a significant and positive impact on NPLs; however, Ekanayake & Azeez (2015) find a negative association between inflation and the NPLR.

Ahmad & Bashir (2013) explored the explanatory power of 9 macroeconomic variables in determining NPLs in Pakistan using a panel and cross country regression analysis, dynamic panel model and a VAR method. To remove heteroscedasticity, a logarithmic transformation was applied to the variables in the econometric model. In contrast to Khemraj & Pasha (2009) and Prasanna (2014), Ahmad & Bashir (2013) find that the interest rate is negatively related to NPLs. The justification is that the interest rate represents a cost of borrowing, causing individuals and investors to reduce or defer consumption and investment in risky projects. This leads to higher savings, declined borrowing and lower NPL levels. Furthermore, a significant negative relationship is found between inflation and NPLs, unlike Curak *et al.* (2013) who find a significant and positive association between these variables.

To forecast loan quality in the Thai corporate and consumer banking sectors, Nualsri, Roengpitya, Sabborriboon and Thanavibul (2015) used an ARMA regression analysis on quarterly data between Q4 of 1999 and Q1 of 2014. Their finding was that movements in NPLs and SM loans (30 days in arrears) can be predicted by factors such as GDP, excess liquidity, loan growth and burden of debtors. However, CPI and oil price only appear to affect SM loans and not NPLs. This is in contrast to Curak *et al.* (2013), Prasanna (2014), Ekanayake & Azeez (2015) and Ahmad & Bashir (2013) who find that NPLs are associated with the price variable, inflation.

Abdullah et al. (2011) use a VECM on 48 monthly data points from Malaysian commercial banks to analyse the relationship between NPLs and the interest and inflation rates. The results show that in the short-term, both macro-variables did not impact NPLs. In the long-term, inflation had a significant relationship with NPLs but the interest rate did not. This is in contrast to Abadi, Achsani & Rachmina (2014) who apply a VECM to Indonesian banking data and find that NPLs have a causal relationship with macroeconomic factors, with the interest rate as one of the most dominant variables affecting NPLs.

Badar & Javid (2013) examine the short run and long run dynamics of macroeconomic variables on NPLs for commercial banks in Pakistan from 2002 to 2011. The VECM explores the short run dynamics and confirms a weak relationship between NPLs with the exchange rate and inflation rate. The analysis further reveals a long run relationship between NPLs with interest rates and money supply. This fails to be supported by Abdullah *et al.* (2011) who find no significant relationship between the interest rate and NPLs in the long-term.

Greenidge and Grosvenor (2009) mention that univariate modelling is useful when the data or determinants are of poor quality or not readily available and when the time series displays persistence. Mukoki & Mapfumo (2015) showed that the previous period's NPL measure had a positive impact on the current NPL levels. Consistent with a time series approach, Makri *et al.* (2014) used the one-period lagged NPLR as an explanatory variable and found that it was positively associated with the current period's NPL. This is similar to Curak *et al.* (2013) and Ekanayake & Azeez (2015), who reveal that the lagged NPL was statistically significant in explaining the current NPLs. Also, Klein (2013) states that NPLs have high autocorrelation, where the previous period's NPL influences the current period's NPL, indicating that a shock to the financial system is likely to persist. To analyse the factors influencing NPLs, Shingjergji (2013) used the natural logarithm of NPLs, as a proportion of total loans four lags ago, as an explanatory variable in an OLS regression model and found that it was a statistically significant predictor of the current period's NPLR. From the discussion above, it is evident that a time series approach to modelling is adequate as NPLs exhibit persistence and the influence of explanatory variables on NPLs often presents conflicting results.

2.3. Fourier series and residual modification

Ludlow and Enders (2000) used Fourier coefficients to estimate non-linear ARMA models, in an effort to capture asymmetric adjustments and conditional volatility in time series data. If there is limited information regarding the type of non-linearity, an error of model misspecification could result. However, to forecast in the out-of-sample window, non-linear models are often approximated with a linear representation. Due to the large number of coefficients in the Fourier ARMA model, the out of sample fit was poor. Forecasting performance improved when only the first frequency was used.

Afshar and Fahmi (2012) used the Fourier series to forecast monthly rainfall for Iran. The authors mention that the AR, ARMA, ARIMA and ANNs are useful and efficient modelling tools but with traditional time series, assumptions of linearity and stationarity have to be satisfied. The mathematical model involved representing periodic functions as a series of sines and cosines and provided reasonably good performance.

Meng, Niu & Sun (2011) forecast electric energy consumption to assist with the planning process of power utilities. The complex characteristics of this time series makes direct modelling difficult. Macroeconomic factors, weather conditions, social development, living habits and other factors influence energy consumption trends, which lead to sub-trends in the series. These sub-trends include a long-term rising trend, a periodical wave trend and a stochastic series. The periodical waves could be forecasted with a Fourier series, and the rising trend with a NN. This combined forecasting technique may yield better results than a single time series model or NN method. NNs have however, been used widely for forecasting monthly data. A limitation of the Fourier series is that it may only be used to simulate waves with constant amplitude. However, due to rapid increases in monthly consumption, mainly from developing nations, the amplitude of periodic waves will increase. A DWT can be used to decompose a signal into low frequency coefficients.

Ejiko and Oladebeye (2015) applied a Fourier series forecasting model to predict the sales volume of a manufacturing firm. The model development coupled a straight line equation with a Fourier series of the cosine odd function, to accommodate for a sinusoidal trend in the series. The developed model was tested and validated with sales data and resulted in accurate predictions. Forecasts generated from the model had high correlation with the actual data series, confirming reliability and dependability. It was proposed that this model be used for products with seasonal fluctuations.

Fumi, Pepe, Scarabotti & Schiraldi (2013) use the Fourier series to forecast demand in the fashion industry. Due to the seasonal nature of product demand and wide range of factors affecting the fashion industry, such as the large variety, short product lifecycles, advertising and marketing campaigns, weather conditions and promotional periods, the Fourier algorithm to forecasting was opted for. Bell, Herbert and Lewis (2002) apply a Fourier analysis to forecasting the incoming call demands of a call centre. These forecasts help to schedule the staff and resources required to satisfy inbound call levels. The call arrival process was identified as a time series with trends and seasonal and cyclic patterns. The FFT was used to fit the periodic data and was shown to be effective.

Omekara, Ekpenyong & Ekerete (2013) modelled monthly Nigerian inflation rates from 2003 to 2011 using a periodogram and Fourier series. The key objective was to identify inflation cycles, fit an adequate model and derive future values. In the time domain, the ARIMA, VAR and error correction models have traditionally been used to forecast economic growth. The authors opt for a periodogram and a Fourier series analysis as it is an easy way to model seasonality and eliminate peaks, overcoming the need for model re-estimation. Periodogram analysis helps in the identification of periods and cycles in a series. One of the reasons why time series models perform well relative to theoretical econometric models is that it does not impose improper restrictions or specifications on variables, allowing the model flexibility to capture the dynamic properties of the data.

From the discussion in chapter one, it was evident that the consumer credit market exhibits seasonal trends. Transunion (2015) confirmed a seasonal spike in distressed borrowing in Q1. The Federal Reserve Bank of Minneapolis (2015) mention that balances on credit cards usually follow a seasonal trend. This was reiterated by Williams (2014), who describes a seasonal peak in revolving credit usage during the festive period. This suggests that a Fourier series has merit in modelling consumer credit related metrics.

Dong, Wang, Wang and Zhao (2012) applied a residual modification to a SARIMA$(1,1,1)(1,1,1)_{12}$ model to forecast electricity demand in China. To improve the accuracy of the SARIMA model, three residual modification techniques were proposed. The results suggest that residual modification models improve the precision of the SARIMA forecasts. Although the SARIMA model already has high precision in fitting data with periodic trends, the analysis revealed that prediction may be enhanced by reducing the error through a Particle Swarm Optimization (PSO) optimal Fourier residual modification approach.

Chen, Hsu, Lai, Nguyen and Shu (2013) forecasted air cargo volume in Taiwan with a Fourier residual modified SARIMA model. The volume of imported air cargo in Taiwan could be forecasted with two SARIMA models, a SARIMA$(1,1,1)(1,1,1)_{12}$ and a SARIMA$(3,1,1)(1,1,1)_{12}$ model. The SARIMA$(3,1,1)(1,1,1)_{12}$ model was chosen due to the lower MAPE measure and a Fourier residual series was then calculated. Six models could potentially be used for forecasting the volume of exported air cargo, including SARIMA$(1,1,1)(1,1,1)_{12}$, SARIMA$(2,1,1)(1,1,1)_{12}$, SARIMA$(3,1,1)(1,1,1)_{12}$, SARIMA$(1,1,3)(1,1,1)_{12}$, SARIMA$(2,1,3)(1,1,1)_{12}$, SARIMA$(3,1,3)(1,1,1)_{12}$ models. The SARIMA$(2,1,3)(1,1,1)_{12}$ and SARIMA$(3,1,3)(1,1,1)_{12}$ models had the lowest MAE and MAPE measures and the Fourier algorithm was used to modify the residual series of both these models. When the traditional SARIMA models were combined with the Fourier residual modified series, the resulting accuracy was higher than the conventional SARIMA models.

Hsu, Hung, Lu, Nguyen & Shu (2014) forecast monthly inbound tourism demand in New Zealand using a Fourier residual modified ARIMA model mainly due to the lack of identifying the key determinants of tourism demand. Based on the ACF and PACF graphs, three possible models could fit the data including SARIMA$(1,0,1)(1,1,1)_{12}$ SARIMA$(1,0,2)(1,1,1)_{12}$ and SARIMA$(1,0,3)(1,1,1)_{12}$ models. The residual series of these three models were modified with a Fourier algorithm and the monthly inbound tourism demand was best forecasted with the modified SARIMA$(1,0,1)(1,1,1)_{12}$. The study concluded that when the SARIMA model was joined with a certain degree of Fourier modification factors, the model performance was significantly better than the conventional SARIMA model. Tsaur & Kuo (2013) propose a Fourier method to revise the residual terms of a time series model in order to enhance forecasting performance, termed a fuzzy model. The fuzzy model was used to forecast the demand of Japanese tourists visiting Taiwan every year. The Fourier series was used to transform the residual terms into frequency spectra, to filter the high frequency terms associated with noise, and select the low frequency terms. The fuzzy model revealed better results, showing smaller forecasting error in the out-of-sample period. Ludlow and Enders (2000) also found that out-of-sample performance improved when only the first frequency was used. Hsu, Huang, Nguyen & Shu (2013) use residual modified models to predict inbound tourism in Vietnam. The conventional models investigated in the study included the ARIMA and Grey GM(*1,1*) models. Among the possible models for selection were SARIMA$(1,1,1)(1,0,2)_{12}$ and SARIMA$(2,1,1)(1,0,2)_{12}$ models. Due to lower MAE and MAPE measures, the SARIMA$(2,1,1)(1,0,2)_{12}$ model's residual series was chosen for the application of Fourier modification. The Grey and SARIMA models were modified with the Fourier series and both models performed well. However, the Fourier modified SARIMA model was better than the Fourier modified Grey model.

2.4. Concluding remarks

Previous studies have predominantly focused on analysing the determinants of NPLs and less emphasis has been placed on exploring different forecasting tools. Both qualitative and quantitative techniques have been used to analyse NPLs, with quantitative preference for dynamic panel regression, VAR and error correction models. Literature has indicated that macroeconomic and bank-specific factors influence NPLs, but several variables were shown to have contrasting impact on NPL levels. The tendency for the previous period's NPLR to influence the current period's NPLR, together with conflicting explanatory factors favours the use of time series models. Evidence from different industries support the application of Fourier residual modification, as forecasting accuracy has seen to improve. To the author's knowledge, there is no prior research on NPL forecasting with a Fourier residual modification model within the context of South African credit providers.

Chapter 3

Statistical tools and Methodology

This chapter provides an introduction to basic concepts in descriptive statistics, time series and frequency domain analysis. The theoretical framework presented here will provide the foundation for the analysis component of this dissertation. The literature review demonstrated the preference to depict NPLs with a time series process, and a Fourier residual modification technique was shown to improve forecasting ability in the energy, tourism and air cargo industries. Seasonal trends characterise the consumer credit market and as such, this chapter will review tests for seasonality and stationarity in addition to time series models, contained in sections two and three. A few basic statistical concepts are introduced in the first section and the fourth describes the Fourier series. Metrics for evaluating model performance are presented in the fifth section and the chapter concludes with the methodology in section six.

3.1. Basic statistical concepts

Definitions 1, 2 and 3 follow from Stolojescu (2011).

Definition 1: Autocovariance and Autocorrelation

The autocovariance of a time series X_t between times *k* and *l*, where *l>k*, is defined as: $\gamma(k,l) = E\{[X_k - \mu][X_l - \mu]\}$. The autocovariance function determines how X_t is related to its previous values in a time series. The autocovariance value becomes more interpretable when it is divided by the variance.

The autocorrelation function (ACF) for a process X_t between times *k* and *l*, where *l>k* is defined as: $\rho(k,l)= \frac{\gamma(k,l)}{\sigma_k \sigma_l} = \frac{E\{[X_k - \mu][X_l - \mu]\}}{\sigma_k \sigma_l}$, where *E* is a statistical expectations operator.

When X_t is a stationary process, the ACF simplifies to: $\rho(k,l)= \frac{\gamma(k,l)}{\sigma^2}$

Definition 2: Partial autocorrelation function (PACF)

The partial autocorrelation at lag *k* is defined as the autocorrelation between X_t and X_{t-k} that is not explained by lower order lags (*1* to *k-1*, inclusive).

$$\emptyset_{kk}=CORR(X_t-P(X_t|X_{t+1}+\ldots+X_{t+k-1}),\ X_{t+k}-P(X_{t+k}|X_{t+1}+\ldots+X_{t+k-1}))$$

Where $P(X_t|X_{t+1}+\ldots+X_{t+k-1})$ is the best linear forecast of X_t over $X_{t+1}+\ldots+X_{t+k-1}$ and *CORR* refers to correlation.

Definition 3: Stationarity

A stochastic process X_t is stationary if it is characterized by a probability distribution with a mean and variance that does not change over time and position. A White noise Gaussian process is an example of a stationary process, as its realizations at every point in a range of time comprise of random variables normally distributed as:

$$P_G(x)=\frac{1}{\sigma_G\sqrt{2\pi}}\ e^{\frac{(x-\mu_G)^2}{2\sigma_G^2}}$$

Brooks (2008) defines a strictly stationary process as one where:

$Fx_{t_1},x_{t_2},\ \ldots,\ x_{t_T}\,(x_1,\ \ldots,x_T)=Fx_{t_{1+j}},x_{t_{2+j}},\ \ldots,\ x_{t_{T+j}}\,(x_1,\ \ldots,x_T)$ *for* $t_1,t_2\ ,\ \ldots,t_T\in Z;\ j\in Z;$

$T=1,2,\ \ldots;$ F is the joint distribution of random variables; Z is a set of natural numbers

As time progresses, the distribution of values in a strictly stationary series remain the same. A weakly stationary process has a constant mean, constant variance and constant autocovariance. A model with non-stationary coefficients will be burdened by an error term whose past values have a non-declining effect on the current values of the time series. In order to prevent spurious results, a series is required to be stationary. A spurious result occurs when a non-stationary variable is regressed on another non-stationary variable yielding a high goodness of fit metric, creating a misleading conclusion of model adequacy.

Definition 4: Maximum likelihood estimation

For a sample $X=(X_1,\ X_2,\ \ldots,\ X_n)$ of random variables chosen with probabilities P_θ, and density function denoted by $f(x/\theta)$, $x=(x_1,\ x_2,\ldots,\ x_n)$ when θ is the true state of nature for the data, the maximum likelihood estimator $\hat{\theta}$ for the parameter θ, yields a value that makes the observed data most probable.

The likelihood function is the density function with respect to a function of θ, expressed as follows: $L(\theta/x)=f(x/\theta),\ \theta\in\ominus$

The maximum likelihood estimator may be expressed as: $\hat{\theta}(x)=argmax_{\theta}L(\theta|x)$ where *argmax* refers to arguments of the maxima, a function in which the outputs are as large as possible for a given set of domain inputs (Watkins, 2011).

3.2. Time series processes

The construction of a time series model is generally a-theoretical. Time series analysis attempts to describe the empirically relevant characteristics of the observed data. Univariate time series models refer to a class of specifications in which the forecasts of a variable are produced using only previous observations of the variable itself and potentially past and current values of the residual term (Brooks, 2008). Definitions 5 to 10 follow Brooks (2008).

From Gerbing (2016), the four underlying components of time series are:

i. Trend: represents the long term growth of the series.
ii. Cyclical variation: describes the alternating cycles of expansion or recession, varying in length and magnitude.
iii. Seasonal variation: reflects a periodical pattern, which repeats itself.
iv. Irregular variation: refers to a random or erratic movement in the series.

<u>Definition 5</u>: Autoregressive process

An autoregressive process is where the current value of a series z_t can be explained by its previous values plus some error term. An autoregressive model of order *p,* AR*(p),* is represented as:

$z_t = \mu + \alpha_1 z_{t-1} + \alpha_2 z_{t-2} + \ldots + \alpha_p z_{t-p} + e_t = \mu + \sum_{k=1}^{p} \alpha_j \; z_{t-k} + e_t$

In the context of the analysis conducted in this analysis, the autoregressive process consists of forecasting the current NPLR with previous values of the series plus an error term, as follows: $NPLR_t = \mu + \alpha_1 NPLR_{t-1} + \alpha_2 NPLR_{t-2} + \ldots + \alpha_p NPLR_{t-p} + e_t$, where the parameter p is chosen by examining the PACF of the NPL series.

<u>Definition 6</u>: White noise process

A white noise process does not have a discernible structure and it is a process where the mean and variance are constant. The autocovariance for a white noise process is different from zero only at lag zero as no observation is correlated with any other observation in the series. This implies that the ACF of a white noise process has a single peak of 1 at lag 0. Let n_t be a white noise process, then:

$E[n_t] = \mu$

$Var[n_t] = \sigma^2$

$$\gamma_{t-i} = \begin{cases} \sigma^2 \text{ for } t=i \\ 0 \text{ otherwise} \end{cases}$$

<u>*Definition 7*</u>: Moving average process

Let w_t $(t = 1, 2, 3 \ldots)$ be a white noise process. A moving average process of order q, MA(q), is represented as:

$z_t = \mu + w_t + \varphi_1 w_{t-1} + \varphi_2 w_{t-2} + \ldots + \varphi_p w_{t-q} = \mu + \sum_{k=1}^{q} \varphi_j \; w_{t-k} + w_t$

Let the backshift operator be $B^k w_t = w_{t-k}$ then $z_t - \mu = \sum_{k=1}^{q} \varphi_k B^k$

In the context of this study, a moving average process will consist of forecasting the current NPL rate with a white noise process, where q is defined by the ACF.

<u>*Definition 8*</u>: Autoregressive Moving Average (ARMA) process

ARMA processes combine the autoregressive, AR(p) and moving average MA(q) models. In an ARMA(p,q) process, the values of a time series depend on previous observations of the series, up to p lags, together with previous values of a white noise residual term, up to q lags ago. The model is expressed as follows:

$$z_t = \mu + \alpha_1 z_{t-1} + \alpha_2 z_{t-2} + \ldots + \alpha_p z_{t-p} + w_t + \varphi_1 w_{t-1} + \varphi_2 w_{t-2} + \ldots + \varphi_p w_{t-q} + e_t$$

$$\text{where } E[e_t] = \mu;\ E\left[e_t^2\right] = \sigma^2;\ E[e_t e_s] = 0,\ t \neq s$$

<u>*Definition 9*</u>: Autoregressive Integrated Moving Average (ARIMA) process

An ARIMA(p,d,q) model is ARMA(p,q) process where the integrated AR(p) process has a characteristic equation with a unit root. The original data series is differenced d times to induce stationarity.

<u>*Definition 10*</u>: Seasonal Autoregressive Integrated Moving Average (SARIMA)

The SARIMA$(p,d,q)(P,D,Q)_s$ model is a combination of two models generated by the ARIMA(p,d,q) model and ARIMA(P,D,Q) model. The SARIMA$(p,d,q)(P,D,Q)_s$ model is denoted as:

$$\alpha_p(B)\theta_P(B^s)(1\text{-}B)^d(1\text{-}B^s)^D z_t = \varphi_q(B)\vartheta_Q(B^s)z_t$$

Where B is the backshift operator and $(1\text{-}B)^d z_t$ becomes a stationary series using the differencing operator. The equations denoting these components are:

$$\alpha_p(B)=1\text{-}\ \alpha_1 B\text{-}\ \alpha_2 B^2\text{-}\ \alpha_3 B^3\text{-}\ \ldots\text{-}\ \alpha_p B^p$$

$$\varphi_q(B)=1\text{-}\ \varphi_1 B\text{-}\ \varphi_2 B^2\text{-}\ \varphi_3 B^3\text{-}\ \ldots\text{-}\ \varphi_q B^q$$

$$\theta_P(B)=1\text{-}\ \theta_1 B^s\text{-}\ \theta_2 B^{2s}\text{-}\ \alpha_3 B^{3s}\text{-}\ \ldots\text{-}\ \theta_P B^{Ps}$$

$$\vartheta_Q(B^s)=1\text{-}\ \vartheta_1 B^s\text{-}\ \vartheta_2 B^{2s}\text{-}\ \vartheta_3 B^{3s}\text{-}\ \ldots\text{-}\ \vartheta_Q B^{Qs}$$

Where p is the order of the AR component; q is the order of the MA component and d is the order of differencing required to make the series stationary. The seasonal AR component has order P, the seasonal MA component has order Q and seasonal differencing is of order D.

3.2.1. Stationarity tests

To assess if a series is stationary, the Phillips-Perron unit root tests may be conducted. The null hypothesis for this test states that a unit root exists in the data. If the null hypothesis fails to be rejected, the series is considered non-stationary. The alternate hypothesis suggests that a unit root does not exist in the data and therefore, when the null hypothesis is rejected, there is evidence to suggest a stationary series. For the condition of stationarity to be statistically met, the null hypothesis should be rejected at a significance level of 5% (or with 95% confidence in the result). Equivalently, a probability value (p-value) less than 0.05, leads to rejection of the null hypothesis and conclusion that the time series is stationary.

3.2.2. Seasonality tests

From Levenbach (2015), seasonality refers to periodic fluctuations that repeat itself with the same intensity and timing. This may refer to monthly, yearly, quarterly, weekly or daily repetitions. Using a seasonal decomposition procedure, the average or typical seasonal patterns in the data may be measured. The additive decomposition is adequate if the magnitude of the seasonal movement is constant and is not reliant on the level of the series. The time series data may then be expressed as: Data = Trend

+ Seasonal index + Irregular component, where the irregular component captures the unexplained variations plus the random error term. When the magnitude of the seasonal movement increases or decreases with the level of the time series, the multiplicative decomposition is appropriate, and is expressed as: Data = Trend * Seasonal factor * Irregular component.

The F-test for the presence of stable seasonality is the quotient of two variances. The numerator represents the variance between months, which is primarily influenced by the magnitude of the seasonal component. The denominator represents the residual variance, which is driven mainly by the irregular component of the series. The null hypothesis of the F-test is that no significant seasonality exists in the data. Because many of the assumptions underlying the F-test are likely to be violated, the significance level at which the F-test is conducted is 0.001. A high F-value for this test is a strong indication for the presence of measurable seasonality. The null hypothesis is rejected when the F-value is greater than its critical F-statistic. In the critical F-statistic, the numerator has degrees of freedom that reflects the between month seasonality and the denominator has degrees of freedom associated with the residual term.

The F-test for moving or YoY seasonality, measures whether the time series portrays gradual changes in its seasonal amplitude but not in its phase. The moving seasonality F-test is the ratio of the between years variance and the residual variance. The numerator captures the YoY movement in seasonality and if this moving seasonality component is too large, it introduces distortion to a model. The denominator is the residual variance, which is what remains after the variance between months and the variance between years are accounted for. The null hypothesis suggests that no moving seasonality is present in the data. A high F-value or low p-value will support the rejection of the null hypothesis, to indicate that moving seasonality exists in the data. In the critical F-statistic for this test, the numerator has degrees of freedom that reflects the between YoY seasonality of the data and the denominator has degrees of freedom corresponding to the residual term. If the null hypothesis is rejected, moving seasonality is present in the data, and the probability of reliably estimating the seasonal factors decreases.

A combined test for the existence of identifiable seasonality combines the moving seasonality F-test and the Kruskal-Wallis chi-squared test for stable seasonality. This test determines if the seasonality in the data is identifiable or not. If the moving seasonal component dominates the stable seasonal component in the process, the chances of accurately estimating the seasonal component is reduced because it will not be properly identified.

The steps for performing a combined test to check for the presence of identifiable seasonality are as follows:

1. Perform the test for stable seasonality. If the p-value is greater than 0.001, the null hypothesis of no significant seasonality is not rejected. This suggests that the series is not seasonal and identifiable seasonality does not exist. If the p-value is less than 0.001, there is evidence of stable seasonality in the data and a test for moving seasonality is then conducted.
2. When performing the moving seasonality test, the significance level returns to 0.05. If the p-value of this test is more than 0.05, there is no evidence of YoY seasonality in the data. If the p-value is less than 0.05, the null hypothesis of moving seasonality is rejected.
3. To establish if identifiable seasonality is present in the series, the following steps are considered:
   - The F values for the stable seasonality and moving seasonality tests, denoted as F_{stb} and F_{mvg} respectively.
   - Compute $X_1 = \frac{7}{F_{stb}}$
   - Compute $X_2 = \frac{3 * F_{mvg}}{F_{stb}}$
   - Calculate the average of these measures: $X = \frac{X_1 + X_2}{2}$

   If the null hypothesis of moving seasonality is rejected from step 2 (p-value less than 0.05) and the value of X is greater than or equal to 1, then the null hypothesis of no identifiable seasonality will fail to be rejected. The conclusion will be that there is no evidence to suggest that identifiable seasonality exists in the series. If either the null hypothesis is rejected or the null hypothesis is not rejected from step 2, but the value of X is less than 1, then the measures for X_1 and X_2 are considered. If the values of X_1 or X_2 are greater than or equal to 1, then the conclusion is that identifiable seasonality is probably not present in the series. If the values for X_1 and X_2 are both strictly less than 1, then the Kruskal-Wallis chi-squared test is performed. If the p-value for the Kruskal-Wallis test is greater than or equal to 0.001, identifiable seasonality is not likely to exist in the data. If the p-value is less than 0.001, the conclusion is that there is statistical evidence to suggest the presence of identifiable seasonality in the series (www.sas.com[2]). A flowchart for this combined seasonality test is contained in Appendix A, Figure A1.

3.3. Phases of a time series model build

The next three subsections follow from Wei (2005). When building a time series model, the phases include:

i. Model identification
ii. Parameter estimation
iii. Diagnostic checking
iv. Model selection

3.3.1. Model identification

The steps for model identification include:

1. Plotting the data and applying the necessary transformations.
   By plotting the data, depiction of trends, seasonality, outliers, variances and non-stationarity in the series can be visually depicted. If the condition of stationarity is not satisfied, the series is differenced. For series with a non-constant variance, a logarithmic transformation is frequently applied.
2. Examining the sample ACF and PACF functions of the original series to establish the degree of differencing required.
   If the sample ACF gradually decays and the sample PACF cuts off after the first lag, it indicates that differencing is required. A unit root test may also be used to identify differencing. Non-stationarity is removed by higher order differencing, but d is commonly 0, 1 or 2 for $(1-B)^d z_t$, $d>0$ where B is the backshift operator.

[2] www.sas.com. Combined Test for the Presence of Identifiable Seasonality. Available: http://support.sas.com/documentation/cdl/en/etsug/60372/HTML/default/viewer.htm#etsug_x12_sect027.htm

3. Examining the sample ACF and PACF of the differenced series and identifying the orders of *p* and *q*, where *p* is the highest order of the AR process and *q* is the highest order of the MA model. The characteristics to theoretically identify the ACF and PACF are given in Table 3.1 and will be used as an identification tool for estimating the parameters *p* and *q*.

Table 3.1: Characteristics of the ACF and PACF for a stationary process

Process	**ACF**	**PACF**
AR(*p*)	Gradually decays exponentially or in a sine wave pattern	Ends after lag *p*
MA(*p*)	Ends after lag *q*	Gradually decays exponentially or in a sine wave pattern
ARMA(*p,q*)	Gradually decays after lag (*q*-*p*)	Gradually decays after lag (*p*-*q*)

Source: Wei (2005)

4. Examining the Inverse Sample Autocorrelation Function (ISACF)
   In an ARMA(*p,q*) model where the AR stationary component is represented as follows: $\alpha_p(B)=(1-\alpha_1 B-\ldots-\alpha_p B^p)$ and the invertible moving average component is: $\varphi_p(B)=(1-\varphi_1 B-\ldots-\varphi_q B^q)$. The ISACF may be used as a model identification tool. The inverse ARMA(*p,q*) process is an ARMA(*q,p*). The inverse process of an AR(*p*) process with ACFs slowly decaying is an MA(*p*) process with ACFs cutting off at lag *p*. Likewise, an MA*(q)* process with ACFs cutting off at lag *q* will have inverse autocorrelations that gradually taper off.
5. Examining the Extended Sample Autocorrelation Function (ESACF)
   In a mixed ARMA model, the ACF, PACF and ISACF may display gradual decays, making the identification of parameters *p* and *q* more difficult. The ESACF is a useful tool in suggesting the orders for *p* and *q* and is derived for an iterated set of regressions. OLS estimation is used in this iteration and may be applied to non-stationary or non-invertible series. When the ESACF is generated from non-differenced data, an ARIMA(*p,d,q*) model is reflected as an ARMA(*p+d,q*) process. When used on a properly transformed stationary data series, the ESACF is a good tool in tentative parameter selection.

3.3.2. Parameter estimation

Once the tentative model has been identified, parameters may be estimated using maximum likelihood, least squares or non-linear estimation. The order of parameters to the AR, namely p and/ P and MA, namely q and/ Q components are established at this stage after testing for statistical significance.

3.3.3. Diagnostic checking

Time series modelling is an iterative procedure involving model identification, parameter estimation and model adequacy. Diagnostic checking establishes whether model assumptions are satisfied by analysing the residual terms. This involves examining the distribution, variance and sample ACF and PACF plots of the errors.

The DW test checks whether the residual or error terms of a series are independent. The residuals can display positive or negative autocorrelation. The null hypothesis of the DW test states that there is no evidence of statistically significant autocorrelation in the residual terms. Failure to reject the null hypothesis will support the conclusion that the error terms are independent. If serial correlation exists in the residual series, the underlying assumption of classical linear regression modelling will be violated, that is: $E[e_t e_s] = 0$ for $t \neq s$ will not hold. This violation makes hypothesis testing and confidence level construction less reliable. The DW test assumes that the model's residual terms are generated by an AR(*1*) process of equally spaced intervals. The diagnostic checks conducted will also test the residual series for constant variance and stationarity.

3.3.4. Model selection criteria

From Brooks (2008), the selection criteria for model comparison include the Akaike's Information Criterion (AIC) and Schwartz's Bayesian Information Criterion (SBIC). The information criteria are functions of the residual sum of squares and the number of parameters in the model. An additional parameter in the model has contrasting effects, it may decrease the residual sum of squares but also increase the penalty term which captures the loss of degrees of freedom by adding the parameter to the model.

Algebraically, these information criteria are denoted as:

$$\text{AIC}= ln(\widehat{\sigma}) + \frac{2n}{T}$$

$$\text{SBIC}= ln(\widehat{\sigma}) + \frac{n}{T}\, ln(T)$$

Where $\widehat{\sigma}$ is the residual variance, equal to the residual sum of squares divided by the total number of data points in the sample, T. The number of parameters in the model $n= p+q+1$. The best model is one where the information criteria measures are lowest.

3.4. Frequency domain analysis

From Wei (2005), the time series approach to modelling studies the autocorrelation and partial autocorrelation functions for constructing a model. When the behaviour of a time series is studied in terms of sinusoidal patterns at different frequencies, it involves frequency domain analysis. Frequency domain analysis involves the representation of signals by its frequency, harmonic components, phase, amplitude and spectrum.

The objective of a Fourier modified residual model is to improve the forecasting ability of conventional time series models. The error terms may be interpreted as a series of sine and cosine waves and the Fourier algorithm helps to identify the most important frequencies. A number of studies have carried out a Fourier residual modification across different industries, as noted in the literature review chapter.

3.4.1. Fourier series

A Fourier series is used to transform signals between two domains and it provides a way to represent how much information is contained at various frequencies for a signal. The Fourier series is a periodic function expansion of an infinite sum of sine and cosine waves. Harmonic analysis is the study of Fourier series (Weisstein, 2016). Definition 11 follows from Weisstein (2016).

Definition 11: Fourier series representation

If f is a piecewise continuous function over $[-\pi,\pi]$, the Fourier series representation of f is: $a_0+\sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx)$

The coefficients are $a_{0,}$ a_n and b_n are calculated as follows:

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)dx$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\, dx$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\ dx$$

3.5. Performance measures

The Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are selected to compare model performance. In order to choose the best model, the value of MAE is minimized (Dong *et al.,* 2012).

The MAE and RMSE measures are summarized below.

$$\mathrm{MAE} = \frac{1}{n}\sum_{k=1}^{n}\left|\frac{\widetilde{z}_k - z_k}{z_k}\right|$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{k=1}^{n}(\widetilde{z}_k - z_k)^2}{n}}$$

Where $\widetilde{z}_k$ refers to the forecasted value and z_k denotes the actual value.

3.6. Methodology

This analysis follows a Box-Jenkins approach to time series modelling, presented in Appendix A, Figure A2. Initial tests for seasonality and stationarity are performed before model identification and parameter estimation. Diagnostic checks will be conducted on the error terms before selecting an adequate model to apply a residual modification technique.

To fit an adequate model to the original series, the following steps are proposed:

1. Conduct tests for stationarity. First differencing and/ or seasonal differencing of the series may be required to induce stationarity.
2. Identify the potential model parameters using the ACF, PACF, ISACF and ESACF plots and estimate the orders of the AR and MA models, namely p, q and/ or P, Q.

3. Conduct diagnostic checks on the residual terms.
4. Evaluate and compare model adequacy.
5. Select the model/s with the best fit for the residual modification stage.

For the residual modification stage, the residual terms from the models with the best fit are chosen. The steps to residual modification follow:

1. Plot the residual series of the original time series model/s. The residual or error term represents the difference between the forecasted and actual value of the original time series.
2. Ensure that the residual series is stationary.
3. Predict the residual series using a time series model. Combine the time series forecasts for the error terms with the forecasts from the original model.
4. Forecast the residual series using a Fourier algorithm. Combine the error forecasts from the Fourier series with the forecasts from the original model.
5. Compare the performance of the unmodified model with the time series residual modified model and Fourier series residual modified model.

Chapter 4

Data and Pre-Analysis

This chapter contains the data and pre-analysis of the paper. The pre-analysis follows a Box-Jenkins modelling approach and applies initial data tests identified in chapter three. This chapter is divided into six subsections. The first section identifies the source and the second explores the descriptive content of the data. Seasonality and stationarity tests are performed in the third section. Contained in the fourth section is a discussion of the estimation of parameters for a univariate time series model. Section five contains the diagnostic checks for the estimated model. A brief comparison of the models selected to fit the original data concludes this chapter in section six. The results produced in this chapter will be applied in the next chapter, where residual modification is explored.

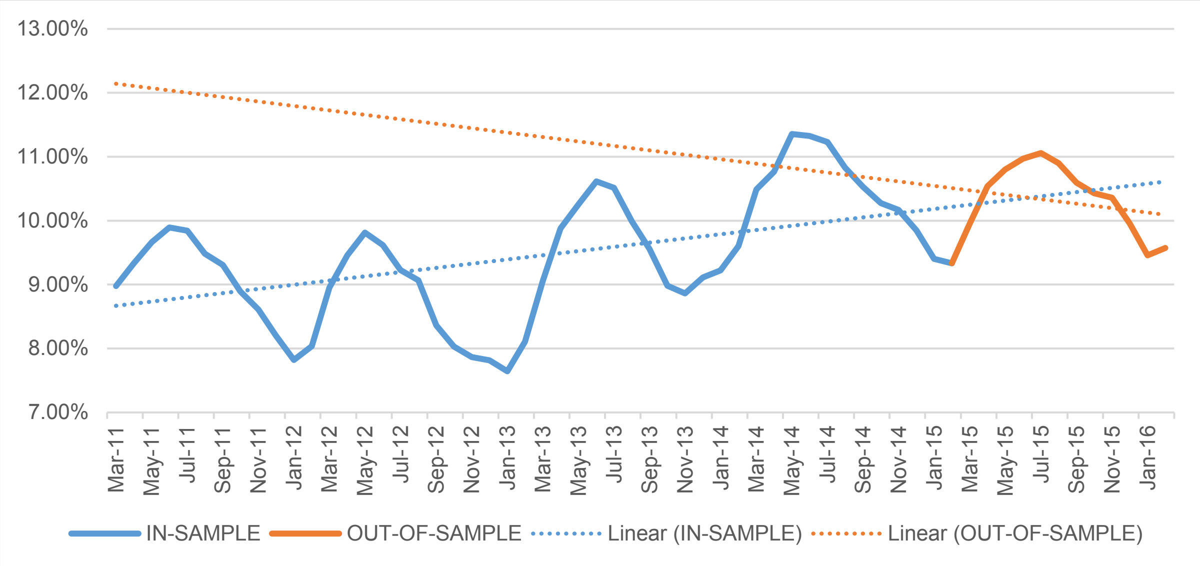
4.1. Data

The market data used in this research project was sourced from an unsecured consumer credit provider and contains sixty monthly observations, from Mar'11 to Feb'16. An in-sample data set was created with the first 48 data points. The last 12 observations were retained as an out-of-sample data set, which serves to validate the results generated from the in-sample period. The out-of-sample period is the forecasting horizon. The NPL measure conforms to the Basel definition of greater than ninety consecutive days in arrears on a credit agreement. The NPLR is balance weighted, where the numerator captures the nonperforming balance and the denominator reflects the balance of total credit advanced (including performing and nonperforming balances). SAS is the statistical software tool used in this analysis.

4.2. Descriptive analysis

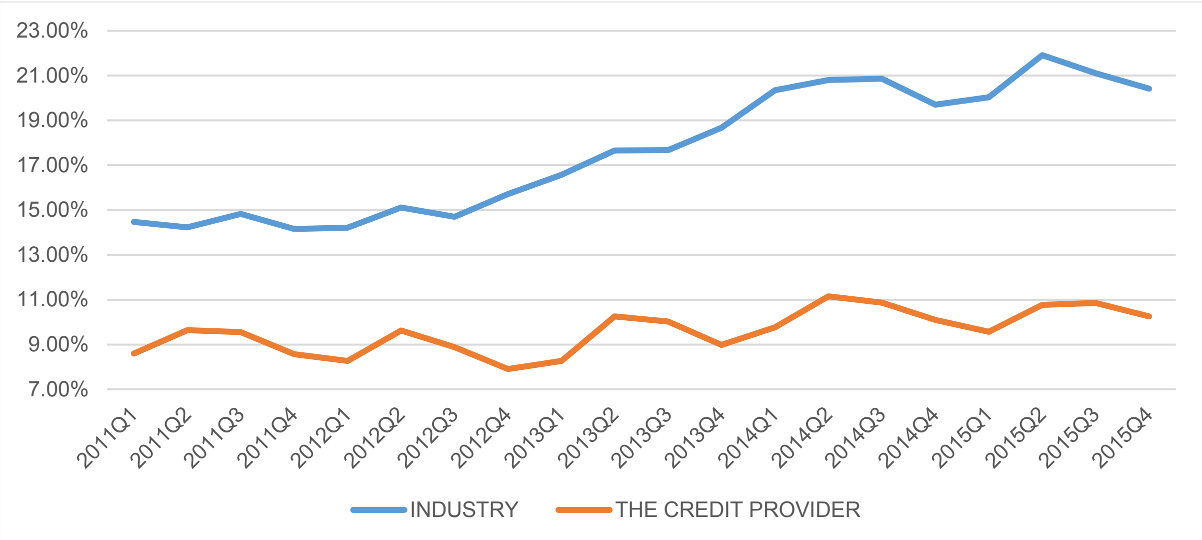
The NPLR for the credit provider was plotted over a five year, from Mar'11 to Feb'16 and is shown in Figure 4.1. The graph shows a seasonal pattern in the data with peaks towards the middle of each year and troughs at beginning of a calendar year. The underlying trend in the NPL rate has a positive gradient during the in-sample period. The hold-out or out-of-sample period represents data from Mar'15 to Feb'16, where a negative trend emerges. Figure 4.1 highlights the contrasting linear slopes of the in-sample and out-of-sample periods.

Figure 4.1: NPLR for the in-sample and out-of-sample period of the credit provider



The quarterly NPLR between the first quarter of 2011 and the final quarter of 2015 for the unsecured credit industry is graphed in Figure 4.2. To facilitate a comparison, the quarterly NPLR of the credit provider was consolidated for the same time period.

Figure 4.2: The quarterly NPLR for the credit provider versus the industry



The NPLR of the industry shows a steady upward trend from the last quarter of 2012, reflecting more aggressive credit policies for unsecured credit providers after the economic recession. The rate was subdued in 2011 and 2012, consistent with tighter credit appetite levels, increased risk averseness of credit businesses and the lower interest rate cycle. The NPLR of the credit provider reveals a distinct cyclical nature, in contrast to the industry, where the cyclical pattern is less pronounced. During a comparable period, the underlying long-term trend of the industry has steeper growth than the credit provider.

Basic descriptive statistics relating to central tendency and variability have been shown for the credit provider relative to the industry. To facilitate comparison, data for the credit provider has been prepared as a quarterly series. The average NPLR for the industry between Q1 2011 and Q4 2015 was 17.7%, almost twice that of the credit provider, at 9.6%. The NPLR for the total unsecured credit industry has experienced greater fluctuation than the credit provider, indicated graphically by Figure 4.2 above and by the variability measures contained in Table 4.1. The industry has a higher standard deviation and inter-quartile range than the credit provider, implying higher variability in its NPLR. A graphical illustration of the industry's higher dispersion measures is contained in the box and whisker plots in Figure 4.3 below.

Table 4.1: Basic statistical measures for the credit provider and the industry

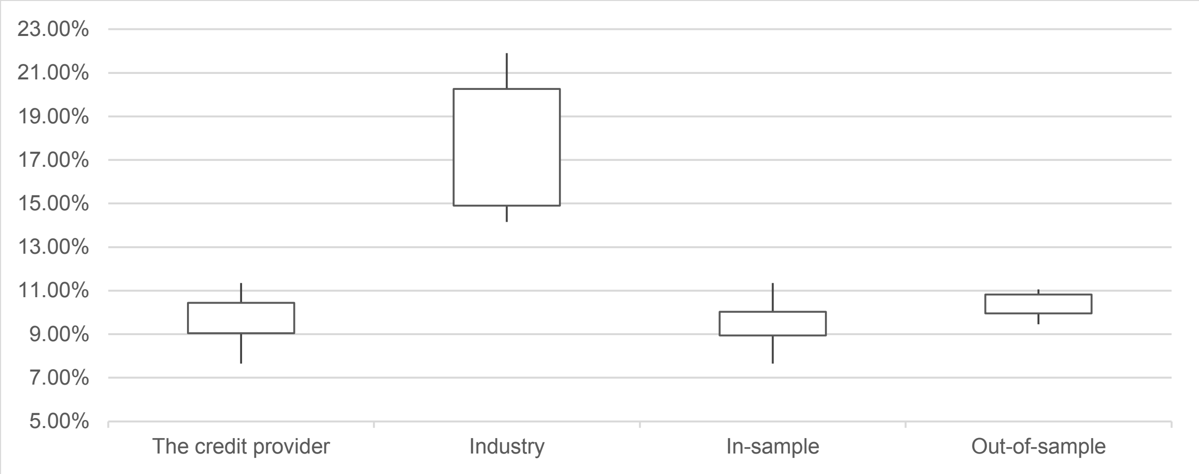
INDUSTRY (2011Q1-2015Q4)				CREDIT PROVIDER (2011Q1-2015Q4)			
Location		Variability		Location		Variability	
Mean	0.17659557	Standard deviation	0.02809398	Mean	0.09591711	Standard deviation	0.00964056
Median	0.17667017	Variance	0.00078927	Median	0.09629642	Variance	0.00009294
		Range	0.07756984			Range	0.03246902
		Interquartile range	0.05572019			Interquartile range	0.01439090

Basic statistical measures for the in-sample and out-of-sample periods of the credit provider's NPLR are included in Table 4.2 below. The NPLR in the 48 month in-sample period shows greater volatility than the out-of-sample period, but this is expected when a four year period is reviewed against a year. However, the 12 month out-of-sample period has a higher average NPLR than the in-sample period, consistent with the upward trending time series. The box and whisker plots for credit provider, industry, in-sample period and out-of-sample period are exhibited in Figure 4.3.

Table 4.2: Basic statistical measures for the in-sample and out-of-sample periods

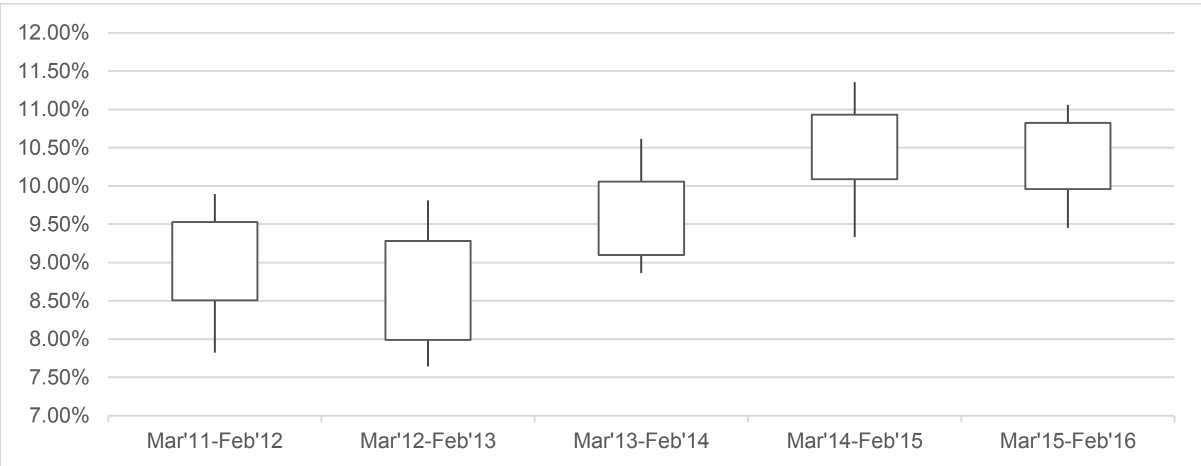
IN-SAMPLE (MAR'11-FEB'15)				OUT-OF-SAMPLE (MAR'15-FEB'16)			
Location		**Variability**		**Location**		**Variability**	
Mean	0.09442203	Standard deviation	0.00968087	Mean	0.10382313	Standard deviation	0.00539199
Median	0.09429971	Variance	0.00009372	Median	0.10485677	Variance	0.00002907
		Range	0.03712161			Range	0.01601330
		Interquartile range	0.01097589			Interquartile range	0.00869192

Figure 4.3: Box and whisker plots for the credit provider, industry, in-sample and out-of-sample periods



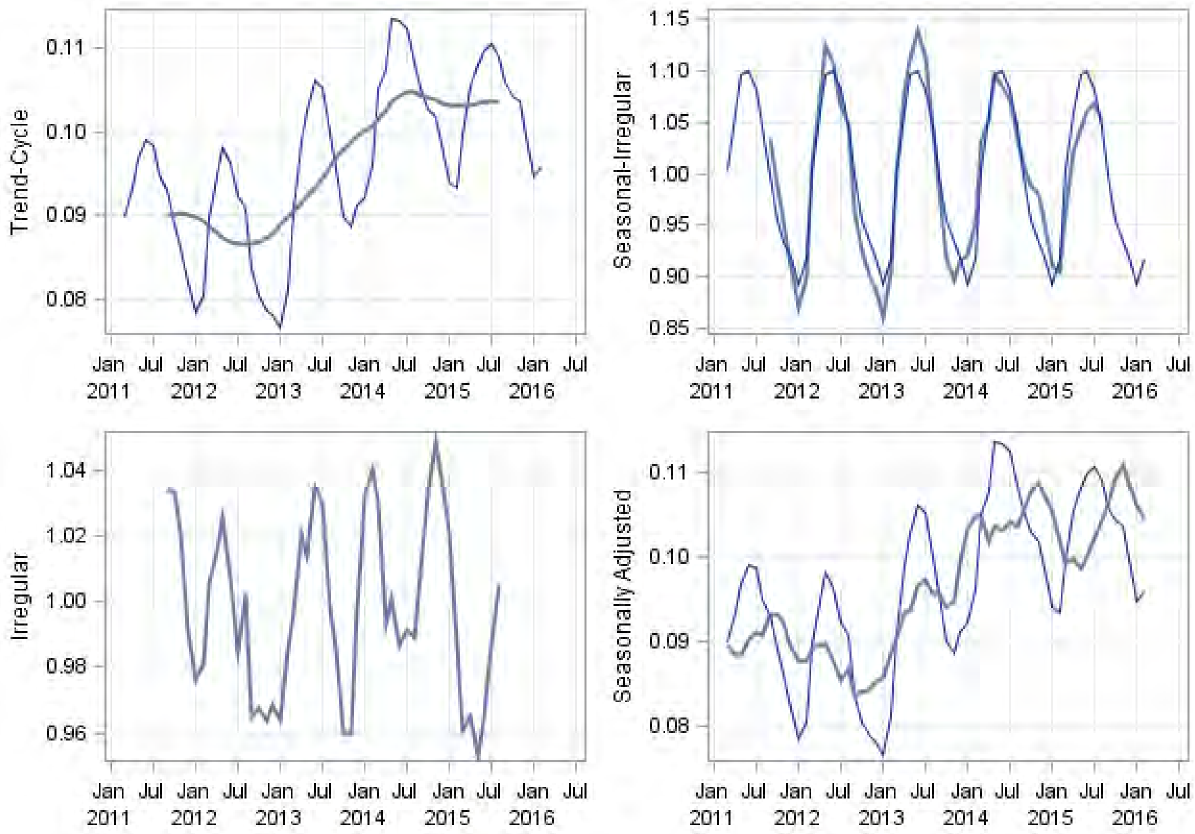
The trend in the NPLR over the five year period may be visualized by the box and whisker plots, displayed in Figure 4.4. The distribution of the NPLR for the credit provider shows the highest range and inter-quartile spread between Mar’12 and Feb’13. There appears to be a sideway shift in the NPLR in the most recent year, after a three year period of steady increases from Mar’12 to Feb’15. This is consistent with a prudent credit risk strategy that aims to keep NPLs below a certain threshold.

Figure 4.4: Box and whisker plot for the credit provider’s yearly NPLR



A seasonal decomposition highlights the trend, seasonal components and irregular terms in a time series. It helps to identify the stable versus the moving seasonality components of the observed data. The seasonal decomposition for the monthly NPLR of the credit provider is seen in Figure 4.5. After initially decreasing YoY, the long-term progression of the NPLR is upward with a flattening over the most recent twelve months. The irregular component reflects the residuals of the series after the trend and seasonal components have been extracted. The seasonal irregular plot shows the movement in the NPLR after accounting for the seasonal trend in the series. The seasonally adjusted plot is where the seasonal variation has been excluded.

Figure 4.5: Seasonal decomposition and adjustment for the NPLR of the credit provider



Based on the descriptive analysis contained in this subsection, the long-term trend in the data is upward and there is visual evidence of seasonal patterns. The credit provider has maintained both a lower average level and lower variation in its NPLR, relative to the industry over the past five years, suggesting a more conservative credit stance in comparison to other financial providers. The following subsection applies statistical tests for seasonality and stationarity.

4.3. Initial tests

The statistical tests of seasonality and stationarity are performed on the entire data series, consisting of sixty data observation points. The data series is labelled as NPL_RATE. In the next chapter, the data is split 80:20, with the 80% of the data used for model development and 20% retained for model validation.

4.3.1. Seasonality test

The initial tests commenced with checking for the presence of stable, moving and identifiable seasonality. The approach adopted for seasonality testing is aligned to the method discussed in Chapter 3. A combined test for the existence of identifiable seasonality was conducted, after performing tests for stable seasonality and moving seasonality. The result of the test for stable seasonality is presented in Table 4.3. The null hypothesis of stable seasonality not being present is rejected at a 0.1% level of significance. This statistically significant result affirms the visual indication of seasonality in the data.

Table 4.3: Test for stable seasonality

Test for the Presence of Seasonality Assuming Stability					
	Sum of Squares	DF	Mean Square	F-Value	
Between Months	3108.932	11	282.6302	236.1574	**
Residual	57.4458	48	1.196787		
Total	3166.378	59			

The test for moving or YoY seasonality is displayed in Table 4.4 below. The null hypothesis states that no moving seasonality is present in the data and is rejected at a 5% level of significance, since the calculated F-statistic, $F_{3,33}$= 3.14781 is larger than the critical F-statistic. This result indicates statistical evidence of YoY change in the seasonal pattern of the series.

Table 4.4: Test for moving seasonality

Moving Seasonality Test					
	Sum of Squares	DF	Mean Square	F-Value	
Between Years	10.55671	3	3.518905	3.14781	
Error	36.89036	33	1.11789		

The Kruskal-Wallis statistic is a nonparametric test for seasonality, as it does not make normality assumptions about the data. Under the null hypothesis, it is assumed that no stable seasonality exists in the time series. The results in Table 4.5 indicate high statistical evidence of the presence of stable seasonality in the series, as the null hypothesis is rejected at less than a 1% level of significance.

Table 4.5: Kruskal-Wallis nonparametric test for stable seasonality

Nonparametric Test for the Presence of Seasonality Assuming Stability		
Kruskal-Wallis Statistic	DF	Probability Level
57.17967	11	.00%

The presence of identifiable seasonality is tested with the procedure outlined in Chapter 3. The results for the combined seasonality test are summarized in Table 4.6. The test verifies that the presence of stable seasonality is not dominated by moving seasonality. The steps for testing identifiable seasonality have been noted below:

- The null hypothesis in the stable seasonality test is rejected at a 0.1% level of significance, observed in the results contained in Table 4.3.
- Since the null hypothesis is rejected from the step above, the quantities, X_1, X_2 and X are calculated, taking the values of F_{stb} and F_{mvg} from Table 4.3 and Table 4.4 as 236.1574 and 3.14781 respectively:
    - $X_1 = \frac{7}{F_{stb}} = \frac{7}{236.1574} = 0.029641$
    - $X_2 = \frac{3*F_{mvg}}{F_{stb}} = \frac{3*3.14781}{236.1574} = 0.039988$
    - $X = \frac{X_1 + X_2}{2} = \frac{0.029641+0.039988}{2} = 0.034815$
    - These calculations are presented in Table 4.6 below.
- The null hypothesis of the stable seasonality test was rejected at a 0.1% level of significance, the null hypothesis of the moving seasonality test was rejected at a 5% level of significance and the null hypothesis for the Kruskal-Wallis test was rejected at a 1% level of significance. These results were displayed in Table 4.3, 4.4 and 4.5. Furthermore, each of the quantities X_1, X_2 and X were calculated to be less than one, which suggests that the null hypothesis of no identifiable seasonality is rejected. The combined test

concludes that there is statistical evidence to suggest the presence of identifiable seasonality in the observed data.

Table 4.6: Combined test for the existence of identifiable seasonality

Summary of Results and Combined Test for the Presence of Identifiable Seasonality	
Seasonality Tests:	Probability Level
Stable Seasonality F-test	0.000
Moving Seasonality F-test	0.038
Kruskal-Wallis Chi-square Test	0.000
Combined Measures:	Value
T1 = 7/F_Stable	0.03
T2 = 3*F_Moving/F_Stable	0.04
T = (T1 + T2)/2	0.03
Combined Test of Identifiable Seasonality:	Present

4.3.2. Stationarity test

The significant statistical evidence of seasonality in the data has implications for stationarity of the data series. In order to avoid spurious results from the time series model development, the requirement for stationarity must be satisfied. From Chapter 3, the property of stationarity implies that the mean, variance and autocorrelation structure of a series does not vary over time. The PP test for stationarity is used in this analysis due to its ease of interpretation from the statistical package applied.

The null hypothesis in the PP test assumes that the series contains a unit root. Failure to reject the null hypothesis suggests that the data is not stationary. Under the zero mean case, the data is assumed to have a mean of 0, irrespective of whether unit roots exist or not. Under the single mean case, it is assumed that the series fluctuates around a constant mean μ. Under the trend assumption, it is possible that the stationary deviations occur around trend, rather than occurring as stationary deviations around a mean. This implies that the tests for unit roots may not be rejected albeit the deviations from the trend are white noise (support.sas.com[3]).

[3] Stationarity= (Phillips) Available: http://support.sas.com/documentation/cdl/en/etsug/60372/HTML/default/viewer.htm#etsug_autoreg_sect013.htm

The result of the PP tests for stationarity conducted on the original data series is shown in the Appendix B, Table B1. The null hypothesis of a unit root fails to be rejected in each of the zero mean, single mean and trend assumption scenarios at a 5% level of significance. At each of the lags, up to lag 12, the probability < *Tau* or probability value (p-value) is greater than 0.05. The conclusion of this test is that the original NPL_RATE is not a stationary series.

To induce stationarity, the first difference of the original series was generated. This involved creating a new series by subtracting the current period's observation ($NPLR_t$) with the observation from the previous period ($NPLR_{t-1}$). The PP unit root test for stationarity was then conducted on this first differenced series and the results are observed in Table B2 of Appendix B. Under the zero mean assumption, the null hypothesis can be rejected at less than a 1% level of significance at each of the lags, up to 12. Under a constant, single mean assumption, the null hypothesis may be rejected at a 10% level of significance. Under the trend assumption, the unit root null hypothesis may only be rejected at a 30% significance level at lags closer to 12.

Following the results of seasonality tests which indicated the presence of identifiable seasonality, the first differenced series was further differenced seasonally. This seasonal differencing involved subtracting the current observation of the first differenced series with the observation of the first differenced series twelve months ago. The PP unit root test result generated for this first seasonally differenced series is contained in Table B3 of Appendix B. The null hypothesis of a unit root is rejected at less than 0.1% significance level under the zero mean assumption, and at a p-value of 0.1% under the single mean and trend assumption types. This highly significant result indicates that there is statistical evidence for rejecting the null hypothesis of a unit root in the series. The condition of stationarity is met under the zero mean, constant mean, and trend assumptions when the NPL_RATE is transformed into a first differenced, seasonally differenced series. Failure to satisfy the requirement for stationarity may lead in spurious results and misleading statistical conclusions.

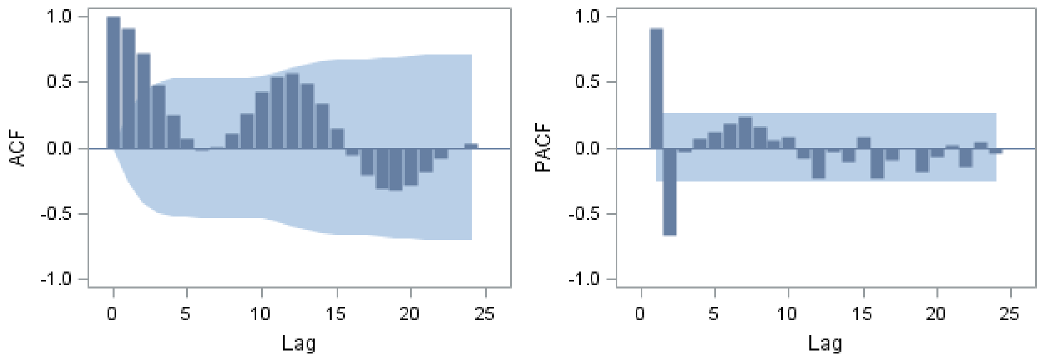
This result will be used in the model building phase of this chapter. The initial tests for seasonality suggest the presence of stable and moving seasonality in the data, which justifies the selection of a seasonal time series model. The seasonal pattern in the data repeats every twelve months. Thus, a SARIMA$(p,d,q)(P,D,Q)_{12}$ will be selected. The test for stationarity indicates that given the selection of SARIMA$(p,d,q)(P,D,Q)_{12}$, the condition for stationarity is satisfied when $d=1$ and $D=1$. It would therefore be appropriate to fit the data series with a SARIMA$(p,1,q)(P,1,Q)_{12}$ model. The parameters p,q, P and Q is estimated using the ACF, PACF and ESACF plots.

4.4. Parameter estimation

Although the initial tests of seasonality and stationarity were performed on the entire data set, the model will be developed using the in-sample period. Model selection will be validated using the out-of-sample period. For model evaluation and adequacy, the performance in both the in-sample and out-of-sample periods will be analysed. The forecasting ability of the model is assessed by its out-of-sample performance.

The primary identification tool used for parameter estimation is the ACF and PACF graphs. The ACF and PACF plots for the original NPL_RATE, before differencing, is shown in Figure 4.6 below. The PACF cut offs after lag 1 but the ACF slowly decays in a sine wave pattern.

Figure 4.6: ACF and PACF plots for the original NPL_RATE series



The analysis from the subsection 4.3 indicated the presence of identifiable seasonality and non-stationarity in the original data series. This result shows that the NPL_RATE may be adequately modelled by a seasonal time series model. The PP tests for unit roots indicated that stationarity is achieved when the first difference of the series is seasonally differenced. It is with this seasonal difference of the first differenced

NPL_RATE that models will be developed, satisfying the requirement for a stationary data series and capturing the seasonal pattern in the data.

4.4.1. Model 1: SARIMA$(0,1,0)(0,1,0)_{12}$

The first model selected assumes only a first difference and a seasonal difference to forecast the NPL_RATE, with no AR or MA terms. This represents the inclusion of only integrated components in the SARIMA model, denoted SARIMA$(0,1,0)(0,1,0)_{12}$. The model results are displayed in Table 4.7 and only contain the information criterion results as no AR or MA terms were specified in the model, i.e. $p=0$, $q=0$, $P=0$ and $Q=0$.

Table 4.7: Model results for the SARIMA$(0,1,0)(0,1,0)_{12}$

Variance Estimate	9.49E-6
Std Error Estimate	0.003081
AIC	-305.459
SBC	-305.459
Number of Residuals	35

4.4.2. Model 2: SARIMA$(1,1,1)(1,1,1)_{12}$

The second model chosen is the SARIMA$(1,1,1)(1,1,1)_{12}$ due to its conventional use in the application of residual modification. In forecasting energy demand for China, Dong *et al.* (2012) apply residual modification to a SARIMA$(1,1,1)(1,1,1)_{12}$ model. Chen *et al.* (2013) also considered a SARIMA$(1,1,1)(1,1,1)_{12}$ model for residual modification to forecast the volume of imported and exported air cargo in Taiwan.

This seasonal time series model contains autoregressive, integrated and moving average terms of order 1. However, the inclusion of a seasonal moving average component leads to unstable parameter estimates and early termination of the iteration process. This creates potentially misleading results. Additionally, from Table 4.8, the AR and MA components are not highly significant in the model.

This model will not be selected as the optimal one but its forecasting performance measures, namely, MAE and RMSE, will be retained to facilitate model comparison.

Table 4.8: Model results for the SARIMA$(1,1,1)(1,1,1)_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.16351	2.55731	0.06	0.9490	1
MA2,1	0.99829	0	Infty	<.0001	12
AR1,1	0.23136	2.51745	0.09	0.9268	1
AR2,1	0.63089	0.39416	1.60	0.1095	12

Variance Estimate	7.499E-6
Std Error Estimate	0.002738
AIC	-303.931
SBC	-297.71
Number of Residuals	35

The third, fourth and fifth models were selected by examining the ACF, PACF and ESACF of the seasonally differenced, first differenced series. The ACF and PACF for this stationary series, shown in Appendix B, Figure B1, cuts off after lag 0, which indicates that an adequate model may contain no AR or MA components, namely, p=0 and q=0. The ESACF was used to suggest the tentative orders for an ARIMA model. These tentative order selection tests can be observed in Table B4 of Appendix B. The table on the left suggests the tentative order for parameters of the original non-stationary series. The table on the right indicates the tentative order for parameters of the first differenced, seasonal differenced stationary series. These tentative order combinations were attempted and only those models with statistically significant parameter estimates were selected.

4.4.3. Model 3: SARIMA$(0,1,0)(1,1,0)_{12}$

Since the ESACF shows a tentative order selection of $p+d=13$ and $q=0$, the third model stipulated assumes a seasonal autoregressive term as $P=1$ in the differenced series. The model is denoted as SARIMA$(0,1,0)(1,1,0)_{12}$. The results exhibited in Table 4.9 below show statistical significance. The p-value is 0.16%, which indicates that modelling the differenced data using a seasonal autoregressive parameter of order 1 ($P=1$) will lead to a good fit.

Table 4.9: Model results for the SARIMA$(0,1,0)(1,1,0)_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
AR1,1	-0.54559	0.17277	-3.16	0.0016	12

Variance Estimate	6.933E-6
Std Error Estimate	0.002633
AIC	-311.22
SBC	-309.664
Number of Residuals	35

4.4.4. Model 4: SARIMA$(0,1,2)(0,1,0)_{12}$

The fourth model selected is one with a moving average term of order 2, denoted as SARIMA$(0,1,2)(0,1,0)_{12}$. Although the ESACF suggests an MA term of order 2 with $p+d=3$, 4 or 5, the inclusion of an AR component of orders 2, 3 or 4 leads to less satisfactory results. When a SARIMA$(2,1,2)(0,1,0)_{12}$ model is considered, the iteration process terminates and the parameter estimates do not converge. For the models: SARIMA$(3,1,2)(0,1,0)_{12}$ and SARIMA$(4,1,2)(0,1,0)_{12}$, the autoregressive and moving average components are not statistically significant at a 5% significance level, observed respectively in Table B5 and Table B6 of Appendix B. From Table 4.10, the p-value for the MA parameter of order 2 is 4.94%, less than a required level of 5%. This result suggests that the SARIMA$(0,1,2)(0,1,0)_{12}$ model is a plausible option.

Table 4.10: Model results for the SARIMA$(0,1,2)(0,1,0)_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.33502	0.17049	-1.96	0.0494	2

Variance Estimate	9.133E-6
Std Error Estimate	0.003022
AIC	-305.578
SBC	-304.023
Number of Residuals	35

4.4.5. Model 5: SARIMA$(0,1,2)(1,1,0)_{12}$

The fifth model combines the parameters of Model 3 and Model 4 and is denoted as SARIMA$(0,1,2)(1,1,0)_{12}$. The statistical significance of this model is seen in Table 4.11. The p-values for both the parameters $P=1$ and $q=2$ are less than 5%, which justifies its selection to estimate the NPL_RATE series. Furthermore, this model has the lowest AIC measure from the preceding models proposed in this subsection.

Table 4.11: Model results for the SARIMA$(0,1,2)(1,1,0)_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.33099	0.16682	-1.98	0.0472	2
AR1,1	-0.54173	0.17056	-3.18	0.0015	12

Variance Estimate	6.696E-6
Std Error Estimate	0.002588
AIC	-311.321
SBC	-308.21
Number of Residuals	35

4.4.6. Summary

Table 4.12 provides a summary of these five models in terms of their information criteria. Model 3: SARIMA$(0,1,0)(1,1,0)_{12}$ has the lowest BIC measure, while Model 5, SARIMA$(0,1,2)(1,1,0)_{12}$ has the lowest AIC measure. These results indicate that Model 3 and Model 5 have better fit the data series, compared to Models 1, 2 and 4. If the diagnostic checks conducted on the residual terms of these models prove to be satisfactory, these models will be selected for the residual modification process.

Table 4.12: Summary of parameter estimation

Model	Notation	AIC	BIC
Model 1	SARIMA$(0,1,0)(0,1,0)_{12}$	-305.459	-305.459
Model 2	SARIMA$(1,1,1)(1,1,1)_{12}$	-303.931	-297.710
Model 3	SARIMA$(0,1,0)(1,1,0)_{12}$	-311.210	-309.664*
Model 4	SARIMA$(0,1,2)(0,1,0)_{12}$	-305.578	-304.023
Model 5	SARIMA$(0,1,2)(1,1,0)_{12}$	-311.321*	-308.210

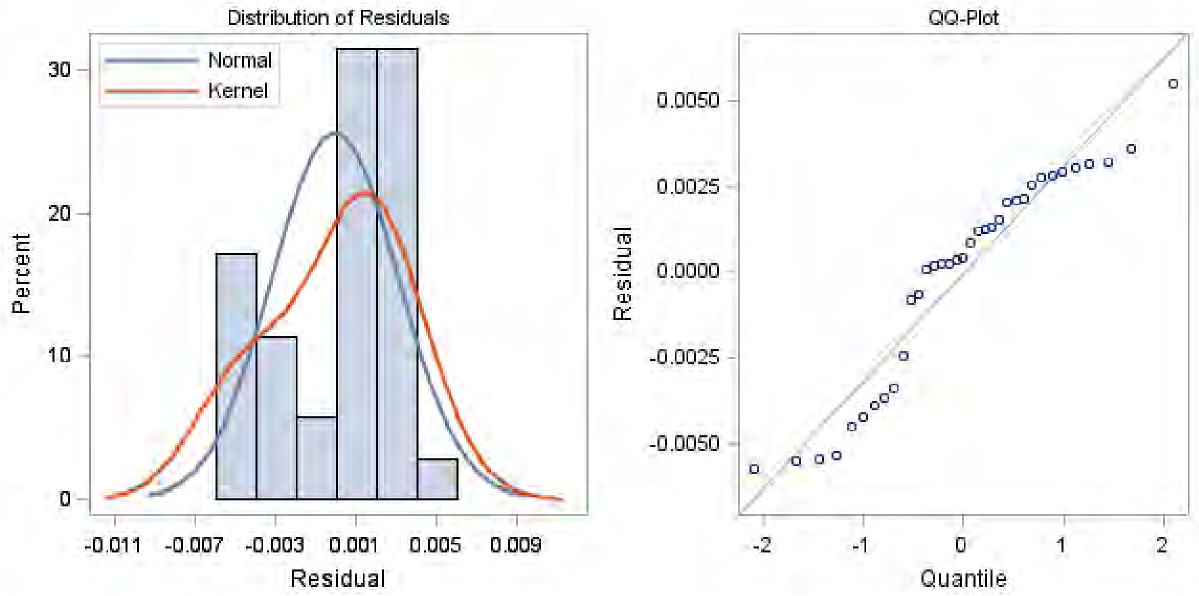
* denotes the lowest value

4.5. Diagnostic checking

The next phase in the model build involves conducting diagnostic tests. In the preceding subsection, the parameters were estimated through a tentative order selection process. This section performs diagnostic tests on the residual terms of the proposed models.

The residual diagnostics for the stationary (first differenced, seasonally differenced) series is plotted in Figure 4.7 below, where the distribution of residuals and the QQ plot reveals that the residuals are almost normally distributed. However, statistical tests for autocorrelation, homoscedasticity and stationarity have been performed on the residual series for each of the potential SARIMA models stipulated in section 4.4. The results of diagnostic tests per model are included in Appendix B.

Figure 4.7: Residual normality diagnostics for the NPL (1,12) series



4.5.1. Model 1: SARIMA(*0,1,0*)*(0,1,0)*$_{12}$

For Model 1, the null hypothesis in the DW test of no autocorrelation in the residual terms, shown in Table B7.1 of Appendix B cannot be rejected at a 1% level for all orders, up to 12, which implies that the error terms show autocorrelation. However, at order 12, the null hypothesis for no negative autocorrelation can be rejected at a 5% level of significance. At order 2 and order 8, the null hypothesis for no positive autocorrelation can be rejected at a 10% level. Based on ARCH tests contained in Table B7.2 of Appendix B, the null hypothesis of homoscedastic variances in the error terms cannot be rejected as the p-values are significantly greater than 5% at every order. This result indicates that residuals have constant variance. The PP unit root test, displayed in Table B7.3 of Appendix B indicates rejection of the null hypothesis for the existence of a unit root in the series. Thus, the conclusion is that the residual terms of Model 1 are stationary.

4.5.2. Model 2: SARIMA(*1,1,1*)*(1,1,1)*$_{12}$

For Model 2, the null hypothesis of no autocorrelation cannot be rejected at a 1% level of significance for all orders, up to 12. However, at order 12, the null hypothesis of no negative autocorrelation in the errors terms is rejected at a 2.5% level of significance, as can be seen in Appendix B, Table B8.1. The result of the ARCH test, in Table B8.2 of Appendix B, indicates that the null hypothesis cannot be rejected as the p-values are more than 5% at every order. This suggests that the residuals of Model 2 have constant variance. The unit root test is rejected at 5% level of significance as the p-values are less than 0.01 under the zero mean, single mean and trend assumptions, evident in Table B8.3 in Appendix B. Thus, the residual series of Model 2 is stationary.

4.5.3. Model 3: SARIMA$(0,1,0)(1,1,0)_{12}$

For Model 3, the null hypothesis of no autocorrelation cannot be rejected at a 1% level of significance for all orders, up to 12, observed in Table B9.1 of Appendix B. At a 2.5% level of significance, the null hypothesis of no positive autocorrelation at order 1 is rejected. The null hypothesis of no positive autocorrelation is rejected at a 10% level of significance at orders 2 and 8. The ARCH test, contained in Table B9.2 in Appendix B, shows that the null hypothesis of constant variance in the residual terms fail to be rejected at 5% significance level at each order. This suggests homoscedastic error terms for Model 3. The unit root null hypothesis in the PP test can be rejected at a 5% level of significance, as seen in Table B9.3 of Appendix B. The requirement for the error terms being stationary is satisfied.

4.5.4. Model 4: SARIMA$(0,1,2)(0,1,0)_{12}$

For Model 4, the DW test is statistically significant at a 1% level of significance, shown in Table B10.1 of Appendix B. At order 8, the null hypothesis for no positive autocorrelation can be rejected at a 10% significance level. At order 12, the null hypothesis for no negative autocorrelation can be rejected at a level of significance of 2%. With these exceptions, the null hypothesis of no autocorrelation in the residuals cannot be rejected with 95% confidence. The ARCH test for constant variance is highly significant. The null hypothesis of homoscedastic variance fails to be rejected at 5% significance level at each order. This is displayed in Table B10.2 of Appendix B. The unit root test is highly significant, observed in Table B10.3 in Appendix B. The null hypothesis can be rejected at a 1% level of significance. The error terms are therefore assumed to be stationary.

4.5.5. Model 5: SARIMA$(0,1,2)(1,1,0)_{12}$

For Model 5, the DW test is statistically significant at a 3% level of significance, seen in Table B11.1 of Appendix B. The null hypothesis for no positive autocorrelation can however, be rejected at a 5% significance level at order 1. With this exception, the null hypothesis of no autocorrelation cannot be rejected with 95% confidence. The ARCH test for constant variance is highly significant. The null hypothesis of homoscedastic variance fails to be rejected at 5% significance level at each order, observed in Table B11.2 of Appendix B. The unit root test is statistically significant, evident in Table B11.3 of Appendix B. The null hypothesis can be rejected at a 5% level of significance. The condition of stationarity in the residual series is met.

4.6. Model selection

The performance measures of each of these models have been summarized in Table 4.13 below. Model 5 has the lowest MAE and RMSE measures, and it showed the best information criterion measures from Table 4.3, albeit only marginally better than Model 3. When the diagnostic checks were conducted on Model 5 and on Model 3, there was evidence to suggest positive autocorrelation at order 1. However, at a 1% level of significance, the null hypothesis of no positive autocorrelation will fail to be rejected and hence, the diagnostic checks will have a satisfactory outcome. Model 4 revealed negative autocorrelation in the error terms at order 12, with a p-value of 0.0158. Additionally, its information criterion measures were weaker than Model 2, Model 3 and Model 5, as seen in Table 4.12. Although the inclusion of a seasonal moving average term in Model 2 led to unstable parameter estimates, this model will be retained in the analysis due to its popularity as a seasonal forecasting tool. The diagnostic tests for Model 2 showed the possibility of negative autocorrelation at order 12, with a p-value of 0.0233. Following the aforementioned discussion, the residual modification technique will be applied to Model 2: SARIMA$(1,1,1)(1,1,1)_{12}$, Model 3: SARIMA$(0,1,0)(1,1,0)_{12}$ and Model 5: SARIMA$(0,1,2)(1,1,0)_{12}$.

Table 4.13: Summary of performance measures for selected SARIMA models

Metric	Model 1: SARIMA$(0,1,0)(0,1,0)_{12}$	Model 2: SARIMA$(1,1,1)(1,1,1)_{12}$	Model 3: SARIMA$(0,1,0)(1,1,0)_{12}$	Model 4: SARIMA$(0,1,2)(0,1,0)_{12}$	Model 5: SARIMA$(0,1,2)(1,1,0)_{12}$
MAE	0.00248338	0.00230212	0.00216751	0.00239479	0.00214101*
RMSE	0.00310	0.00284	0.00279	0.00301	0.00272*

Chapter 5

Analysis and Results

The aim of this chapter is to address two research questions. Firstly, the chapter explores whether there is benefit in selecting the optimal time series model to a given data set, before applying a residual modification technique. Secondly, it seeks to establish if residual modification improves forecasting performance. The residual modification applied in this analysis involves estimating the error terms with firstly, a time series model and secondly, a Fourier series. This chapter is partitioned into four subsections. The first section contains the in-sample and out-of-sample forecasting performance of the seasonal models estimated in the previous chapter, before modifying the error terms. The second section shows the results of the time series residual modification, while the results of the Fourier residual modification is contained in section three. Concluding remarks follow in section four.

5.1. Seasonal model performance

This subsection explores the performance of the seasonal models selected in the previous chapter. The first 48 months of the series is used for model estimation. The most recent 12 months of the series is used to measure forecasting ability. The in-sample and out-of-sample forecasting performance of the seasonal model before applying a residual modification technique is shown in Figure 5.1, 5.2 and 5.3 for SARIMA$(1,1,1)(1,1,1)_{12}$, SARIMA$(0,1,0)(1,1,0)_{12}$, and SARIMA$(0,1,2)(1,1,0)_{12}$.

Figure 5.1: Forecasting performance for Model 2: SARIMA$(1,1,1)(1,1,1)_{12}$

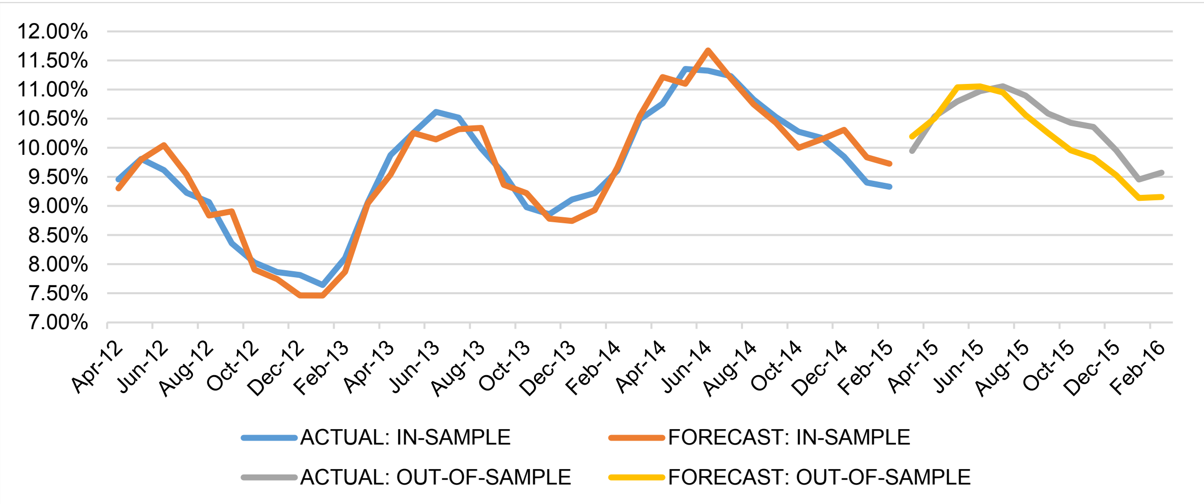


Figure 5.2: Forecasting performance for Model 3: SARIMA$(0,1,0)(1,1,0)_{12}$

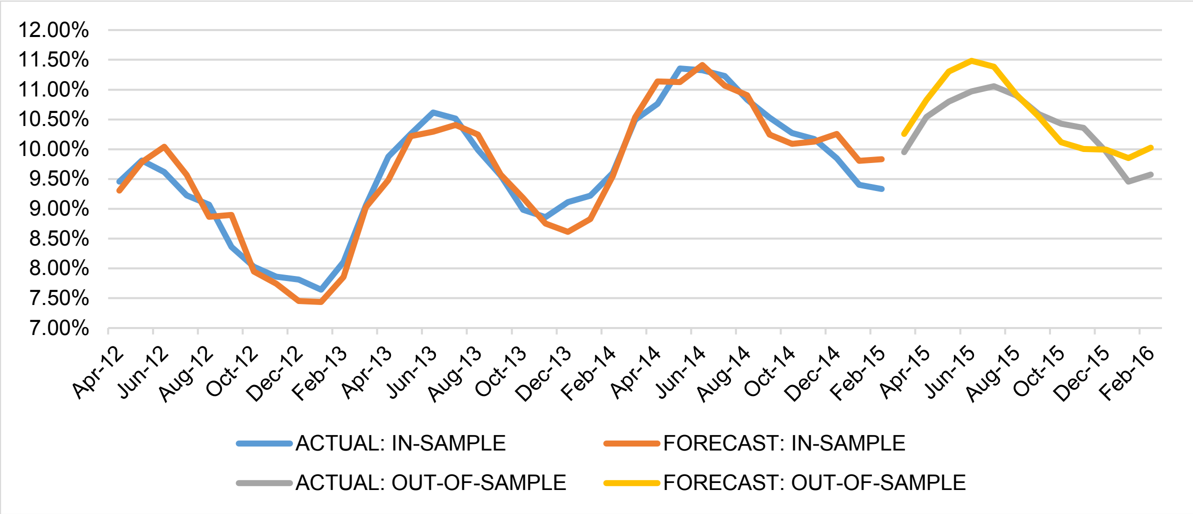
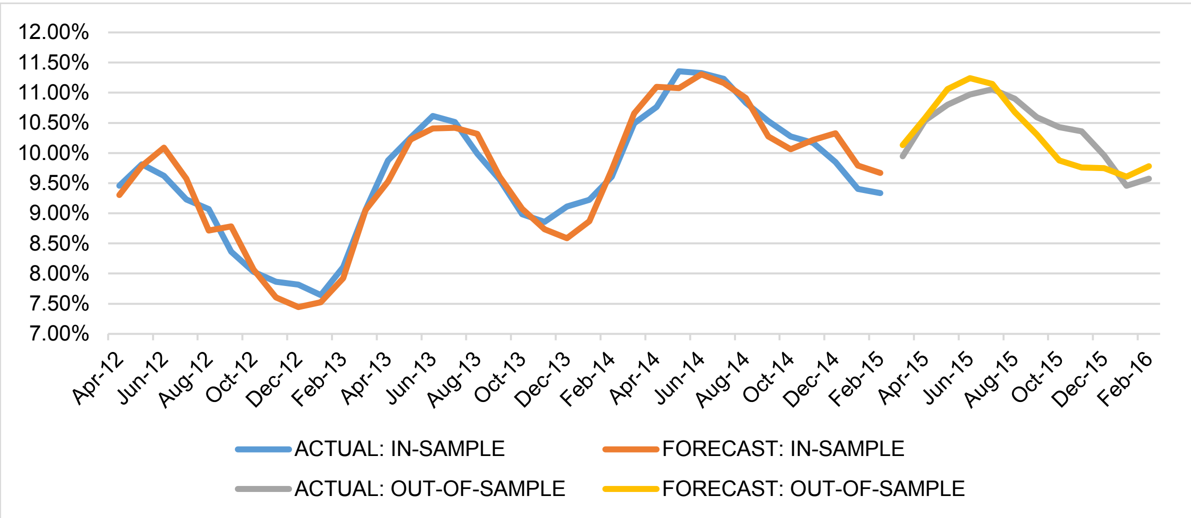


Figure 5.3: Forecasting performance for Model 5: SARIMA$(0,1,2)(1,1,0)_{12}$



Graphically, all three models appear to fit the data well as it captures the seasonal pattern of the series. Based only on visual inspection, it is difficult to identify which model is the best describes the data series. Thus, the performance measures discussed in section 3.5 are calculated to facilitate model comparison. The fit of the model is evaluated for the in-sample and out-of-sample periods. Due to the inclusion of a seasonal autoregressive component in the model, together with first and seasonal differencing of the series, thirteen observation points were lost, namely March 2011 to March 2012 (inclusive). From the performance measures summarized in Table 5.1, Model 5: SARIMA$(0,1,2)(1,1,0)_{12}$ shows the lowest error rates in both the in-sample and out-of-sample forecasting periods, thereby suggesting relatively better fit than Model 2 and Model 3. The lowest value per row is denoted with a *.

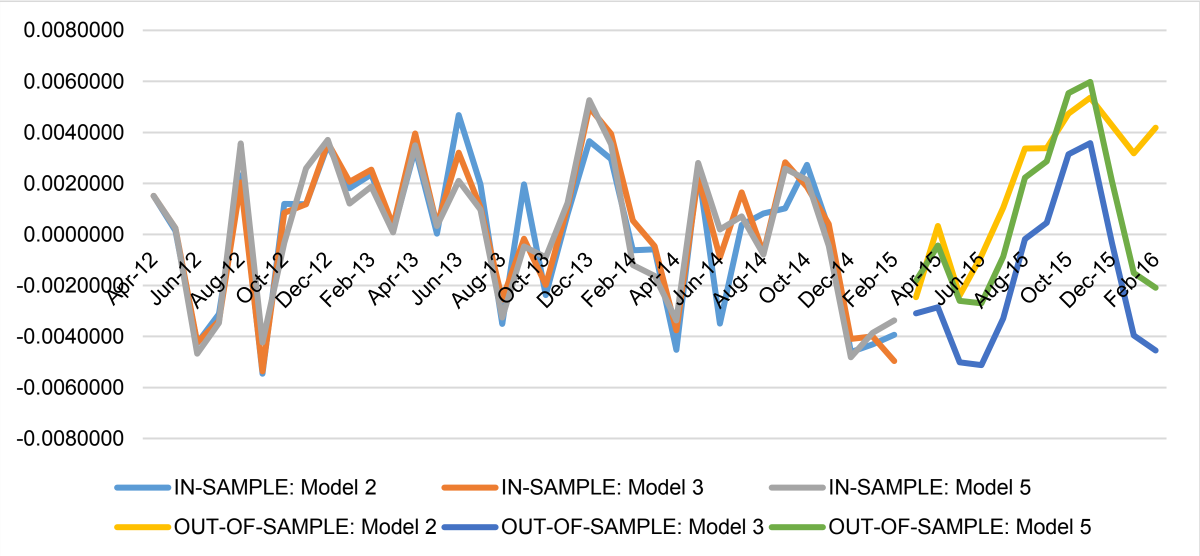
Table 5.1: Summary of performance for unmodified SARIMA models

Period	**Metric**	**Model 2:** SARIMA$(1,1,1)(1,1,1)_{12}$	**Model 3:** SARIMA$(0,1,0)(1,1,0)_{12}$	**Model 5:** SARIMA$(0,1,2)(1,1,0)_{12}$
In-sample	MAE	0.002345	0.002261	0.002199*
	RMSE	0.002805	0.00274	0.002666*
Out-of-sample	MAE	0.002964	0.002966	0.002561*
	RMSE	0.003336	0.003408	0.003013*

Although Model 5 has the lowest MAE and RMSE, a prudent credit risk strategy may prefer the forecasts of Model 3. In the out-of-sample period, Model 3 overestimates the NPL_RATE in several of the 12 months forecasted, while Model 5 underestimates the NPL_RATE in more months than it overstates it. Even though the overprovision of capital reserves for withstanding NPLs has an associated opportunity cost, it is a more conservative approach than projecting provisions that fall short of actual losses.

The residual terms of Models 2, 3, and 5 are plotted in Figure 5.4. Only 35 data points remain after the original seasonal model is fitted to the data. The residual series of the models are randomly distributed and oscillate around zero, which is further indication of the goodness of fit for these models. The out-of-sample fit of these models tend to be poorer than the in-sample fit, indicated by larger deviations from zero. This is confirmed in Table 5.1, where the MAE and RMSE measures in the out-of-sample period were greater than the in-sample period measures for each model.

Figure 5.4: Residual series of the selected models



In preparation for residual modification, the error terms of the selected models must be stationary. The diagnostic checks performed in section 4.5 confirmed stationarity and constant variance in the residual series of each seasonal model. The check for no significant autocorrelation was satisfied at a 1% level but failed at a 5% significance level for certain lags. To forecast the error terms for the models, no differencing of the series was required as the condition for stationarity has been met. The method to parameter estimation and model selection was outlined in chapter three and has been implemented in the pre-analysis phase. A similar approach is followed to select the time series models that adequately fit the residual terms of each seasonal model.

5.2. Time series residual modification

The first residual modification technique applied in this analysis involves estimating the residual terms of each model with a basic time series model.

5.2.1. Model 2: SARIMA$(1,1,1)(1,1,1)_{12}$

Using the tentative order selection tests from the ESACF, shown in Appendix C, Table C1, an AR(*12*) or SARIMA$(0,0,0)(1,0,0)_{12}$ model revealed the best fit to the residual terms of Model 2: SARIMA$(1,1,1)(1,1,1)_{12}$. The model result of the AR(*12*) process is shown in Table 5.2, with a highly significant p-value of 0.0025. Another possibility was to fit an AR(*8*) model to Model 2's error terms, but the p-value was 10.02%, seen in Table C2.1 of Appendix C. An ARMA(*4,2*) model was also tentatively tested; however, the AR and MA parameters were not significant at a 15% level, which can be observed in Table C2.2 of Appendix C. The information criterion measures for the AR(*12*) model were also better than the AR(*8*) and ARMA(*4,2*) models. Twelve further observation points were lost due to the inclusion of only an AR component of order 12 to forecast the error terms.

Table 5.2: Model results of AR(*12*) *for the residual series of Model 2:* SARIMA$(1,1,1)(1,1,1)_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
AR1,1	-0.50998	0.16855	-3.03	0.0025	12

Variance Estimate	6.022E-6
Std Error Estimate	0.002454
AIC	-316.78
SBC	-315.225
Number of Residuals	35

5.2.2. Model 3: SARIMA$(0,1,0)(1,1,0)_{12}$

The residual series for Model 3: SARIMA$(0,1,0)(1,1,0)_{12}$, was fitted with an ARMA$(8,1)$ time series model. The parameter estimates are observed in Table 5.3. The tentative order selection tests, presented in Appendix C, Table C3, reveal that autoregressive components of order 2, 3, 4, 5 and 8 could be considered. Compared to the results of the AR(2), AR(3), AR(4), AR(5) and AR(8) models contained Appendix C, Table C4.1 to Table C4.5, the standard error estimate and AIC measure is lowest for the ARMA$(8,1)$ model, seen in Table 5.3. Although the p-values for these parameters are greater than 10%, the p-values for the AR orders of 2, 3, 4, 5 and 8 are higher still at 0.2596, 0.3136, 0.3765, 0.2866 and 0.2127, respectively. The AIC measure for these models were also greater than the ARMA$(8,1)$ model.

Table 5.3: Model results of ARMA$(8,1)$ *for the residual series of Model 3:* SARIMA$(0,1,0)(1,1,0)_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.23510	0.17320	-1.36	0.1746	1
AR1,1	0.30979	0.18888	1.64	0.1010	8

Variance Estimate	6.905E-6
Std Error Estimate	0.002628
AIC	-313.785
SBC	-310.674
Number of Residuals	35

5.2.3. Model 5: SARIMA$(0,1,2)(1,1,0)_{12}$

The residuals for Model 5: SARIMA$(0,1,2)(1,1,0)_{12}$, was fitted with an ARMA$(12,1)$ or equivalently, a SARIMA$(0,0,1)(1,0,0)_{12}$ model. The tentative orders chosen included an AR(3), ARMA$(5,1)$, ARMA$(11,1)$ and an ARMA$(8,1)$ model. The results of these models are contained in Appendix C, Table C5.1 to Table C5.4. The ARMA$(12,1)$ model generated the lowest AIC, BIC and standard error measures, even though the parameters were only statistically significant at an 11% level. These results are highlighted in Table 5.4.

Table 5.4: Model results of ARMA$(12,1)$ *for the residual series of Model 5:* SARIMA$(0,1,2)(1,1,0)_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.30353	0.16798	-1.81	0.0708	1
AR1,1	-0.31175	0.19241	-1.62	0.1052	12

Variance Estimate	6.372E-6
Std Error Estimate	0.002524
AIC	-316.137
SBC	-313.027
Number of Residuals	35

5.2.4. Combined results

By presenting a combination model, it can be determined if there is an improvement in performance when the residual terms are modified with a time series model. The combined model forecast is a straight addition of the original seasonal model forecast and the residual model forecast.

The time series residual modified models for Model 2: TS_SARIMA$(1,1,1)(1,1,1)_{12}$, Model 3: TS_SARIMA$(0,1,0)(1,1,0)_{12}$, and Model 5: TS_SARIMA$(0,1,2)(1,1,0)_{12}$ are shown below, in Figure 5.5, Figure 5.6 and Figure 5.7, respectively. The forecasts generated from these modified models have been overlaid with the original, unmodified seasonal model forecasts to facilitate comparison. As seen in Figures 5.5 to Figure 5.7, the in-sample performance of the time series residual modified models appears to be better than the unmodified, seasonal model counterparts. This observation is confirmed by the lower MAE and RMSE measures in Table 5.2.

Figure 5.5: Forecasting performance of the TS_SARIMA$(1,1,1)(1,1,1)_{12}$

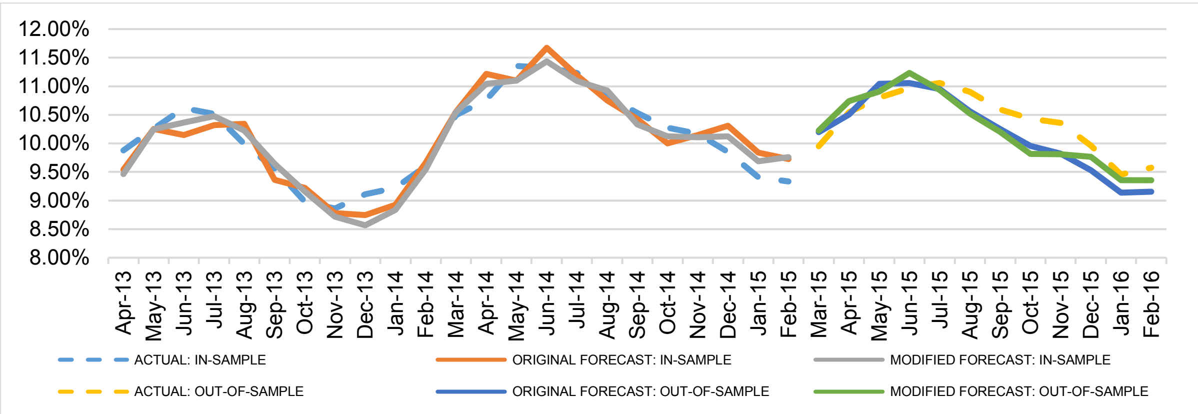


Figure 5.6: Forecasting performance of the TS_SARIMA$(0,1,0)(1,1,0)_{12}$

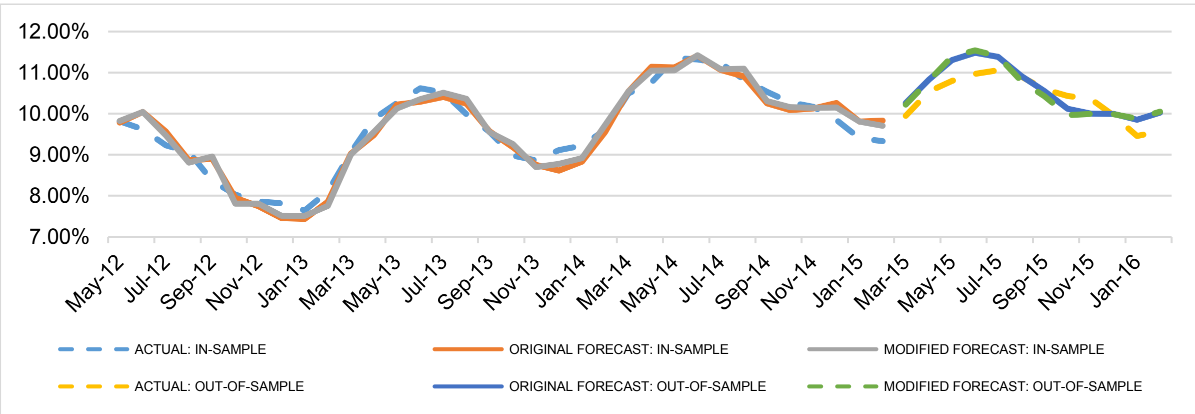
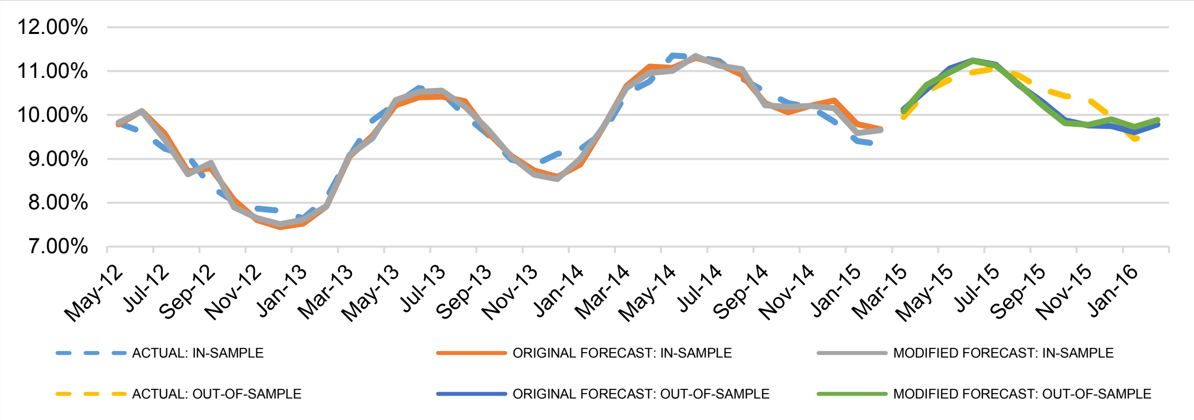


Figure 5.7: Forecasting performance of the TS_SARIMA($0,1,2$)$(1,1,0)_{12}$



The MAE and RMSE were computed for each time series residual modified model. It is evident from Table 5.5 that the residual modification technique improves the in-sample forecasting performance for each model. This is indicated by the relatively lower MAE and RMSE statistics from the time series residual modified model compared to the original, unmodified seasonal models. In addition, the out-of-sample performance for the residual modified TS_SARIMA($1,1,1$)$(1,1,1)_{12}$ model is slightly better than the unmodified SARIMA($1,1,1$)$(1,1,1)_{12}$. This is consistent with the finding of Dong *et al.* (2012). For the other two time series residual modified models, the out-of-sample performance is worse than the unmodified, seasonal model. It is possible that the failure to select a highly statistical significant time series models to fit the error terms may result in less than optimal performance. The difficulty in model selection arises because the error terms are uncorrelated and does not show persistence from one period to the next.

Table 5.5: Summary of performance measures for the time series residual modified models

Period	**Metric**	**Model 2:** TS_SARIMA($1,1,1$)$(1,1,1)_{12}$	**Model 3:** TS_SARIMA($0,1,0$)$(1,1,0)_{12}$	**Model 5:** TS_SARIMA($0,1,2$)$(1,1,0)_{12}$
In-sample	MAE	0.002026*	0.002216	0.002042
	RMSE	0.002455*	0.002617	0.002530
Out-of-sample	MAE	0.002839	0.003395	0.002667*
	RMSE	0.003264	0.003798	0.003189*

Another result of this exercise is that the time series residual modified model, TS_SARIMA($0,1,2$)$(1,1,0)_{12}$ shows superior out-of-sample performance than both the TS_SARIMA($1,1,1$)$(1,1,1)_{12}$ and TS_SARIMA($0,1,0$)$(1,1,0)_{12}$. This result suggests that it is beneficial to select the most adequate model to fit the data prior to applying a time series residual modification to the seasonal model forecasts.

5.3. Fourier series residual modification

The coefficients generated from the Fourier series were used to forecast the residual terms of each of the seasonal models selected. The performance of this residual modification technique was compared to the results obtained from the original seasonal model and the time series residual modification model to establish if the Fourier residual modification technique is better.

When representing the data as a mixture of sine and cosine waves, the Fourier representation is apt as it identifies peaks and troughs in the series. It is an appealing method to fit randomly distributed error terms that oscillate around zero. The literature review confirmed that the Fourier series produces good forecasting results when the determinants of the underlying series are difficult to identify.

In order to generate Fourier forecasts, the sine and cosine terms were summed, as per *Definition 11*. For each model, the cosine and sine coefficients were generated from the Fourier algorithm for frequencies between *0* and *π*, where the frequency is represented as: $\frac{2*\pi*t}{N}$; for $t=1, 2\ldots \frac{N}{2}$ and $N= 35$ in the analysis. The results of the Fourier algorithm are contained in Appendix C, in Table C6 for Model 2, Table C7 for Model 3 and Table C8 for Model 5. The periodograms which plot the magnitude or power of the frequencies (on the vertical axis) ranging from *0* to *π* (on the horizontal axis) are shown in Figures C1, C2 and C3 for the residual series of Model 2, 3 and 5, respectively. In each graph, the second frequency tends to have relatively good power.

Ludlow and Enders (2000) found that due to the large number of coefficients in the Fourier algorithm, the out-of-sample performance was poor. However, forecasting performance improved when only the first frequency was used in estimating non-linear ARMA models. As such, this analysis further tests if forecasting accuracy could be enhanced by considering a lower number of frequencies. The impact of reducing the number of Fourier frequencies in the residual forecast is shown in Appendix C, Table C9 for Model 2, in Table C10 for Model 3 and in Table C11 for Model 5. With the residual terms of Model 2, the out-of-sample performance is the best when only 3 frequencies are used in the forecast (excluding the average of the series, a_0), shown by the lowest MAE and RMSE measures. For forecasting Model 3 residual terms, the MAE for the out-of-sample period is the lowest when three frequencies are used, while the RMSE is smallest when nine frequencies are considered. The forecasting performance for the residual series of Model 5 is the best when six Fourier frequencies are used. However, the MAE and RMSE statistics for Model 5 are still lower when 3 frequencies are used, compared to when 16 frequencies are used. Furthermore, the periodograms shown in Figures C1, C2 and C3 reveal high power at the second frequency. Thus, the Fourier residual modification technique also evaluates forecasting performance when only three frequencies are included in the forecast. A summary of model performance is contained in Table 5.6. The Fourier series modified models have been denoted as Model 2: FS_SARIMA$(1,1,1)(1,1,1)_{12}$, Model 3 as FS_SARIMA$(0,1,0)(1,1,0)_{12}$ and Model 5 as FS_SARIMA$(0,1,2)(1,1,0)_{12}$.

Table 5.6: Summary of performance measures for the Fourier series forecasts of the residual terms

	Period	Metric	Model 2: FS_SARIMA$(1,1,1)(1,1,1)_{12}$	Model 3: FS_SARIMA$(0,1,0)(1,1,0)_{12}$	Model 5: FS_SARIMA$(0,1,2)(1,1,0)_{12}$
Fourier with *N/2* terms	In-sample	MAE	0.0000000000000001925	0.0000000000000001366*	0.0000000001293
		RMSE	0.0000000000000002674	0.0000000000000001991*	0.0000000002066
	Out-of-sample	MAE	0.002774	0.003073	0.002718*
		RMSE	0.003508	0.003714	0.003149*
Fourier with 3 terms (excl. a_0)	In-sample	MAE	0.002035	0.001920*	0.001954
		RMSE	0.002457	0.002328*	0.002349
	Out-of-sample	MAE	0.002164	0.002214	0.001998*
		RMSE	0.002496	0.003027	0.002498*

From Table 5.6, the in-sample performance of each model improves when all the Fourier coefficients are included. This represents a significant lift from the in-sample performance of the seasonal models, summarised in Table 5.1, as well as the time series residual models, seen in Table 5.5. When all the Fourier coefficients are used to generate forecasts, the accuracy is poorer than the seasonal, unmodified model forecasts and the results are mixed when compared against the time series residual models. When only the three lowest Fourier frequencies are used to forecast, excluding the average or a_0 term, the out-of-sample performance improves for each model, indicated by lower MAE and RMSE figures. The forecast with three Fourier frequencies appears to capture the general, underlying trend of the series. The results generated for the out-of-sample performance is superior to both, the time series residual modified model results, in Table 5.5, and the seasonal, unmodified model results in Table 5.1 across all three models.

Another important result from this analysis is that the out-of-sample performance using the Fourier series to forecast the residual terms for Model 5 is better than Model 3 and Model 2. This suggests that the initial step of selecting the most adequate model to fit the data is beneficial. The full Fourier series representation of the in-sample data is extremely good but relatively poor in the out-of-sample period. The better out-of-sample or forecasting performance occurs when only the lowest three frequencies are combined. Instead of projecting distinct peaks and troughs in the data, the sum of the three lowest frequencies plot a gentle oscillation around the x-axis, and serves to better represent the unknown, uncorrelated errors.

From Figures 5.8 to 5.10, it can be observed that in the out-of-sample or forecasting window, the Fourier residual technique decreases the seasonal model estimate to bring it closer to the actual NPL_RATE. Towards the latter of the twelve month period, the Fourier residual model inflates the original seasonal model projections which are closer to the actual data points. These figures provide graphical evidence of the improved forecasting accuracy derived by only including a few, low frequencies.

Figure 5.8: Forecasting performance for Fourier series residual modified Model 2: FS_SARIMA$(1,1,1)(1,1,1)_{12}$

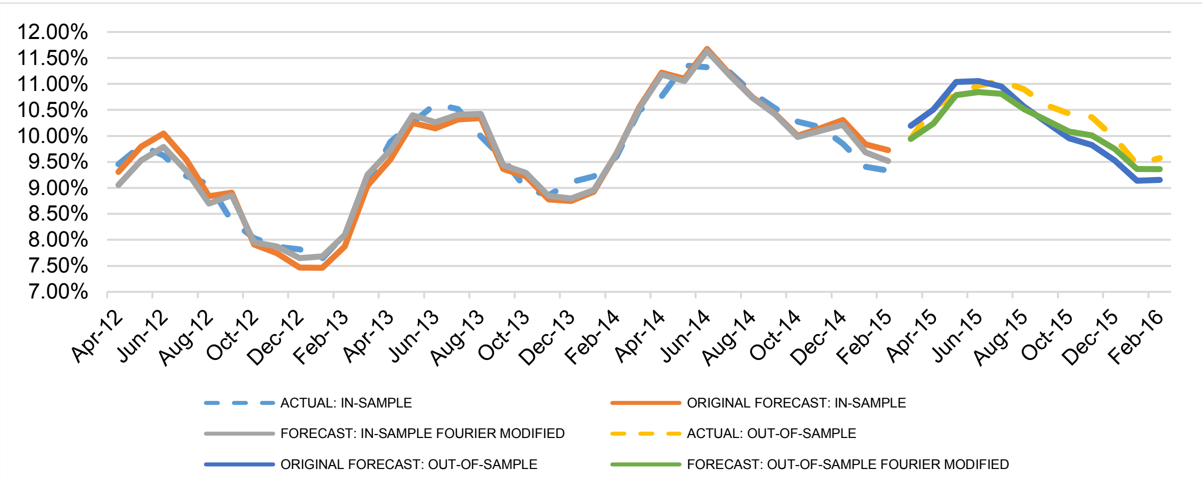


Figure 5.9: Forecasting performance for Fourier series residual modified Model 3: FS_SARIMA(*0,1,0*)(*1,1,0*)$_{12}$

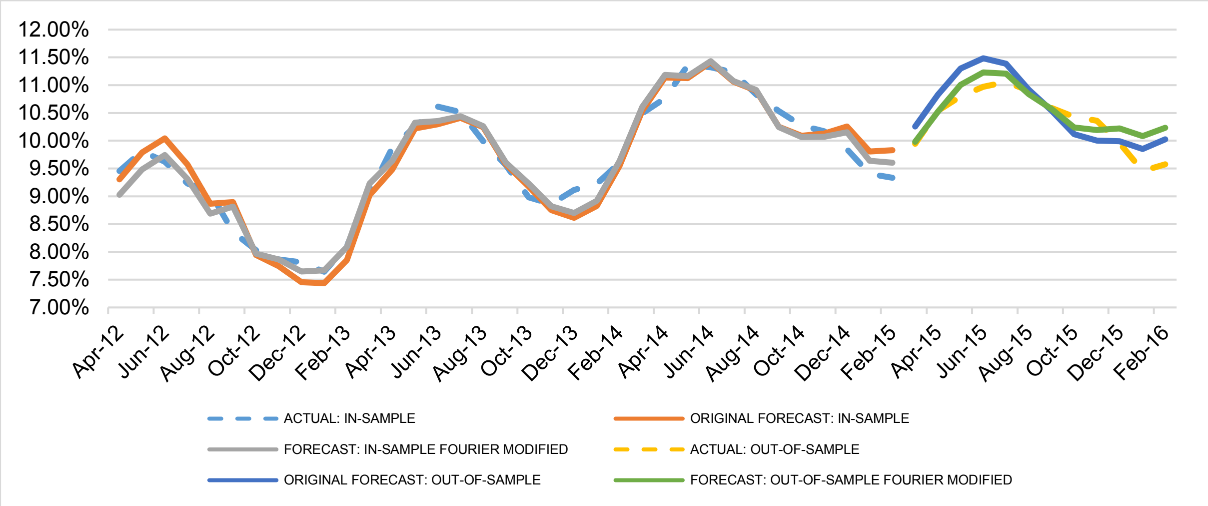
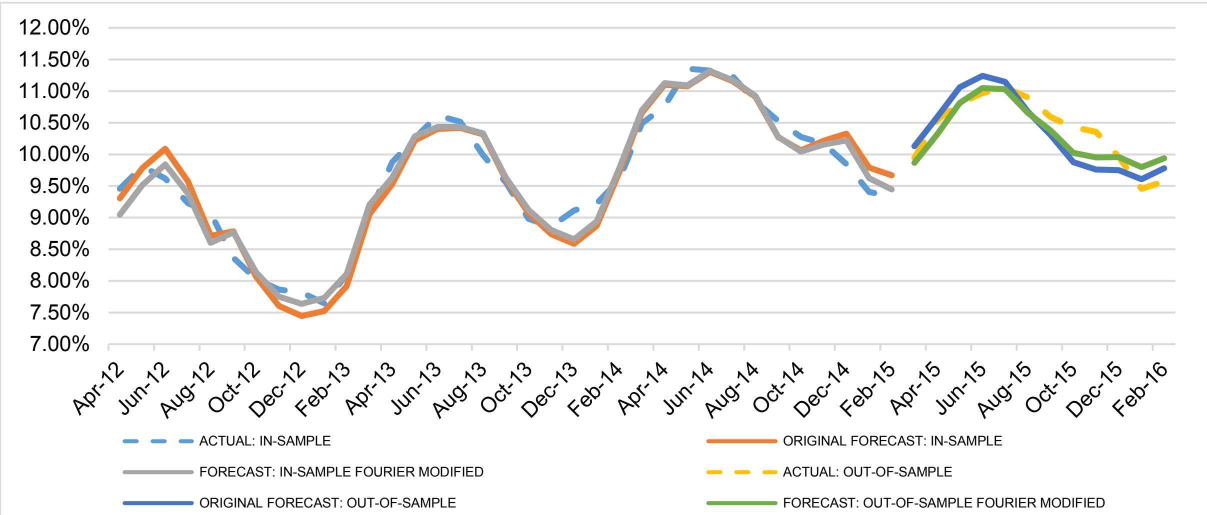


Figure 5.10: Forecasting performance for Fourier series residual modified Model 5: FS_SARIMA(*0,1,2*)(*1,1,0*)$_{12}$



5.4. Concluding remarks

The results discussed in the preceding subsection identified Model 5 as best fitting the data series, which suggests that there is merit in choosing the most adequate model before applying a residual modification method. However, Model 3 has practical advantage in that it overstates the NPL_RATE in more months than understating it, which supports a prudent credit risk strategy. When producing forecasts for the NPL_RATE, the seasonal, unmodified time series model performs well, showing low MAE and RMSE measures in the out-of-sample period. The in-sample performance may be improved by modifying the residual terms with a time series model or a Fourier series algorithm. The out-of-sample performance for the time series residual modified model was poor because the error terms do not display persistence or autocorrelation.

The inclusion of a high number of Fourier coefficients also led to less than satisfactory out-of-sample performance, consistent with the finding of Ludlow and Enders (2000). When the three lowest frequencies were combined to generate a modified forecast, the out-of-sample ability for each model was markedly enhanced. Table 5.7 summarises the in-sample and out-of-sample performance for each model, across the various forecasting techniques.

Table 5.7: Summary of the different forecasting models

Model	**Period**	**Metric**	**Model 2:** SARIMA$(1,1,1)(1,1,1)_{12}$	**Model 3:** SARIMA$(0,1,0)(1,1,0)_{12}$	**Model 5:** SARIMA$(0,1,2)(1,1,0)_{12}$
Original unmodified seasonal ARIMA	In-sample	MAE	0.002345	0.002261	0.002199*
		RMSE	0.002805	0.00274	0.002666*
	Out-of-sample	MAE	0.002964	0.002966	0.002561*
		RMSE	0.003336	0.003408	0.003013*
Time series residual modified	In-sample	MAE	0.002026*	0.002216	0.002042
		RMSE	0.002455*	0.002617	0.002530
	Out-of-sample	MAE	0.002839	0.003395	0.002667*
		RMSE	0.003264	0.003798	0.003189*
Fourier series residual modified (*N*/2 terms)	In-sample	MAE	0.0000000000000001925	0.0000000000000001366*	0.0000000001293
		RMSE	0.0000000000000002674	0.0000000000000001991*	0.0000000002066
	Out-of-sample	MAE	0.002774	0.003073	0.002718*
		RMSE	0.003508	0.003714	0.003149*
Fourier series residual modified (3 terms)	In-sample	MAE	0.002035	0.001920*	0.001954
		RMSE	0.002457	0.002328*	0.002349
	Out-of-sample	MAE	0.002164	0.002214	0.001998*
		RMSE	0.002496	0.003027	0.002498*

Chapter 6

Conclusion and areas for further research

The purpose of this dissertation was to examine the use of a Fourier residual modification technique in forecasting NPLs. An upward trend in balances for the unsecured consumer credit market has been evident over the past few years, which highlights the demand and competitive pressures characterising the industry. NPLs have regulatory and capital implications for credit providers, as governed by the Basel accord. The Greek crisis and collapse of African Bank, discussed chapter one, illustrate the detriment associated with the lack of sound risk management.

A review of existing literature revealed less focus on designing an effective forecasting tool and more emphasis of analysing the determinants of NPLs. In examining the drivers of NPLs in chapter two, both macroeconomic and bank-specific characteristics were found to influence the NPLR. However, there was contrasting results with the impact of explanatory variables on the NPLR. Studies have shown that NPLs tend to persist as previous periods' rates can be used to explain the current period's NPLR. Also, the application of a Fourier series residual modification has shown enhanced forecasting ability across different industries.

This dissertation presented an alternative method to forecasting NPLs rather than the conventional correlation analyses, panel regressions, VAR methods and error correction models. The use of a time series model was justified as NPLs tend to linger in a system and its underlying determinants have conflicting impact. Chapter four illustrated the graphical and statistical evidence of the presence of identifiable seasonality in the data. Hence, the series was first differenced and then seasonally differenced to induce stationarity. The parameters of a seasonal time series model were selected by examining the tentative orders generated from the ESACF. Only the models with statistically significant AR and MA terms, with the lowest information criteria were retained for residual modification. In chapter five, the error terms of three seasonal time series models selected from chapter four were modified. The residual series was first predicted with a time series model and then predicted with a Fourier algorithm. These forecasts were then combined with the original, unmodified seasonal time series forecasts. Both residual modification approaches showed an improvement in the in-sample performance but deterioration in the out-of-sample period, relative to the results of the original models. However, when the three lowest Fourier frequencies were included in the forecast, the out-of-sample performance was markedly better. The analysis concluded that the selection of an adequate model before modifying the residual terms is beneficial. Secondly, a traditional unmodified seasonal time series model performs well, but may be enhanced with a Fourier series residual modification technique that contains a few low frequencies.

This dissertation applied the market data of an unsecured credit provider in South Africa. As such, the results may suffer model specificity bias. In order to test the robustness of the Fourier residual modified technique, the methodology could be carried out on different financial market participants. A limitation to this study was that the credit industry data is not available for a sufficiently long period of time. The NCR contains quarterly data post the NCA, from 2007 onwards for SA. Furthermore, domestic macroeconomic variables such as interest and inflation rates have not varied enough to sufficiently explain movements in NPLs.

Areas for further research include conducting a periodogram and spectral density analysis in the frequency domain to better understand the decomposition of time series data. The use of the FFT and DWT could also be explored. Spectral densities produce efficient information about the dynamic behaviour of time series processes but are not popular in the field of finance and economics, mainly due to sample size limitations. In order to successfully apply frequency domain tools, a large amount of historical data is required. Due to small sample provided by the unsecured lender, a spectral density analysis is of out of the scope of this dissertation. Interested readers with access to larger econometric data sets are directed to Steehouwer (2009) for further detail to frequency domain methods and its applications.

Regulatory requirements stipulate the need for explicit inclusion of historical and forward looking macroeconomic indicators when undertaking exercises in forecasting and stress testing of credit portfolios. This may be modelled with a multivariate time series model, including a causality analysis.

The unsecured credit industry in SA is fast growing as more customers seek access to credit. It is important for lenders to make good credit decisions and to monitor portfolio health through credit risk quality indicators such as the NPLR. Selecting an adequate time series model and modifying its residual terms using a Fourier algorithm could improve forecasting accuracy and assist financial institutions to better provide for bad debts.

Bibliography

1. Abadi, S., Achsani, N. A., Rachmina, D. (2014). The Dynamics of Non-Performing Loan in Indonesian Banking Industry: A Sensitivity Analysis using VECM Approach. *International Journal of Education and Research*. Vol. 2, No. 8, pp. 123-140. Available: http://www.ijern.com/journal/2014/August-2014/13.pdf
2. Abdullah, N., Ahmad, W., Asari, F. F. A. H., Jusoff, K., Latif, N. I. A., Muhamad, N. A. (2011). An Analysis of Non-Performing Loan, Interest Rate and Inflation Rate Using Stata Software. *World Applied Sciences Journal*. Vol. 1, No. 12, pp. 41-48. Available: http://www.idosi.org/wasj/wasj12(BES)11/7.pdf
3. Adjei-Mensah, G. (2014). Executive Compensation, Ownership Structure and Loan Quality of Banks in Ghana. Available: http://ugspace.ug.edu.gh/bitstream/handle/123456789/7469/Gifty%20Adjei-Mensah%20_Executive%20Compensation,%20Ownership%20Structure%20And%20Loan%20Quality%20Of%20Banks%20In%20Ghana_2014.pdf?sequence=1
4. Afshar, N. R., Fahmi, H. (2012). Rainfall Forecasting Using Fourier Series. *Journal of Civil Engineering and Architecture*. Vol. 6, No. 9, pp. 1258-1262.
5. Ahmad, F., Bashir, T. (2013). Explanatory Power of Bank Specific Variables as Determinants of Non-Performing Loans: Evidence from Pakistan Banking Sector. *World Applied Sciences Journal*. Vol. 22, No. 9, pp. 1220-1231. Available: http://idosi.org/wasj/wasj22(9)13/4.pdf
6. Ahmad, N. H., Nor, A. M. (2015). Impaired Financing Determinants of Islamic Banks in Malaysia. *Information Management and Business Review*. Vol. 7, No. 3, pp. 17-25. Available: http://www.ifrnd.org/Research%20Papers/I7(3)2.pdf
7. Akinlo, O., Emmanuel, M. (2014). Determinants of Non-performing Loans in Nigeria. *Accounting & Taxation*. Vol. 6, No. 2, pp. 21-28. Available: http://www.theibfr2.com/RePEc/ibf/acttax/at-v6n2-2014/AT-V6N2-2014-3.pdf

8. Aman, H., Miyazaki, H. (2006). Valuation Effects of New Equity Issues by Banks: Evidence from Japan. Available: http://www.fep.up.pt/conferencias/pfn2006/Conference%20Papers/492.pdf
9. Badar, M., Javid, A. Y. (2013). Impact of Macroeconomic Forces on Nonperforming Loans: An Empirical Study of Commercial Banks in Pakistan. *WSEAS Transactions on Business and Economics*. Vol. 10, No. 1, pp. 40-48. Available: http://www.wseas.org/multimedia/journals/economics/2013/56-259.pdf
10. Bank of England. (2015). Credit Conditions Review 2015 Q4. Available: http://www.bankofengland.co.uk/publications/Documents/creditconditionsreview/2016/ccrq415.pdf
11. Bell, R. D., Herbert, R. D., Lewis, B. G. (2002). The Application of Fourier Analysis to Forecasting the Inbound Call Time Series of a Call Centre. Available: http://www.mssanz.org.au/MODSIM03/Volume_03/B10/06_Lewis.pdf
12. Bloem, A. M., Freeman, R. (2005). The Treatment of Nonperforming Loans. Eighteenth Meeting of the IMF Committee on Balance of Payments Statistics, June 27–July 1, 2005, Washington, D.C. Available: https://www.imf.org/external/pubs/ft/bop/2005/05-29.pdf
13. Bloem, A. M., Gorter, C. N. (2001). The Treatment of Nonperforming Loans in Macroeconomic Statistics, IMF Working Paper, No. 01/209. Available: https://www.imf.org/external/pubs/ft/wp/2001/wp01209.pdf
14. Bofondi, M., Ropele, T. (2011). Macroeconomic determinants of bad loans: evidence from Italian banks, Questioni di Economia e Finanza (Occasional Papers), Bank of Italy, No. 89. Available: https://www.bancaditalia.it/pubblicazioni/qef/2011-0089/QEF_89.pdf
15. Boudriga, A., Taktak, N. B., Jellouli, S. (2009). Banking supervision and nonperforming loans: a cross-country analysis. *Journal of Financial Economic Policy*. Vol. 1, No. 4, pp. 286-318. Available: http://www.bct.gov.tn/bct/siteprod/documents/Conference_Taktak.pdf
16. Brooks, C. (2008). *Introductory Econometrics for Finance*. Reading. Cambridge University Press.

17. Cheang, N. (2009). Early Warning System for Financial Crises. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.493.9683&rep=rep1&type=pdf
18. Chen, P., Hsu, B., Lai, Y., Nguyen, T., Shu, M. (2013). Forecasting with Fourier Residual Modified ARIMA Model- The Case of Air Cargo in Taiwan. International Conference on Technology Innovation and Industrial Management, 29-31 May 2013, Phuket, Thailand. Available: http://www.toknowpress.net/ISBN/978-961-6914-07-9/papers/S5_135-146.pdf
19. Clementina, K., Isu, H. O. (2014). The Rising Incidence of Non-performing Loans and the Nexus of Economic Performance in Nigeria: An investigation. *European Journal of Accounting Auditing and Financial Research*. Vol. 2, No. 5, pp. 87-96. Available: http://www.eajournals.org/wp-content/uploads/The-Rising-Incidence-of-Non-Performing-Loans-and-the-Nexus-of-Economic-Performance-in-Nigeria.pdf
20. Crédit Agricole Consumer Finance. (2013). Consumer Credit Worldwide at year end 2012. Available: https://www.ca-consumerfinance.com/uploads/media/Consumer_Credit_Worldwide_at_year_end_2012.pdf
21. Curak, M., Pepur, S., Poposki, K. (2013). Determinants of non-performing loans – evidence from Southeastern European banking systems. *Banks and Bank Systems*. Vol. 8, No. 1, pp. 45-53. Available: http://businessperspectives.org/journals_free/bbs/2013/BBS_en_2013_01_Curak.pdf
22. Cuza, A. I., Thu, T. N. T. (2012). Basel III and Impacts on Credit Risk Management. Available: http://www.mbf-eu.info/Files/f7992176-6431-4c4a-94c4-b51bcde4dfe7/Essays_NGUYEN%20THI%20THU%20THAO.pdf
23. Dong, Y., Wang, J., Wang, Y., Zhao, G. (2012). Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China. *Energy Policy*. Vol. 48, No. 1, pp. 284-294.
24. Ejiko, O. S., Oladebeye, D. H. (2015). Development of a Fourier Series Forecasting Model for Predicting the Sales Volume of Selected Manufacturing Company. *Industrial Engineering Letters*. Vol. 5, No. 1, pp. 32-40. Available: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=B496EC16021A34737FCDDF75CD036041?doi=10.1.1.672.831&rep=rep1&type=pdf

25. Ekanayake, E. M. N. N., Azeez, A. A. (2015). Determinants of non-performing loans in licensed commercial banks: Evidence from Sri Lanka. *Asian Economic and Financial Review*. Vol. 5, No. 6, pp. 868-882. Available: http://www.aessweb.com/pdf-files/aefr-2015-5(6)-868-882.pdf
26. Espinoza, R., Prasad, A. (2010). Non-performing Loans in the GCC Banking System and their Macroeconomic Effects, IMF Working Paper, No. 10/224.
27. Federal Bank of Minneapolis. (2015). Consumer Credit Conditions in the Ninth District. Available: https://www.minneapolisfed.org/community/community-development/credit/delinquencies
28. FRB. (2007). FOMC: Transcripts and Other Historical Materials, 2007. Available: https://www.federalreserve.gov/monetarypolicy/files/FOMC20070509material.pdf
29. Fofack, H. (2005). Nonperforming loans in Sub-Saharan Africa: causal analysis and macroeconomic implications, World Bank Policy Research Working Paper, No. 3769. Available: https://core.ac.uk/download/files/153/6615276.pdf
30. Fumi, A., Pepe, A., Scarabotti, L., Schiraldi, M. M. (2013). Fourier Analysis for Demand Forecasting in a Fashion Company. *International Journal of Engineering Business Management*. Special Issue on Innovations in Fashion Industry. Vol. 5, No. 30, pp. 1-10. Available: http://cdn.intechopen.com/pdfs/45558.pdf
31. Gerbing, D. W. (2016). Time Series Components. Available: http://web.pdx.edu/~gerbing/515/Resources/ts.pdf
32. Giakoumis, M. (2014). NPLs: The Achilles heel of the Greek banking system. Available: http://www.macropolis.gr/?i=portal.en.the-agora.1375
33. Government of Japan. (2001). Annual Report on Japan's Economy and Public Finance. Cabinet Office, 4 December 2001. Available: http://www5.cao.go.jp/keizai3/2001/1204wp-keizai/summary.pdf
34. Greenidge, K., Grosvenor, T. (2010). Forecasting non-performing loans in Barbados. *Journal of Business, Finance and Economics in Emerging Economies*. Vol. 5, No. 1, pp. 79-108. Available: http://www.ccmf-uwi.org/files/publications/journal/2010_1_5/79_108.pdf

35. Gumata, N., Klein, N., and Ndou, E. (2012). A Financial Conditions Index for South Africa. IMF Working Paper, No. 12/196. Available: https://www.imf.org/external/pubs/ft/wp/2012/wp12196.pdf
36. Harrison, V., Liakos, C. (2015). Greece defaults on $1.7 billion IMF payment. Available: http://money.cnn.com/2015/06/30/news/economy/greece-imf-default/
37. Hsu, B., Hung, W., Lu, C., Nguyen, T., Shu, M. (2014). Forecasting with Fourier Residual Modified ARIMA Model- An Empirical Case of Inbound Tourism Demand in New Zealand. *WSEAS Transactions on Mathematics*. Vol. 13, No. 1, pp. 12-21. Available: http://www.wseas.org/multimedia/journals/mathematics/2014/c045706-310.pdf
38. Hsu, B., Huang, Y., Nguyen, T., Shu, M. (2013). Accurate forecasting models in predicting the inbound tourism demand in Vietnam. *Journal of Statistics and Management Systems*. Vol. 16, No. 1, pp. 25-43.
39. Inaba, N., Kozu, T., Sekine, T. (2005). Non-performing loans and the real economy: Japan’s experience. BIS Papers, No. 22, part 7. Available: http://www.bis.org/publ/bppdf/bispap22g.pdf
40. Islam, M. S., Shil, N. C., Mannan, M. A. (2008). Non performing loans – its causes, consequences and some learning. MPRA Paper No. 7708. Available: https://mpra.ub.uni-muenchen.de/7708/1/Non_Performing_Loans-Its_causes_consequences_and_some_l.pdf
41. Jeon, B. N. (2010). From the 1997-98 Asian Financial Crisis to the 2008-09 Global Economic Crisis: Lessons from Korea’s Experience. *East Asia Law Review*. Vol. 5, No. 1, pp. 103-154. Available: http://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1041&context=ealr
42. Joseph, M.T., Edson, G., Manuere, F., Clifford, M., Michael, K. (2012), Non Performing Loans in Commercial Banks: A Case Of CBZ Bank Limited In Zimbabwe. *Interdisciplinary Journal of Contemporary Research In Business*. Vol. 4, No. 7. pp, Vol. 4.

43. Khemraj, T., Pasha, S. (2009). The determinants of non-performing loans: An econometric case study of Guyana. The Caribbean Centre for Banking and Finance Bi-annual Conference on Banking and Finance, St. Augustine, Trinidad.
44. Klein, N. (2013). Non-Performing Loans in CESEE: Determinants and Impact on Macroeconomic Performance. IMF Working Paper, No. 13/72. Available: https://www.imf.org/external/pubs/ft/wp/2013/wp1372.pdf
45. Koutras, A. (2015). The Greek NPL issue and a possible resolution path. Available: http://greekeconomistsforreform.com/wp-content/uploads/The-Greek-NPL-issue-and-a-possible-resolution_v1.pdf
46. Legae Securities. (2010). African Bank Inv. Ltd. Non-retail deposits taking strategy clouds long term growth outlook. Available: https://www.scribd.com/document/32754614/ABIL-Non-Retail-Deposit-Taking-Strategy-Clouds-Long-Term-Growth-Outlook-Initiating-With-a-HOLD
47. Levenbach, H. (2015). Dealing with Seasonal Influences and Seasonal Adjustments. Available: http://cpdftraining.org/downloads/Levenbach_Seasonality2015.pdf
48. Li, F., Zou, Y. (2014). The Impact of Credit Risk Management on Profitability of Commercial Banks: A Study of Europe. Umea School of Business and Economics. Available: https://www.diva-portal.org/smash/get/diva2:743402/FULLTEXT01.pdf
49. Ludlow, J., Enders, W. (2000). Estimating non-linear ARMA models using Fourier coefficients. *International Journal of Forecasting*. Vol. 16, No. 3, pp. 333-347.
50. Makri, V., Tsagkanos, A., Bellas, A. (2014). Determinants of Non-Performing Loans: The Case of Eurozone. Panoeconomicus. Vol. 1, No. 2, pp. 193-206. Available: http://www.doiserbia.nb.rs/img/doi/1452-595X/2014/1452-595X1402193M.pdf
51. Mendoza, E. G., Terrones, M. E. (2012). An anatomy of credit booms and their demise. National Bureau of Economic Research, Working Paper 18379. Available: http://www.sas.upenn.edu/~egme/wp/w18379.pdf
52. Meng, M., Niu, D., Sun, W. (2011). Forecasting Monthly Electric Energy Consumption Using Feature Extraction. *Energies*. Vol. 1, No. 4, pp. 1495-1507.

53. Messai, A. S., Jouini, F. (2013). Micro and Macro Determinants of Non-performing Loans. *International Journal of Economics and Financial Issues*. Vol. 3, No. 4, pp. 852-860. Available: http://www.econjournals.com/index.php/ijefi/article/viewFile/517/pdf
54. Mesnard, B., Margerit, A., Power, C., Magnus, M. (2016). Non-performing loans in the Banking Union: stocktaking and challenges. Economic Governance Support Unit, European Parliament Briefing. 18 March 2016. Available: http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/574400/IPOL_BRI(2016)574400_EN.pdf
55. Mileris, R. (2012). Macroeconomic Determinants of Loan Portfolio Credit Risk in Banks. *Inzinerine Ekonomika-Engineering Economics*. Vol. 23, No. 5, pp. 496-504. Available: http://www.eejournal.ktu.lt/index.php/EE/article/viewFile/1890/2218
56. Moody's Investors Service. (2011). Banking Account & Ratio Definitions. Available: https://www.moodys.com/sites/products/ProductAttachments/Banking%20Account%20and%20Ratio%20Definitions.pdf
57. Mukoki, P. G. V., Mapfumo, A. (2015). The Effect of Dollarization on the Growth of Non-Performing Loans in the Zimbabwe Banking System: An Autoregressive Distributed Lag (ARDL) Bound Test Approach. *Journal of Economics and Sustainable Development.* Vol. 6, No. 10, pp. 82-92. Available: http://iiste.org/Journals/index.php/JEDS/article/viewFile/22672/23326
58. Mwengei, K. B. (2013). Assessing the Factors Contributing to Non – Performance Loans in Kenyan Banks. *European Journal of Business and Management*. Vol. 5, No. 32, pp. 155-162. Available: http://www.iiste.org/Journals/index.php/EJBM/article/viewFile/9576/9702
59. NCR (2016). Consumer Credit Market Report. Available: http://www.ncr.org.za/index.php?option=com_content&view=article&id=42
60. Nelson, R. M., Belkin, P., Mix, D. E. (2011). Greece's Debt Crisis: Overview, Policy Responses, and Implications. Congressional Research Service. Available: https://www.fas.org/sgp/crs/row/R41167.pdf

61. Nualsri, A., Roengpitya, R., Sabborriboon, W., Thanavibul, N. (2015). Forecasting the Quality of Corporate and Consumer Loans in the Thai Banking Sector: Methodology and Policy Implications. Bank of Thailand, Discussion Paper. Available: https://www.bot.or.th/Thai/MonetaryPolicy/ArticleAndResearch/DiscussionPaper/DP12015.pdf
62. Omekara, C. O., Ekpenyong, E. J., Ekerete, M. P. (2013). Modeling the Nigerian Inflation Rates Using Periodogram and Fourier Series Analysis. *CBN Journal of Applied Statistics*. Vol. 4, No. 2, pp. 51-68. Available: https://www.cbn.gov.ng/out/2014/sd/modeling%20the%20nigerian%20inflation%20rates%20using%20periodogram%20and%20fourier%20series%20analysis.pdf
63. Podpiera, R. (2006). Does Compliance with Basel Core Principles Bring Any Measurable Benefits? IMF Staff Papers. Vol. 53, No. 2, pp. 306-326. Available: https://www.imf.org/External/Pubs/FT/staffp/2006/02/pdf/podpiera.pdf
64. Prasanna, P. K. (2014). Determinants of Non-Performing Loans in Indian Banking System. 3rd International Conference on Management, Behavioural Sciences and Economics Issues 11-12 February 2014, Singapore. Available: http://psrcentre.org/images/extraimages/27%20214306.pdf
65. PWC. (2015). Navigating a volatile landscape, Major banks analysis – South Africa. Available: https://www.pwc.co.za/en/assets/pdf/bank-analysis-march-2015.pdf
66. Rajput, N., Arora, A. P., Kaur, B. (2011). Non-performing assets in the Indian public sector banks: an analytical study. *Bank and Bank Systems*. Vol. 6, No. 4, pp. 84-89. Available: http://businessperspectives.org/journals_free/bbs/2011/BBS_en_2011_04_Rajput.pdf
67. Saba, I., Kouser, R., Azeem, M. (2012). Determinants of Non Performing Loans: Case of US Banking Sector. *The Romanian Economic Journal*. Vol. 15, No. 44, pp. 141-152. Available: http://www.rejournal.eu/sites/rejournal.versatech.ro/files/issues/2012-06-02/556/15-determinantsofnon-performingloanscaseofusbankingsector.pdf

68. Sanchez, D. (2014). Why It Failed: African Bank Gave Credit To The Poor. Available: http://afkinsider.com/70304/failed-s-african-bank-gave-credit-to-the-poor/
69. SARB. (2014). Remarks by the Governor of the South African Reserve Bank, Gill Marcus. Press Conference, 10 August 2014: African Bank Limited. Available: https://www.resbank.co.za/Lists/Speeches/Attachments/414/Governor's%20Address%20-%20ABIL.pdf
70. Serwa, D. (2013). Measuring Non-Performing Loans During (and After) Credit Booms. *Central European Journal of Economic Modelling and Econometrics*. Vol. 1, No. 5, pp. 163-183. Available: http://cejeme.eu/publishedarticles/2013-15-19-635230665380156250-4784.pdf
71. Shingjergji, A. (2013). An Analysis of the Nonperforming Loans in the Albanian Banking System. *International Journal of Business and Commerce*. Vol. 2, No. 6, pp. 1-11. Available: http://www.ijbcnet.com/2-6/IJBC-13-2602.pdf
72. Škarica, B. (2014). Determinants of non-performing loans in Central and Eastern European countries. *Financial Theory and Practice*. Vol. 1, No. 38, pp. 37-59.
73. Steehouwer, H. (2009). A Frequency Domain Methodology for Time Series Modelling. Methodological Working Paper No. 2008-02. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2229313
74. Stolojescu, C. (2011). A Wavelets Based Approach for Time Series Mining. Available: http://cmpicsu.utt.ro/cercetare/CNCSIS_Idei/TezaOctombrie.pdf
75. Tsaur, R., Kuo, T. (2013). Tourism Demand Forecasting Using a Novel High-Precision Fuzzy Time Series Model. International Journal of Innovative Computing, Information and Control. Vol. 10, No. 2, pp. 695-701. Available: http://tkuir.lib.tku.edu.tw/dspace/bitstream/987654321/99125/1/Tourism%20demand%20forecasting%20using%20a%20novel%20high-precision%20fuzzy%20time%20series%20model.pdf
76. Transunion. (2015). Consumer credit health continued to improve in the 1st quarter of 2015 – Further moderate relief for consumers. Available: http://www.transunion.co.za/abouttudocs/newsroom/Consumer%20credit%20

health%20continued%20to%20improve%20in%20the%201st%20quarter%20of%202015.pdf

77. Viswanadham, N., Nahid, B. (2015). Determinants of Non Performing Loans in Commercial Banks: A Study of NBC Bank Dodoma Tanzania. *International Journal of Finance & Banking Studies*. Vol. 4, No. 1, pp. 70-94.
78. Watkins, J. C. (2011). Maximum Likelihood Estimation. Available: http://math.arizona.edu/~jwatkins/o-mle.pdf
79. Wei, W. W. S. (2005). *Time Series Analysis Univariate and Multivariate Methods*. Philadelphia. Pearson.
80. Weisstein, E. (2016). Fourier Series from Wolfram MathWorld. Available: http://mathworld.wolfram.com/FourierSeries.html
81. Westernhagen, N., Harada, E., Nagata, T., Vale, B., Ayuso, J., Saurina, J., Daltung, S., Ziegler, S., Kent, E., Reidhill, J., Peristiani, S. (2004). Bank Failures in Mature Economies. Basel Committee on Banking Supervision. Working Paper No. 13. Available: http://www.bis.org/publ/bcbs_wp13.pdf
82. Williams, F. (2014). Seasonal debt cycle can make costs grow. Available: http://blogs.creditcards.com/2014/05/test-for-chart.php
83. World Bank. (2015). Bank nonperforming loans to total gross loans %. Available: http://data.worldbank.org/indicator/FB.AST.NPER.ZS

Appendix A

Figure A1: Combined seasonality test flowchart

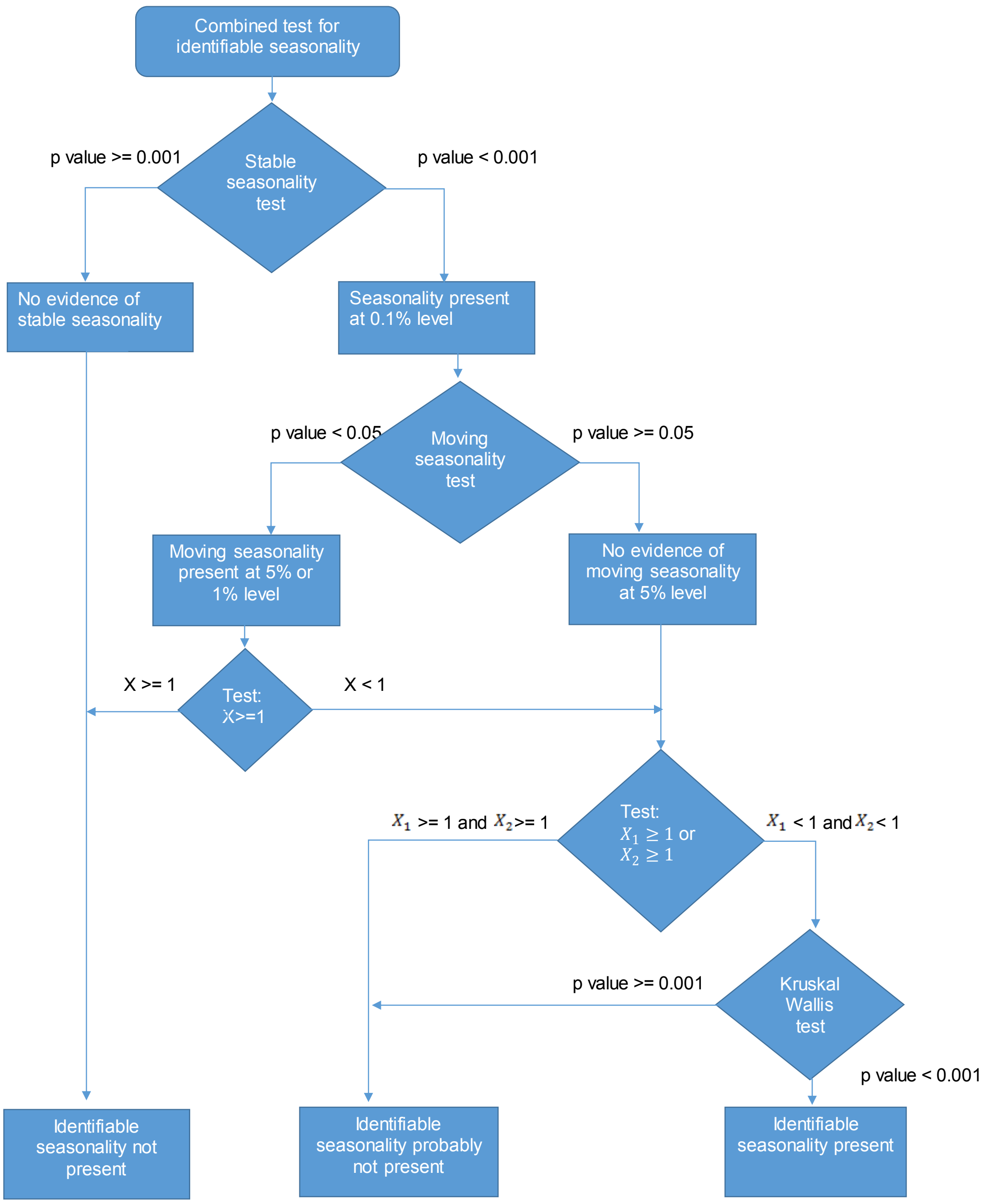
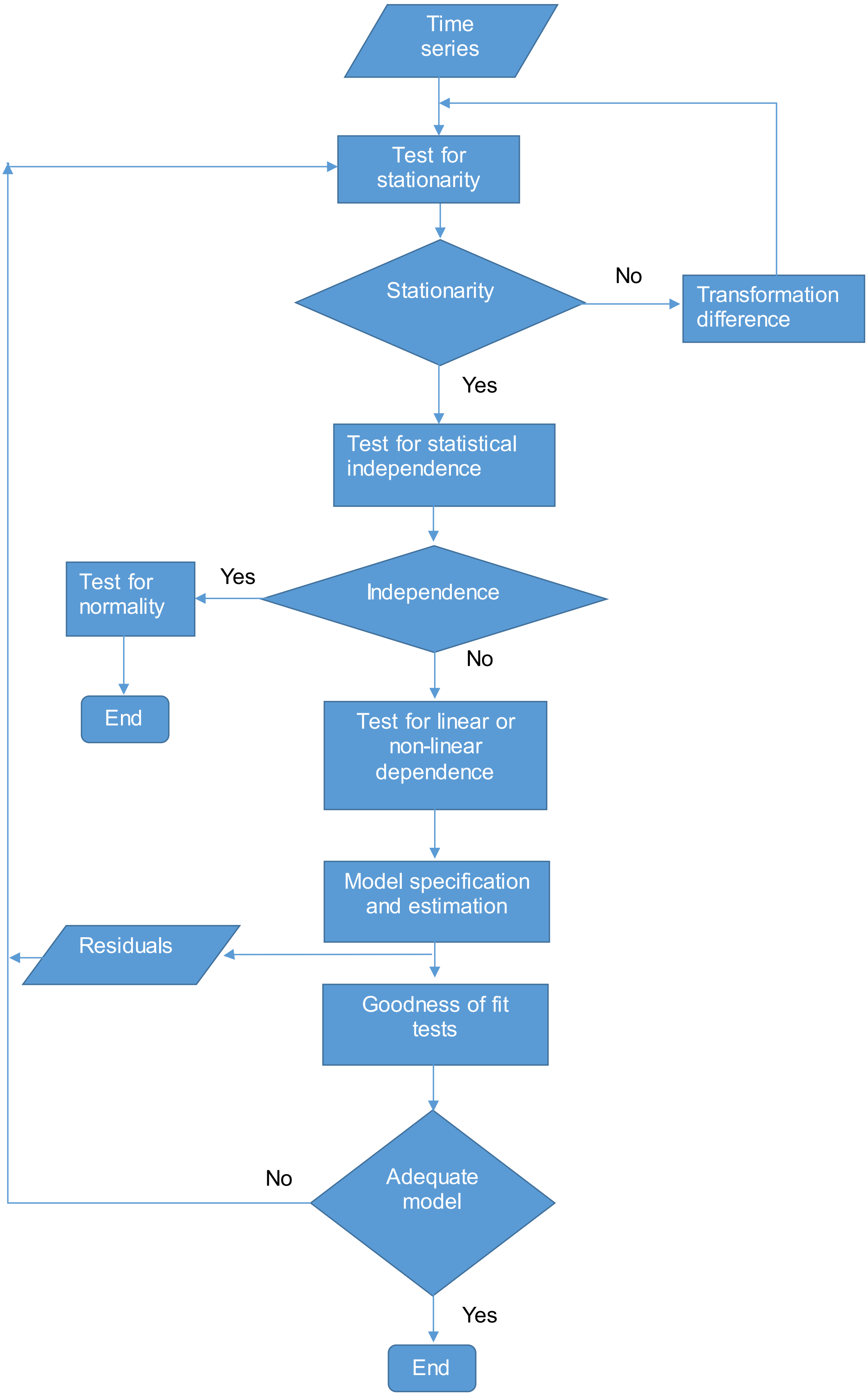


Figure A2: Box-Jenkin's modelling approach



Appendix B

Table B1: PP unit root tests for the original series

Phillips-Perron Unit Root Tests					
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	0	0.0079	0.6811	0.02	0.6872
	1	-0.0249	0.6737	-0.06	0.6590
	2	-0.0455	0.6691	-0.10	0.6458
	3	-0.0546	0.6670	-0.11	0.6406
	4	-0.0533	0.6673	-0.11	0.6413
	5	-0.0433	0.6695	-0.10	0.6470
	6	-0.0273	0.6731	-0.07	0.6572
	7	-0.0084	0.6774	-0.02	0.6716
	8	0.0098	0.6815	0.03	0.6892
	9	0.0249	0.6850	0.10	0.7096
	10	0.0351	0.6873	0.16	0.7296
	11	0.0390	0.6881	0.20	0.7396
	12	0.0373	0.6878	0.18	0.7350
Single Mean	0	-5.2719	0.3947	-1.67	0.4409
	1	-8.4407	0.1785	-2.09	0.2491
	2	-10.4810	0.1042	-2.32	0.1686
	3	-11.4521	0.0802	-2.42	0.1396
	4	-11.4517	0.0802	-2.42	0.1396
	5	-10.6282	0.1001	-2.34	0.1639
	6	-9.2315	0.1451	-2.18	0.2144
	7	-7.5490	0.2247	-1.98	0.2941
	8	-5.9248	0.3374	-1.77	0.3941
	9	-4.5703	0.4644	-1.56	0.4952
	10	-3.6672	0.5659	-1.41	0.5703
	11	-3.3493	0.6044	-1.36	0.5978
	12	-3.5410	0.5811	-1.39	0.5812
Trend	0	-6.7957	0.6601	-1.75	0.7164
	1	-11.5273	0.3028	-2.33	0.4121
	2	-14.5510	0.1634	-2.63	0.2679
	3	-15.9531	0.1200	-2.76	0.2167
	4	-15.8897	0.1217	-2.76	0.2188
	5	-14.5890	0.1621	-2.64	0.2664
	6	-12.4236	0.2542	-2.42	0.3642
	7	-9.8280	0.4135	-2.14	0.5141
	8	-7.3201	0.6147	-1.82	0.6814
	9	-5.2227	0.7926	-1.51	0.8146
	10	-3.8118	0.8912	-1.26	0.8879
	11	-3.2864	0.9196	-1.15	0.9107
	12	-3.5304	0.9070	-1.20	0.9004

Table B2: PP unit root tests for the first differenced series

Phillips-Perron Unit Root Tests					
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	0	-20.5489	0.0008	-3.54	0.0006
	1	-22.9798	0.0003	-3.70	0.0004
	2	-24.0739	0.0002	-3.77	0.0003
	3	-24.1555	0.0002	-3.77	0.0003
	4	-23.3047	0.0003	-3.72	0.0003
	5	-21.2118	0.0006	-3.58	0.0005
	6	-18.4239	0.0017	-3.39	0.0010
	7	-15.0102	0.0052	-3.16	0.0021
	8	-12.1453	0.0126	-2.96	0.0037
	9	-9.6230	0.0277	-2.81	0.0057
	10	-7.5783	0.0528	-2.76	0.0066
	11	-6.8222	0.0672	-2.79	0.0061
	12	-7.1292	0.0609	-2.77	0.0064
Single Mean	0	-20.5455	0.0057	-3.51	0.0110
	1	-22.9745	0.0027	-3.67	0.0070
	2	-24.0669	0.0019	-3.74	0.0057
	3	-24.1478	0.0018	-3.74	0.0056
	4	-23.2969	0.0024	-3.69	0.0066
	5	-21.2039	0.0046	-3.55	0.0097
	6	-18.4164	0.0108	-3.36	0.0164
	7	-15.0035	0.0296	-3.12	0.0306
	8	-12.1404	0.0662	-2.91	0.0497
	9	-9.6203	0.1307	-2.76	0.0710
	10	-7.5782	0.2228	-2.69	0.0821
	11	-6.8246	0.2698	-2.71	0.0789
	12	-7.1339	0.2495	-2.69	0.0811
Trend	0	-20.5027	0.0404	-3.46	0.0533
	1	-22.9363	0.0216	-3.63	0.0361
	2	-24.0268	0.0162	-3.70	0.0302
	3	-24.1016	0.0158	-3.70	0.0299
	4	-23.2461	0.0199	-3.65	0.0343
	5	-21.1465	0.0343	-3.51	0.0481
	6	-18.3525	0.0683	-3.31	0.0747
	7	-14.9306	0.1499	-3.06	0.1255
	8	-12.0617	0.2723	-2.85	0.1872
	9	-9.5376	0.4341	-2.67	0.2515
	10	-7.4902	0.5997	-2.58	0.2894
	11	-6.7317	0.6652	-2.59	0.2872
	12	-7.0392	0.6386	-2.58	0.2901

Table B3: PP unit root tests for the first differenced, seasonally differenced series

Phillips-Perron Unit Root Tests					
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	0	-40.2222	<.0001	-5.90	<.0001
	1	-39.7102	<.0001	-5.89	<.0001
	2	-42.5790	<.0001	-5.95	<.0001
	3	-43.8452	<.0001	-5.98	<.0001
	4	-43.1488	<.0001	-5.96	<.0001
	5	-42.0543	<.0001	-5.94	<.0001
	6	-41.4104	<.0001	-5.93	<.0001
	7	-40.1588	<.0001	-5.90	<.0001
	8	-40.1319	<.0001	-5.90	<.0001
	9	-39.7037	<.0001	-5.89	<.0001
	10	-39.4065	<.0001	-5.89	<.0001
	11	-39.8688	<.0001	-5.90	<.0001
	12	-38.7930	<.0001	-5.88	<.0001
Single Mean	0	-40.2244	0.0004	-5.84	0.0001
	1	-39.7150	0.0004	-5.83	0.0001
	2	-42.5812	0.0004	-5.89	0.0001
	3	-43.8408	0.0004	-5.92	0.0001
	4	-43.1375	0.0004	-5.90	0.0001
	5	-42.0338	0.0004	-5.88	0.0001
	6	-41.3799	0.0004	-5.86	0.0001
	7	-40.1190	0.0004	-5.83	0.0001
	8	-40.0846	0.0004	-5.83	0.0001
	9	-39.6516	0.0004	-5.83	0.0001
	10	-39.3511	0.0004	-5.82	0.0001
	11	-39.8119	0.0004	-5.83	0.0001
	12	-38.7363	0.0004	-5.81	0.0001
Trend	0	-40.3695	<.0001	-5.77	0.0001
	1	-39.8894	<.0001	-5.76	0.0001
	2	-42.7297	<.0001	-5.83	0.0001
	3	-43.9161	<.0001	-5.86	0.0001
	4	-43.1131	<.0001	-5.84	0.0001
	5	-41.8745	<.0001	-5.81	0.0001
	6	-41.0735	<.0001	-5.79	0.0001
	7	-39.6533	<.0001	-5.76	0.0001
	8	-39.4613	<.0001	-5.75	0.0001
	9	-38.8645	<.0001	-5.74	0.0001
	10	-38.4086	<.0001	-5.73	0.0001
	11	-38.7127	<.0001	-5.74	0.0001
	12	-37.4781	0.0001	-5.72	0.0001

Figure B1: Correlation plots for the first differenced, seasonally differenced NPL_RATE

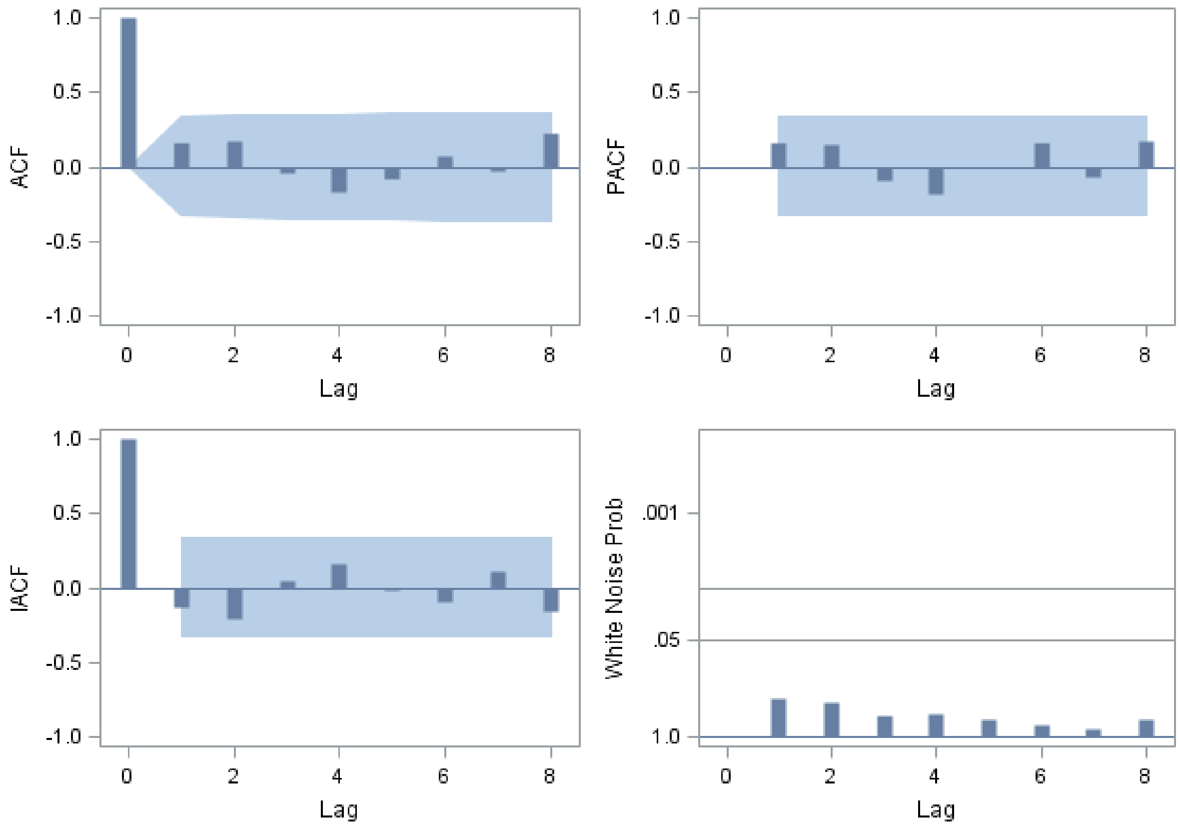


Table B4: Tentative order selection tests for the NPL_RATE non-stationary and stationary series

ARMA(p+d,q) Tentative Order Selection Tests	
ESACF	
p+d	q
2	0
4	1
8	3
0	5
12	6
13	0

ARMA(p+d,q) Tentative Order Selection Tests	
ESACF	
p+d	q
0	0
6	0
3	2
7	0
4	2
5	2
8	0
10	1
12	0
13	0

Table B5: Model results for SARIMA$(3,1,2)(0,1,0)_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.33065	0.17322	-1.91	0.0563	2
AR1,1	-0.03172	0.19926	-0.16	0.8735	3

Table B6: Model results for SARIMA*(4,1,2)(0,1,0)*$_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.26790	0.18437	-1.45	0.1462	2
AR1,1	-0.24436	0.20287	-1.20	0.2284	4

Figure B7.1: DW statistics for Model 1: SARIMA(0,1,0)(0,1,0)$_{12}$

Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	1.6791	0.1251	0.8749
2	1.5574	0.0910	0.9090
3	1.8519	0.4008	0.5992
4	2.0830	0.7351	0.2649
5	1.8397	0.5367	0.4633
6	1.4587	0.1903	0.8097
7	1.6611	0.4761	0.5239
8	1.1169	0.0544	0.9456
9	1.6301	0.5956	0.4044
10	1.3558	0.3384	0.6616
11	1.2267	0.2618	0.7382
12	2.0109	0.9823	0.0177

Table B7.2: ARCH test for Model 1: SARIMA*(0,1,0)(0,1,0)*$_{12}$

Tests for ARCH Disturbances Based on OLS Residuals				
Order	Q	Pr > Q	LM	Pr > LM
1	1.5068	0.2196	0.7521	0.3858
2	1.5277	0.4659	0.7558	0.6853
3	1.8300	0.6084	1.0216	0.7960
4	2.8104	0.5900	1.7186	0.7873
5	6.6752	0.2459	5.3047	0.3798
6	6.8013	0.3396	5.6241	0.4666
7	7.5042	0.3783	6.0347	0.5357
8	7.9656	0.4368	6.3202	0.6114
9	10.3667	0.3216	6.9630	0.6410
10	10.5189	0.3962	7.2222	0.7043
11	10.5236	0.4840	7.2326	0.7799
12	13.3686	0.3428	7.8496	0.7968

Table B7.3: PP unit root test for Model 1: SARIMA*(0,1,0)(0,1,0)*$_{12}$

Phillips-Perron Unit Root Tests					
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	0	-28.1034	<.0001	-4.68	<.0001
	1	-27.6121	<.0001	-4.67	<.0001
	2	-29.7089	<.0001	-4.74	<.0001
	3	-30.5491	<.0001	-4.77	<.0001
	4	-29.8343	<.0001	-4.75	<.0001
	5	-29.0977	<.0001	-4.72	<.0001
	6	-28.9650	<.0001	-4.71	<.0001
	7	-28.4507	<.0001	-4.70	<.0001
	8	-28.9902	<.0001	-4.72	<.0001
	9	-28.9910	<.0001	-4.72	<.0001
	10	-29.1091	<.0001	-4.72	<.0001
	11	-29.5138	<.0001	-4.73	<.0001
	12	-28.8031	<.0001	-4.71	<.0001
Single Mean	0	-28.0747	0.0002	-4.61	0.0007
	1	-27.5607	0.0002	-4.59	0.0008
	2	-29.6868	0.0002	-4.67	0.0006
	3	-30.5974	0.0002	-4.71	0.0006
	4	-29.9627	0.0002	-4.68	0.0006
	5	-29.3172	0.0002	-4.66	0.0007
	6	-29.2766	0.0002	-4.66	0.0007
	7	-28.8488	0.0002	-4.64	0.0007
	8	-29.4668	0.0002	-4.66	0.0006
	9	-29.5426	0.0002	-4.67	0.0006
	10	-29.7272	0.0002	-4.67	0.0006
	11	-30.1991	0.0002	-4.69	0.0006
	12	-29.5516	0.0002	-4.67	0.0006
Trend	0	-28.9719	0.0014	-4.70	0.0033
	1	-28.5798	0.0016	-4.69	0.0034
	2	-30.5308	0.0007	-4.76	0.0029
	3	-31.1984	0.0005	-4.78	0.0027
	4	-30.3160	0.0008	-4.75	0.0029
	5	-29.2802	0.0012	-4.71	0.0032
	6	-28.7720	0.0015	-4.69	0.0034
	7	-27.8489	0.0022	-4.66	0.0036
	8	-27.9551	0.0021	-4.67	0.0036
	9	-27.5712	0.0024	-4.65	0.0037
	10	-27.3170	0.0026	-4.65	0.0038
	11	-27.3847	0.0026	-4.65	0.0038
	12	-26.4024	0.0037	-4.62	0.0040

Table B8.1: DW statistics for Model 2: SARIMA$(1,1,1)(1,1,1)_{12}$

Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	1.7985	0.2154	0.7846
2	1.5864	0.1065	0.8935
3	1.8853	0.4397	0.5603
4	2.0781	0.7303	0.2697
5	1.8531	0.5526	0.4474
6	1.4572	0.1890	0.8110
7	1.7198	0.5472	0.4528
8	1.1141	0.0534	0.9466
9	1.6466	0.6152	0.3848
10	1.3044	0.2798	0.7202
11	1.1516	0.1861	0.8139
12	1.9766	0.9767	0.0233

Table B8.2: ARCH test for Model 2: SARIMA$(1,1,1)(1,1,1)_{12}$

Tests for ARCH Disturbances Based on OLS Residuals				
Order	Q	Pr > Q	LM	Pr > LM
1	2.2251	0.1358	1.2095	0.2714
2	2.2864	0.3188	1.2235	0.5424
3	2.8405	0.4169	1.5377	0.6736
4	3.3739	0.4973	1.8676	0.7601
5	6.6349	0.2492	4.9416	0.4230
6	6.8628	0.3337	5.4898	0.4827
7	7.4677	0.3819	6.2628	0.5094
8	7.9065	0.4427	6.7498	0.5639
9	9.0145	0.4359	6.8264	0.6552
10	9.0418	0.5281	6.8464	0.7399
11	9.0430	0.6179	6.8529	0.8109
12	10.7727	0.5485	7.0589	0.8537

Table B8.3: PP unit root test for Model 2: SARIMA$(1,1,1)(1,1,1)_{12}$

Phillips-Perron Unit Root Tests					
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	0	-30.5706	<.0001	-5.06	<.0001
	1	-30.2829	<.0001	-5.06	<.0001
	2	-32.4529	<.0001	-5.11	<.0001
	3	-33.2275	<.0001	-5.14	<.0001
	4	-32.4839	<.0001	-5.12	<.0001
	5	-31.7337	<.0001	-5.09	<.0001
	6	-31.5483	<.0001	-5.09	<.0001
	7	-30.9444	<.0001	-5.07	<.0001
	8	-31.3663	<.0001	-5.08	<.0001
	9	-31.2707	<.0001	-5.08	<.0001
	10	-31.3480	<.0001	-5.08	<.0001
	11	-31.7428	<.0001	-5.09	<.0001
	12	-31.0759	<.0001	-5.08	<.0001
Single Mean	0	-30.5458	0.0002	-4.98	0.0003
	1	-30.2488	0.0002	-4.97	0.0003
	2	-32.4399	0.0002	-5.03	0.0003
	3	-33.2489	0.0002	-5.06	0.0003
	4	-32.5402	0.0002	-5.04	0.0003
	5	-31.8295	0.0002	-5.01	0.0003
	6	-31.6856	0.0002	-5.01	0.0003
	7	-31.1198	0.0002	-4.99	0.0003
	8	-31.5796	0.0002	-5.01	0.0003
	9	-31.5190	0.0002	-5.01	0.0003
	10	-31.6283	0.0002	-5.01	0.0003
	11	-32.0562	0.0002	-5.02	0.0003
	12	-31.4179	0.0002	-5.00	0.0003
Trend	0	-31.1226	0.0006	-5.01	0.0015
	1	-30.9141	0.0006	-5.01	0.0015
	2	-32.9637	0.0002	-5.06	0.0013
	3	-33.5816	0.0002	-5.08	0.0012
	4	-32.6854	0.0003	-5.06	0.0013
	5	-31.6619	0.0004	-5.03	0.0014
	6	-31.1366	0.0006	-5.01	0.0015
	7	-30.1650	0.0009	-4.99	0.0016
	8	-30.2089	0.0008	-4.99	0.0016
	9	-29.7821	0.0010	-4.98	0.0016
	10	-29.5492	0.0011	-4.98	0.0016
	11	-29.6711	0.0010	-4.98	0.0016
	12	-28.7941	0.0015	-4.96	0.0017

Table B9.1: DW statistics for Model 3: SARIMA$(0,1,0)(1,1,0)_{12}$

Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	1.4123	0.0236	0.9764
2	1.4938	0.0626	0.9374
3	2.0414	0.6235	0.3765
4	1.9558	0.5980	0.4020
5	1.9798	0.6953	0.3047
6	1.4513	0.1840	0.8160
7	1.7355	0.5658	0.4342
8	1.1652	0.0751	0.9249
9	1.4627	0.3906	0.6094
10	1.2433	0.2165	0.7835
11	1.2488	0.2864	0.7136
12	1.6543	0.8291	0.1709

Table B9.2: ARCH test for Model 3: SARIMA*(0,1,0)(1,1,0)*$_{12}$

Tests for ARCH Disturbances Based on OLS Residuals				
Order	Q	Pr > Q	LM	Pr > LM
1	0.3882	0.5332	0.1755	0.6753
2	0.6976	0.7055	0.2501	0.8825
3	1.4143	0.7022	0.6257	0.8905
4	1.5437	0.8189	0.7989	0.9386
5	2.6167	0.7588	1.7374	0.8841
6	3.0775	0.7991	3.6544	0.7233
7	3.3269	0.8532	4.5118	0.7193
8	3.5569	0.8947	4.8298	0.7756
9	3.8145	0.9232	4.9166	0.8415
10	4.0705	0.9441	5.1766	0.8791
11	4.5679	0.9503	6.1890	0.8605
12	4.6208	0.9695	6.3400	0.8980

Table B9.3: PP unit root test for Model 3: SARIMA*(0,1,0)(1,1,0)*$_{12}$

Phillips-Perron Unit Root Tests					
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	0	-24.3928	<.0001	-4.09	0.0002
	1	-23.7075	0.0001	-4.06	0.0002
	2	-26.0909	<.0001	-4.18	0.0001
	3	-25.9365	<.0001	-4.17	0.0001
	4	-25.3151	<.0001	-4.14	0.0001
	5	-24.1797	0.0001	-4.08	0.0002
	6	-23.8245	0.0001	-4.06	0.0002
	7	-22.7608	0.0002	-4.01	0.0002
	8	-22.7032	0.0002	-4.01	0.0002
	9	-22.3672	0.0002	-3.99	0.0002
	10	-22.2732	0.0003	-3.98	0.0002
	11	-22.4459	0.0002	-3.99	0.0002
	12	-21.9953	0.0003	-3.97	0.0002
Single Mean	0	-24.4539	0.0008	-4.01	0.0036
	1	-23.7767	0.0011	-3.97	0.0040
	2	-26.1767	0.0004	-4.10	0.0028
	3	-25.9848	0.0004	-4.09	0.0029
	4	-25.3115	0.0006	-4.06	0.0032
	5	-24.1105	0.0009	-3.99	0.0038
	6	-23.7005	0.0011	-3.97	0.0041
	7	-22.5743	0.0017	-3.91	0.0048
	8	-22.4714	0.0018	-3.90	0.0048
	9	-22.0895	0.0020	-3.88	0.0051
	10	-21.9577	0.0021	-3.88	0.0052
	11	-22.0967	0.0020	-3.88	0.0051
	12	-21.6061	0.0024	-3.86	0.0055
Trend	0	-24.5908	0.0071	-3.99	0.0188
	1	-23.9176	0.0089	-3.95	0.0205
	2	-26.2344	0.0040	-4.08	0.0153
	3	-25.9658	0.0044	-4.06	0.0158
	4	-25.2503	0.0056	-4.02	0.0173
	5	-23.9467	0.0088	-3.95	0.0204
	6	-23.3860	0.0107	-3.92	0.0219
	7	-22.1004	0.0162	-3.85	0.0257
	8	-21.8359	0.0176	-3.84	0.0265
	9	-21.3119	0.0208	-3.81	0.0283
	10	-21.0419	0.0226	-3.80	0.0292
	11	-21.0611	0.0225	-3.80	0.0291
	12	-20.4919	0.0267	-3.77	0.0311

Table B10.1: DW statistics for Model 4: SARIMA$(0,1,2)(0,1,0)_{12}$

Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	1.7508	0.1757	0.8243
2	2.0566	0.5684	0.4316
3	1.7738	0.3135	0.6865
4	2.0101	0.6592	0.3408
5	1.8087	0.4996	0.5004
6	1.4792	0.2081	0.7919
7	1.5474	0.3416	0.6584
8	1.1735	0.0791	0.9209
9	1.6018	0.5615	0.4385
10	1.2326	0.2062	0.7938
11	1.1776	0.2108	0.7892
12	2.0246	0.9842	0.0158

Table B10.2: ARCH test for Model 4: SARIMA$(0,1,2)(0,1,0)_{12}$

Tests for ARCH Disturbances Based on OLS Residuals				
Order	Q	Pr > Q	LM	Pr > LM
1	0.0006	0.9797	0.0138	0.9066
2	0.5407	0.7631	0.1287	0.9377
3	0.6626	0.8820	0.2963	0.9607
4	0.8481	0.9319	0.6274	0.9600
5	2.4833	0.7790	2.3663	0.7965
6	3.1342	0.7918	3.4946	0.7447
7	4.6684	0.7004	5.2936	0.6242
8	5.4358	0.7101	5.6055	0.6913
9	5.6900	0.7705	5.6072	0.7785
10	6.1237	0.8048	6.6025	0.7624
11	6.2597	0.8555	6.6076	0.8299
12	8.6721	0.7306	8.1289	0.7750

Table B10.3: PP unit root test for Model 4: SARIMA$(0,1,2)(0,1,0)_{12}$

Phillips-Perron Unit Root Tests					
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	0	-28.9449	<.0001	-4.92	<.0001
	1	-29.1125	<.0001	-4.92	<.0001
	2	-28.6823	<.0001	-4.91	<.0001
	3	-28.7182	<.0001	-4.91	<.0001
	4	-27.6379	<.0001	-4.89	<.0001
	5	-26.6087	<.0001	-4.87	<.0001
	6	-26.1389	<.0001	-4.86	<.0001
	7	-25.6177	<.0001	-4.85	<.0001
	8	-25.8660	<.0001	-4.86	<.0001
	9	-25.6694	<.0001	-4.85	<.0001
	10	-25.7825	<.0001	-4.86	<.0001
	11	-26.2035	<.0001	-4.86	<.0001
	12	-25.5411	<.0001	-4.85	<.0001
Single Mean	0	-28.9683	0.0002	-4.85	0.0004
	1	-29.1163	0.0002	-4.86	0.0004
	2	-28.7094	0.0002	-4.85	0.0004
	3	-28.8152	0.0002	-4.85	0.0004
	4	-27.8124	0.0002	-4.82	0.0004
	5	-26.8709	0.0003	-4.80	0.0005
	6	-26.4908	0.0004	-4.80	0.0005
	7	-26.0510	0.0004	-4.79	0.0005
	8	-26.3691	0.0004	-4.79	0.0005
	9	-26.2380	0.0004	-4.79	0.0005
	10	-26.4060	0.0004	-4.79	0.0005
	11	-26.8799	0.0003	-4.80	0.0005
	12	-26.2675	0.0004	-4.79	0.0005
Trend	0	-29.8544	0.0010	-4.92	0.0019
	1	-30.0387	0.0009	-4.92	0.0019
	2	-29.3851	0.0012	-4.90	0.0020
	3	-29.1863	0.0013	-4.90	0.0020
	4	-27.9292	0.0021	-4.87	0.0021
	5	-26.6571	0.0034	-4.85	0.0022
	6	-25.8966	0.0045	-4.84	0.0023
	7	-25.0666	0.0060	-4.84	0.0023
	8	-25.0097	0.0061	-4.84	0.0023
	9	-24.5492	0.0072	-4.84	0.0023
	10	-24.3981	0.0076	-4.84	0.0023
	11	-24.5731	0.0071	-4.84	0.0023
	12	-23.7065	0.0096	-4.84	0.0023

Table B11.1: DW statistics for Model 5: SARIMA$(0,1,2)(1,1,0)_{12}$

Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	1.4762	0.0376	0.9624
2	1.9951	0.4954	0.5046
3	2.0323	0.6131	0.3869
4	1.8714	0.4984	0.5016
5	1.8136	0.5055	0.4945
6	1.5136	0.2401	0.7599
7	1.6991	0.5222	0.4778
8	1.2490	0.1230	0.8770
9	1.4178	0.3375	0.6625
10	1.1950	0.1722	0.8278
11	1.2055	0.2391	0.7609
12	1.6101	0.7911	0.2089

Table B11.2: ARCH test for Model 5: SARIMA(*0,1,2*)*(1,1,0)*$_{12}$

Tests for ARCH Disturbances Based on OLS Residuals				
Order	Q	Pr > Q	LM	Pr > LM
1	0.2033	0.6521	0.2413	0.6233
2	1.1965	0.5498	0.6273	0.7308
3	1.3354	0.7207	0.6933	0.8748
4	1.5118	0.8246	0.7039	0.9508
5	1.7045	0.8883	1.1747	0.9473
6	3.4600	0.7493	3.6512	0.7238
7	5.8853	0.5532	5.8969	0.5518
8	5.8865	0.6599	5.9729	0.6503
9	5.8938	0.7505	6.1259	0.7273
10	6.8603	0.7386	6.7780	0.7462
11	7.9234	0.7202	7.6842	0.7413
12	8.9624	0.7061	7.8204	0.7990

Table B11.3: PP unit root test for Model 5: SARIMA(*0,1,2*)*(1,1,0)*$_{12}$

Phillips-Perron Unit Root Tests					
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	0	-25.1112	<.0001	-4.31	<.0001
	1	-25.5355	<.0001	-4.33	<.0001
	2	-25.3470	<.0001	-4.32	<.0001
	3	-24.4283	<.0001	-4.28	<.0001
	4	-23.4736	0.0002	-4.24	0.0001
	5	-22.5007	0.0002	-4.20	0.0001
	6	-22.0050	0.0003	-4.18	0.0001
	7	-21.0199	0.0004	-4.15	0.0001
	8	-20.7549	0.0005	-4.14	0.0001
	9	-20.3568	0.0006	-4.13	0.0001
	10	-20.2423	0.0006	-4.12	0.0001
	11	-20.3968	0.0006	-4.13	0.0001
	12	-20.0075	0.0007	-4.12	0.0001
Single Mean	0	-25.1125	0.0006	-4.23	0.0020
	1	-25.5395	0.0005	-4.25	0.0019
	2	-25.3495	0.0006	-4.24	0.0019
	3	-24.4245	0.0008	-4.20	0.0022
	4	-23.4633	0.0012	-4.16	0.0025
	5	-22.4835	0.0017	-4.11	0.0027
	6	-21.9837	0.0021	-4.09	0.0029
	7	-20.9918	0.0030	-4.06	0.0032
	8	-20.7240	0.0033	-4.05	0.0033
	9	-20.3226	0.0039	-4.03	0.0034
	10	-20.2065	0.0040	-4.03	0.0035
	11	-20.3609	0.0038	-4.03	0.0034
	12	-19.9686	0.0044	-4.02	0.0036
Trend	0	-25.3998	0.0053	-4.22	0.0109
	1	-25.8358	0.0046	-4.24	0.0104
	2	-25.5219	0.0051	-4.23	0.0107
	3	-24.4557	0.0074	-4.18	0.0121
	4	-23.4131	0.0106	-4.13	0.0135
	5	-22.3126	0.0151	-4.08	0.0150
	6	-21.6459	0.0187	-4.06	0.0160
	7	-20.4913	0.0267	-4.02	0.0177
	8	-20.0770	0.0303	-4.00	0.0183
	9	-19.5508	0.0354	-3.98	0.0190
	10	-19.3120	0.0380	-3.98	0.0193
	11	-19.3547	0.0375	-3.98	0.0192
	12	-18.8855	0.0430	-3.97	0.0198

Appendix C

Table C1: Tentative order selection tests using the ESACF for Model 2: SARIMA(*1,1,1*)(*1,1,1*)$_{12}$

ARMA(p+d,q) Tentative Order Selection Tests	
ESACF	
p+d	q
0	0
6	0
3	2
7	0
4	2
5	2
8	0
10	1
12	0
13	0

Table C2.1: Information criteria of an AR(*8*) *for fitting the residuals of Model 2:* SARIMA(*1,1,1*)(*1,1,1*)$_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
AR1,1	0.30063	0.18286	1.64	0.1002	8

Variance Estimate	7.385E-6
Std Error Estimate	0.002717
AIC	-312.495
SBC	-310.94
Number of Residuals	35

Table C2.2: Information criteria of an ARMA(*4,2*) *for fitting the residuals of Model 2:* SARIMA(*1,1,1*)(*1,1,1*)$_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.24234	0.18288	-1.33	0.1851	2
AR1,1	-0.22358	0.19566	-1.14	0.2532	4

Variance Estimate	7.561E-6
Std Error Estimate	0.00275
AIC	-311.093
SBC	-307.982
Number of Residuals	35

Table C3: Tentative order selection tests using the ESACF for Model 3: SARIMA(*0,1,0*)(*1,1,0*)$_{12}$

ARMA(p+d,q) Tentative Order Selection Tests	
ESACF	
p+d	q
0	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
11	0
13	0

Table C4.1: Information criteria of an AR(2) *for fitting the residuals of Model 3:* SARIMA$(0,1,0)(1,1,0)_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
AR1,1	0.20572	0.18249	1.13	0.2596	2

Variance Estimate	7.415E-6
Std Error Estimate	0.002723
AIC	-313.022
SBC	-311.466
Number of Residuals	35

Table C4.2: Information criteria of an AR(3) *for fitting the residuals of Model 3:* SARIMA$(0,1,0)(1,1,0)_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
AR1,1	-0.19172	0.19025	-1.01	0.3136	3

Variance Estimate	7.477E-6
Std Error Estimate	0.002734
AIC	-312.704
SBC	-311.149
Number of Residuals	35

Table C4.3: Information criteria of an AR(4) *for fitting the residuals of Model 3:* SARIMA$(0,1,0)(1,1,0)_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
AR1,1	-0.16885	0.19093	-0.88	0.3765	4

Variance Estimate	7.53E-6
Std Error Estimate	0.002744
AIC	-312.453
SBC	-310.898
Number of Residuals	35

Table C4.4: Information criteria of an AR(5) *for fitting the residuals of Model 3:* SARIMA$(0,1,0)(1,1,0)_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
AR1,1	-0.20335	0.19083	-1.07	0.2866	5

Variance Estimate	7.433E-6
Std Error Estimate	0.002726
AIC	-312.813
SBC	-311.258
Number of Residuals	35

Table C4.5: Information criteria of an AR(*8*) *for fitting the residuals of Model 3:* SARIMA(*0,1,0*)(*1,1,0*)$_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
AR1,1	0.24048	0.19299	1.25	0.2127	8

Variance Estimate	7.311E-6
Std Error Estimate	0.002704
AIC	-313.127
SBC	-311.572
Number of Residuals	35

Table C5.1: Information criteria of an AR(*3*) *for fitting the residuals of Model 5:* SARIMA(0,1,2)(1,1,0)$_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
AR1,1	-0.18487	0.18806	-0.98	0.3256	3

Variance Estimate	7.099E-6
Std Error Estimate	0.002664
AIC	-314.528
SBC	-312.973
Number of Residuals	35

Table C5.2: Information criteria of an ARMA(*5,1*) *for fitting the residuals of Model 5:* SARIMA(*0,1,2*)(*1,1,0*)$_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.27401	0.16890	-1.62	0.1047	1
AR1,1	-0.14475	0.18720	-0.77	0.4394	5

Variance Estimate	6.887E-6
Std Error Estimate	0.002624
AIC	-314.553
SBC	-311.443
Number of Residuals	35

Table C5.3: Information criteria of an ARMA(*11,1*) *for fitting the residuals of Model 5:* SARIMA(*0,1,2*)(*1,1,0*)$_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.28023	0.16920	-1.66	0.0977	1
AR1,1	0.15582	0.20516	0.76	0.4476	11

Variance Estimate	6.862E-6
Std Error Estimate	0.002619
AIC	-314.517
SBC	-311.406
Number of Residuals	35

Table C5.4: Information criteria of an ARMA(*8,1*) *for fitting the residuals of Model 5:* SARIMA(*0,1,2*)(*1,1,0*)$_{12}$

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.30755	0.16813	-1.83	0.0674	1
AR1,1	0.25645	0.18760	1.37	0.1716	8

Variance Estimate	6.605E-6
Std Error Estimate	0.00257
AIC	-315.56
SBC	-312.45
Number of Residuals	35

Table C6: Results of the Fourier algorithm for the residual series of Model 2: SARIMA(*1,1,1*)(*1,1,1*)$_{12}$

Frequency	Period	Cosine coefficients	Sine coefficients	Power
0		0.00002885	0	0.00000001
0.17951958	35.00000000	-0.00131963	0.00073023	0.00003981
0.35903916	17.50000000	-0.00088312	-0.00050801	0.00001816
0.53855874	11.66666667	-0.00030615	-0.00050925	0.00000618
0.71807832	8.75000000	-0.00022807	0.00098563	0.00001791
0.89759790	7.00000000	0.00132194	0.00047724	0.00003457
1.07711748	5.83333333	-0.00012423	0.00069644	0.00000876
1.25663706	5.00000000	0.00092823	-0.00012766	0.00001536
1.43615664	4.37500000	0.00062474	-0.00050196	0.00001124
1.61567622	3.88888889	0.00039396	-0.00001789	0.00000272
1.79519580	3.50000000	0.00103026	0.00018002	0.00001914
1.97471538	3.18181818	-0.00006017	0.00040775	0.00000297
2.15423496	2.91666667	0.00017428	0.00021694	0.00000136
2.33375454	2.69230769	-0.00085847	0.00144161	0.00004927
2.51327412	2.50000000	0.00038486	-0.00033423	0.00000455
2.69279370	2.33333333	-0.00013800	-0.00018318	0.00000092
2.87231328	2.18750000	-0.00066477	-0.00065527	0.00001525
3.05183286	2.05882353	0.00121549	0.00027626	0.00002719

Figure C1: Frequency analysis for the residual series of Model 2: SARIMA*(1,1,1)(1,1,1)*$_{12}$

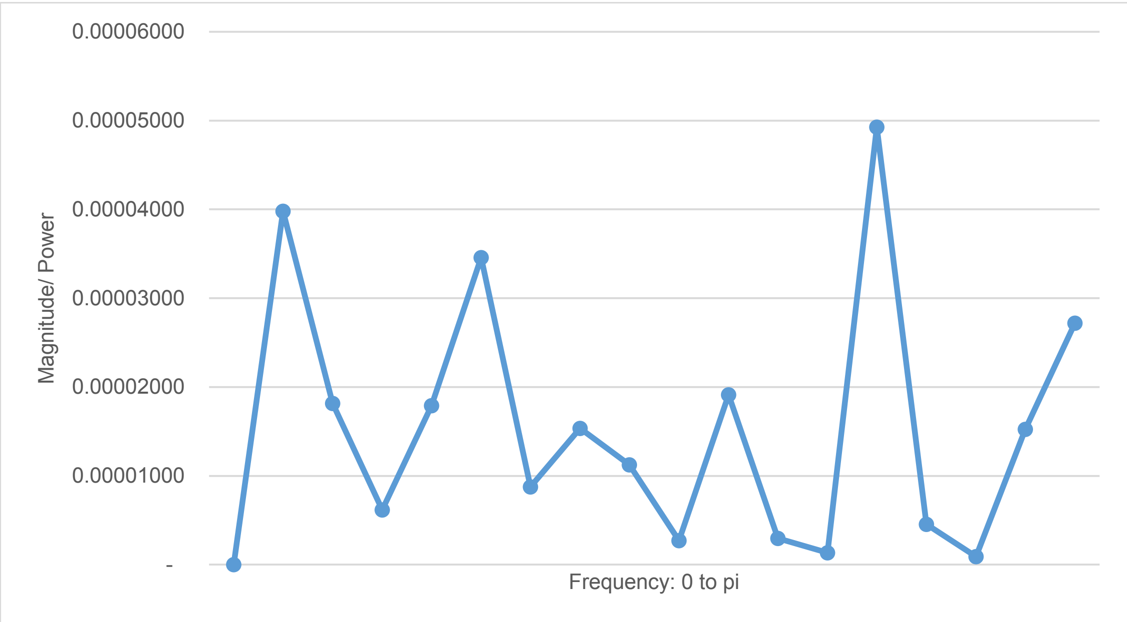


Table C7: Results of the Fourier algorithm for the residual series of Model 3: SARIMA*(0,1,0)(1,1,0)*$_{12}$

Frequency	Period	Cosine coefficients	Sine coefficients	Power
0		0.00033688	0	0.00000199
0.17951958	35.00000000	-0.00143075	0.00035629	0.00003804
0.35903916	17.50000000	-0.00119660	-0.00028858	0.00002651
0.53855874	11.66666667	-0.00013364	-0.00068220	0.00000846
0.71807832	8.75000000	-0.00022262	0.00121022	0.00002650
0.89759790	7.00000000	0.00132510	0.00027467	0.00003205
1.07711748	5.83333333	-0.00035039	0.00074052	0.00001175
1.25663706	5.00000000	0.00095035	0.00017540	0.00001634
1.43615664	4.37500000	0.00065011	-0.00058814	0.00001345
1.61567622	3.88888889	0.00046055	0.00003743	0.00000374
1.79519580	3.50000000	0.00061256	0.00038561	0.00000917
1.97471538	3.18181818	0.00019652	0.00062270	0.00000746
2.15423496	2.91666667	0.00035961	0.00043160	0.00000552
2.33375454	2.69230769	-0.00063956	0.00093860	0.00002258
2.51327412	2.50000000	0.00014472	-0.00030207	0.00000196
2.69279370	2.33333333	-0.00028427	-0.00011543	0.00000165
2.87231328	2.18750000	-0.00045586	-0.00021459	0.00000444
3.05183286	2.05882353	0.00135129	0.00008281	0.00003207

Figure C2: Frequency analysis for the residual series of Model 3: SARIMA*(0,1,0)(1,1,0)*$_{12}$

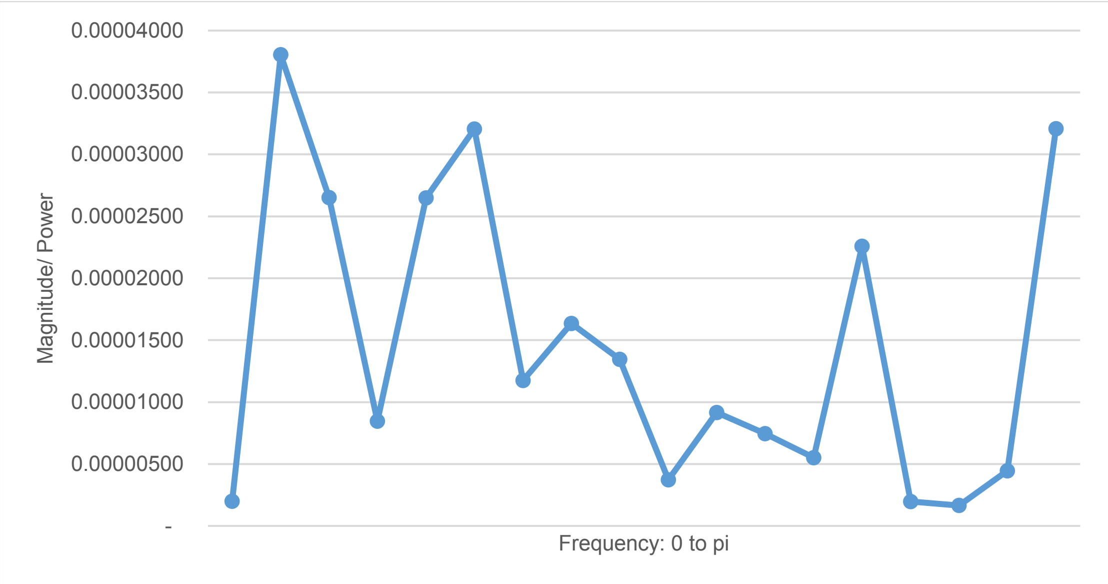


Table C8: Results of the Fourier algorithm for the residual series of Model 5: SARIMA*(0,1,2)(1,1,0)*$_{12}$

Frequency	Period	Cosine coefficients	Sine coefficients	Power
0		0.00018827	0	0.00000062
0.17951958	35.00000000	-0.00117218	0.00040732	0.00002695
0.35903916	17.50000000	-0.00110491	-0.00008281	0.00002148
0.53855874	11.66666667	-0.00033220	-0.00055089	0.00000724
0.71807832	8.75000000	0.00004250	0.00114461	0.00002296
0.89759790	7.00000000	0.00121193	-0.00016994	0.00002621
1.07711748	5.83333333	-0.00024777	0.00088588	0.00001481
1.25663706	5.00000000	0.00116764	-0.00013803	0.00002419
1.43615664	4.37500000	0.00067828	-0.00100865	0.00002586
1.61567622	3.88888889	0.00054965	-0.00004329	0.00000532
1.79519580	3.50000000	0.00069824	0.00057662	0.00001435
1.97471538	3.18181818	-0.00002549	0.00074208	0.00000965
2.15423496	2.91666667	0.00016965	0.00050133	0.00000490
2.33375454	2.69230769	-0.00088618	0.00058683	0.00001977
2.51327412	2.50000000	0.00021205	-0.00025378	0.00000191
2.69279370	2.33333333	-0.00021743	-0.00014086	0.00000117
2.87231328	2.18750000	-0.00035704	-0.00023962	0.00000324
3.05183286	2.05882353	0.00102471	0.00008396	0.00001850

Figure C3: Frequency analysis for the residual series of Model 5: SARIMA$(0,1,2)(1,1,0)_{12}$

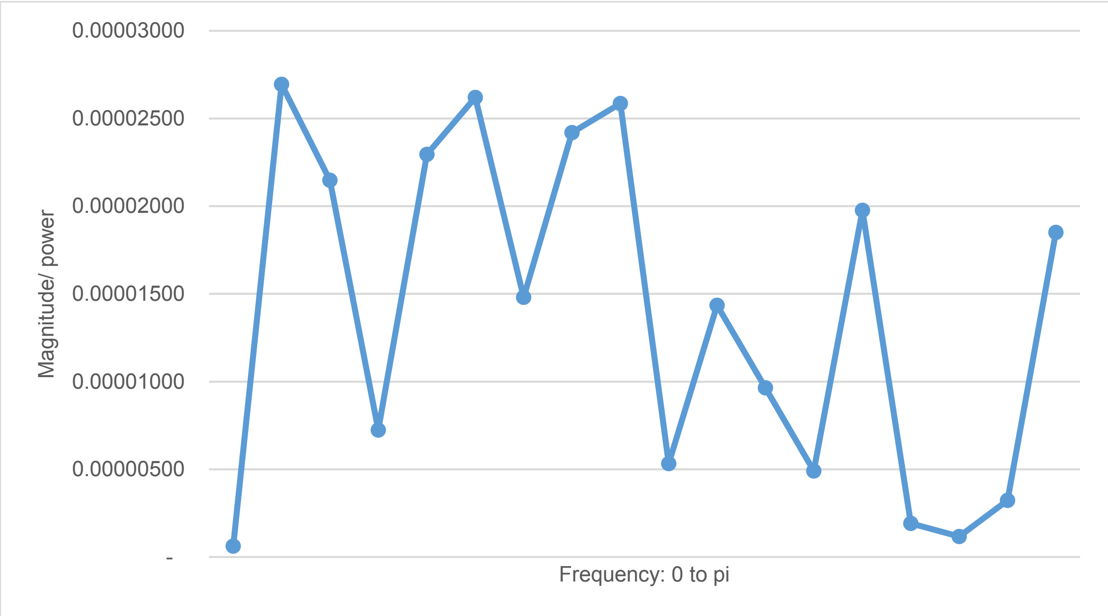


Table C9: Impact of reducing Fourier terms for residual series of Model 2: SARIMA$(1,1,1)(1,1,1)_{12}$

Number of Fourier terms	Period	Measure	Model 2
Number of Fourier terms: 1	In-sample	MAE	0.002271083
		RMSE	0.002594245
	Out-of-sample	MAE	0.002550454
		RMSE	0.002904968
Number of Fourier terms: 2	In-sample	MAE	0.002108653
		RMSE	0.002492213
	Out-of-sample	MAE	0.002270469
		RMSE	0.002633562
Number of Fourier terms: 3	In-sample	MAE	0.002034753
		RMSE	0.002456541
	Out-of-sample	MAE	0.002163589*
		RMSE	0.002496222*
Number of Fourier terms: 4	In-sample	MAE	0.001977965
		RMSE	0.002350074
	Out-of-sample	MAE	0.002171225
		RMSE	0.002676727
Number of Fourier terms: 5	In-sample	MAE	0.001742118
		RMSE	0.002129601
	Out-of-sample	MAE	0.002348259
		RMSE	0.002621276
Number of Fourier terms: 6	In-sample	MAE	0.001671519
		RMSE	0.002070017
	Out-of-sample	MAE	0.002446306
		RMSE	0.002767631
Number of Fourier terms: 7	In-sample	MAE	0.0015053
		RMSE	0.001961128
	Out-of-sample	MAE	0.002372208
		RMSE	0.002652399
Number of Fourier terms: 8	In-sample	MAE	0.001491984
		RMSE	0.001877469
	Out-of-sample	MAE	0.002402177
		RMSE	0.002719828
Number of Fourier terms: 9	In-sample	MAE	0.001448548
		RMSE	0.001856644
	Out-of-sample	MAE	0.002452053
		RMSE	0.002749638
Number of Fourier terms: 10	In-sample	MAE	0.001409704
		RMSE	0.001703
	Out-of-sample	MAE	0.002519188
		RMSE	0.002872817
Number of Fourier terms: 11	In-sample	MAE	0.001358689
		RMSE	0.001677876
	Out-of-sample	MAE	0.002483146

		RMSE	0.00290201
Number of Fourier terms: 12	In-sample	MAE	0.001348126
		RMSE	0.001666298
	Out-of-sample	MAE	0.002512575
		RMSE	0.002942326
Number of Fourier terms: 13	In-sample	MAE	0.001048357
		RMSE	0.001170018
	Out-of-sample	MAE	0.002571924
		RMSE	0.003172952
Number of Fourier terms: 14	In-sample	MAE	0.000965108
		RMSE	0.001113117
	Out-of-sample	MAE	0.002561678
		RMSE	0.003154759
Number of Fourier terms: 15	In-sample	MAE	0.000991981
		RMSE	0.00110124
	Out-of-sample	MAE	0.002544034
		RMSE	0.003158348
Number of Fourier terms: 16	In-sample	MAE	0.000793387*
		RMSE	0.00088152*
	Out-of-sample	MAE	0.00260606
		RMSE	0.003202933

Table C10: Impact of reducing Fourier terms for residual series of Model 3: SARIMA$(0,1,0)(1,1,0)_{12}$

Number of Fourier terms	**Period**	**Measure**	**Model 3**
Number of Fourier terms: 1	In-sample	MAE	0.002203739
		RMSE	0.002533452
	Out-of-sample	MAE	0.002686441
		RMSE	0.003119118
Number of Fourier terms: 2	In-sample	MAE	0.002003271
		RMSE	0.002379245
	Out-of-sample	MAE	0.002386523
		RMSE	0.003080747
Number of Fourier terms: 3	In-sample	MAE	0.001920407
		RMSE	0.002327914
	Out-of-sample	MAE	0.002213971*
		RMSE	0.00302743
Number of Fourier terms: 4	In-sample	MAE	0.001864964
		RMSE	0.002159186
	Out-of-sample	MAE	0.00290578
		RMSE	0.003709471
Number of Fourier terms: 5	In-sample	MAE	0.001560524
		RMSE	0.001935565
	Out-of-sample	MAE	0.002499242
		RMSE	0.003097209
Number of Fourier terms: 6	In-sample	MAE	0.001504445
		RMSE	0.001846845
	Out-of-sample	MAE	0.002421606
		RMSE	0.002985321
Number of Fourier terms: 7	In-sample	MAE	0.001311568
		RMSE	0.001715771
	Out-of-sample	MAE	0.002560319
		RMSE	0.003163293
Number of Fourier terms: 8	In-sample	MAE	0.001319872
		RMSE	0.001599874
	Out-of-sample	MAE	0.002551049
		RMSE	0.002975362
Number of Fourier terms: 9	In-sample	MAE	0.001270844
		RMSE	0.001566155
	Out-of-sample	MAE	0.002557394
		RMSE	0.002964362*
Number of Fourier terms: 10	In-sample	MAE	0.001219274
		RMSE	0.001480161
	Out-of-sample	MAE	0.002639571
		RMSE	0.003131727
Number of Fourier terms: 11	In-sample	MAE	0.001143239
		RMSE	0.001406303
	Out-of-sample	MAE	0.0027331

		RMSE	0.003287729
Number of Fourier terms: 12	In-sample	MAE	0.001130395
		RMSE	0.001349033
	Out-of-sample	MAE	0.002815497
		RMSE	0.003323651
Number of Fourier terms: 13	In-sample	MAE	0.000956572
		RMSE	0.001083921
	Out-of-sample	MAE	0.003079738
		RMSE	0.003486084
Number of Fourier terms: 14	In-sample	MAE	0.000935167
		RMSE	0.001057729
	Out-of-sample	MAE	0.003021989
		RMSE	0.003438346
Number of Fourier terms: 15	In-sample	MAE	0.000972308
		RMSE	0.001035241
	Out-of-sample	MAE	0.003039848
		RMSE	0.00346859
Number of Fourier terms: 16	In-sample	MAE	0.000868057*
		RMSE	0.000972006*
	Out-of-sample	MAE	0.003011581
		RMSE	0.003439321

Table C11: Impact of reducing Fourier terms for residual series of Model 5: SARIMA$(0,1,2)(1,1,0)_{12}$

Number of Fourier terms	Period	Measure	Model 5
Number of Fourier terms: 1	In-sample	MAE	0.002143489
		RMSE	0.002517782
	Out-of-sample	MAE	0.002397194
		RMSE	0.002926802
Number of Fourier terms: 2	In-sample	MAE	0.001998994
		RMSE	0.002392778
	Out-of-sample	MAE	0.002174938
		RMSE	0.002681741
Number of Fourier terms: 3	In-sample	MAE	0.001954345
		RMSE	0.002349142
	Out-of-sample	MAE	0.001997727
		RMSE	0.002498082
Number of Fourier terms: 4	In-sample	MAE	0.001873384
		RMSE	0.002205109
	Out-of-sample	MAE	0.002541795
		RMSE	0.003106312
Number of Fourier terms: 5	In-sample	MAE	0.001677071
		RMSE	0.002028221
	Out-of-sample	MAE	0.002010488
		RMSE	0.002471661
Number of Fourier terms: 6	In-sample	MAE	0.001505207
		RMSE	0.001921092
	Out-of-sample	MAE	0.001895822*
		RMSE	0.002276416*
Number of Fourier terms: 7	In-sample	MAE	0.001312654
		RMSE	0.001731871
	Out-of-sample	MAE	0.001972118
		RMSE	0.002466284
Number of Fourier terms: 8	In-sample	MAE	0.001210357
		RMSE	0.00150355
	Out-of-sample	MAE	0.001906407
		RMSE	0.002320881
Number of Fourier terms: 9	In-sample	MAE	0.001195654
		RMSE	0.001452126
	Out-of-sample	MAE	0.001997711
		RMSE	0.002381511
Number of Fourier terms: 10	In-sample	MAE	0.001093766
		RMSE	0.001303323
	Out-of-sample	MAE	0.002259847
		RMSE	0.00263495
Number of Fourier terms: 11	In-sample	MAE	0.000950758
		RMSE	0.00119289
	Out-of-sample	MAE	0.002439902

		RMSE	0.002825725
Number of Fourier terms: 12	In-sample	MAE	0.000925374
		RMSE	0.001132666
	Out-of-sample	MAE	0.00254442
		RMSE	0.00287978
Number of Fourier terms: 13	In-sample	MAE	0.000739878
		RMSE	0.000847402
	Out-of-sample	MAE	0.002627804
		RMSE	0.003012684
Number of Fourier terms: 14	In-sample	MAE	0.000713962
		RMSE	0.000814497
	Out-of-sample	MAE	0.00256387
		RMSE	0.002983611
Number of Fourier terms: 15	In-sample	MAE	0.000744249
		RMSE	0.000793629
	Out-of-sample	MAE	0.002579054
		RMSE	0.003002371
Number of Fourier terms: 16	In-sample	MAE	0.000657424*
		RMSE	0.000733075*
	Out-of-sample	MAE	0.002604675
		RMSE	0.002987838