

# A New Zealand study of association between crime and the state of the economy

by

Shuhan Xu

A thesis

submitted to the Victoria University of Wellington

in fulfilment of the requirements for the degree of

Master of Science in Statistics

Victoria University of Wellington

2017



# Abstract

The aim of this thesis is to investigate whether there are associations between economically motivated crimes and macroeconomic variables. Economically motivated crimes include burglary, fraud and theft. Non-traffic offences are used as the measurement of overall crime levels, and an association between non-traffic offences and macroeconomic variables is analysed as well. Forecasting the number of people charged with burglary, fraud, theft and non-traffic offences is another objective of this thesis. Association between economically motivated crimes and the unemployment rate is also analysed at a regional level.

Methods used in this thesis include Vector Autoregressive (VAR) models, Vector Error Correction Models (VECM) and Autoregressive Integrated Moving Average (ARIMA) models. VECM and VAR models are used to produce Granger-Causality tests and impulse responses in order to summarise the associations between crimes and macroeconomic variables. All modelling methods are used to generate forecasts.

The conclusion from this thesis is that there are associations between crime and some macroeconomic variables at a national level. The biggest impact on crime is its own value in the past. The impact of macroeconomic variables is minor, and this makes the sign of the impact less important. In fact, the sign of the impact is hard to conclude because it moves between positive and negative in different periods. At a national level, the growth rate of unemployment causes the growth rate of burglary, theft and non-traffic charges. The association between unemployment and crime becomes insignificant once all macroeconomic variables are included. Overall, the growth rate of personal weekly average income or household debt and disposable income ratio (both measuring personal or household financial condition) causes an increase in the growth rate of burglary, theft and non-traffic charges. Movement of inflation causes an increase in the growth rate of fraud charges. At a regional level, growth in the unemployment rate causes an increase in theft charges in Auckland and Northland. In Nelson/Marlborough/West Coast, growth in the unemployment rate causes growth in burglary charges and vice versa. Growth in the unemployment rate causes growth in the rate of fraud charges, but this is found in Northland only. Forecasts produced by this study suggest that the number of people charged with burglary, theft, fraud and non-traffic offences will continue to decrease up until 2019, but at a lower rate of reduction.

# Acknowledgement

I would like to express my gratitude to my supervisor, Dr John Haywood for his invaluable guidance, positive encouragement and feedback throughout my Masters study at Victoria University of Wellington. Special thanks to Dr Russil Durrant for his advice in Criminology.

I would like to deeply thank my parents for their support and encouragement throughout the many years of my school and adult life.

This research is only possible because of the data provided by Ministry of Justice. My thanks go to the team at Ministry of Justice especially Adrian Angus, Jo Fink, Suzanne Jones and Rory McRae.

Thanks go to my colleagues as well, especially my manager Greg Evans for his support in the past year. Last but not least, special thanks to go to Jason Ford and Paul Yan for doing the proofreading.

# Contents

<b>Chapter 1</b>	<b><i>Introduction</i></b>	<b>1</b>
1.1	Background and purpose of this study	1
1.2	Structure of this study	5
<b>Chapter 2</b>	<b><i>Literature review</i></b>	<b>7</b>
<b>Chapter 3</b>	<b><i>Statistical Methodology</i></b>	<b>10</b>
3.1	Stationary and Test for unit root	11
3.2	ARIMA and linear regression model	12
3.3	VAR and VECM Modelling	17
3.3.1	VAR	17
3.3.2	Cointegration and VECM	18
3.3.3	Granger-Causality test	19
3.3.4	Impulse Response	21
3.4	Seasonality Modelling	22
3.5	Principal Components Analysis	23
3.6	Model diagnostics	25
<b>Chapter 4</b>	<b><i>Results</i></b>	<b>29</b>
4.1	Model 1: Unemployment and crime	29
4.1.1	National Level Analysis	29
4.1.1.1	Summary Result	32
4.1.1.2	Impulse response	35
4.1.2	Regional Analysis	37
4.1.2.1	Summary Result	39
4.2	Model 2: Crime and the state of the economy	42
4.2.1	Non-traffic and the state of the economy model	44
4.2.2	Burglary and the state of the economy model	49
4.2.3	Fraud and the state of the economy model	53
4.2.4	Theft and the state of the economy model	57
4.3	Model 3: Association between Theft, Burglary and Fraud	60

4.4	Model 4: Crime forecasting using ARIMA model .....	63
<b>Chapter 5</b>	<b><i>Model Comparison and Forecasting</i></b> .....	<b>68</b>
5.1	In-Sample and Out-of-sample comparison .....	69
5.2	Forecasting .....	79
<b>Chapter 6</b>	<b><i>Discussion and Conclusions</i></b> .....	<b>85</b>
	<b><i>Bibliography</i></b> .....	<b>92</b>
<b>Appendix A</b>	<b><i>Unemployment and crime model</i></b> .....	<b>95</b>
	Non-traffic and unemployment model .....	95
	Burglary and unemployment model .....	96
	Fraud and unemployment rate .....	97
	Theft and Unemployment rate and model .....	98
<b>Appendix B</b>	<b><i>Crime association model</i></b> .....	<b>99</b>
	Number of people charged .....	99
	Crime rate .....	100
<b>Appendix C</b>	<b><i>Crime and the state of the economy model</i></b> .....	<b>102</b>
	Non-traffic and the state of the economy model .....	102
	Full model .....	102
	Reduced Model .....	107
	Theft and the state of the economy model .....	108
	Full Model .....	108
	Reduced Model .....	113
	Burglary and the state of the economy model .....	115
	Full Model1 .....	115
	Full Model2 .....	120
	Reduced Model .....	125
	Fraud and the state of the economy model .....	128
	Full Model .....	128
	Reduced model .....	134
<b>Appendix D</b>	<b><i>ARIMA model</i></b> .....	<b>138</b>

<b>Seasonal Dummy Model .....</b>	<b>138</b>
Non-traffic .....	138
Burglary .....	139
Theft .....	140
Fraud .....	141
<b>Seasonal Differencing .....</b>	<b>143</b>
Non-traffic .....	143
Burglary .....	144
Theft .....	145
Fraud .....	146
<b><i>Appendix E   SAS code .....</i></b>	<b><i>148</i></b>

# List of Figures

Figure 4.1: Quarterly unemployment rate in New Zealand, 1992-2012 .....	30
Figure 4.2: Quarterly unemployment rate and number of people charged with theft in logarithm form, 1992-2012.....	30
Figure 4.3: Quarterly unemployment rate and number of people charged with burglary and fraud in logarithm form, 1992-2012 .....	31
Figure 4.4: Quarterly unemployment rate and number of people charged with Non-traffic offences in logarithm form, 1992-2012 .....	31
Figure 4.5: Impulse response of crime variables to one unit shock to 'LnU' from unemployment and crime models.....	35
Figure 4.6: Impulse response of 'LnU' to one unit shock to crime variables from unemployment and crime models.....	35
Figure 4.7: Impulse response of crime variables to one unit shock to itself from unemployment and crime models.....	36
Figure 4.8: Quarterly unemployment rate by region, 1992-2012 .....	38
Figure 4.9: Quarterly number of people charged by offence categories in Gisborne/Hawke's Bay, 1992-2012 .....	38
Figure 4.10: Quarterly number of people charged by offence categories in Nelson/Marlborough/West Coast, 1992-2012 .....	38
Figure 4.11: Quarterly number of people charged by offence categories in Northland, 1992-2012...	39
Figure 4.12: Quarterly number of people charged by offence categories in Auckland, 1992-2012 ....	39
Figure 4.13: Quarterly crime rate by offence type, 1992-2012 .....	43
Figure 4.14: Quarterly crime rate for non-traffic offences, 1992-2012.....	43
Figure 4.15 Weekly personal average income and Household debt and disposable income ratio, 1992-2012 .....	43
Figure 4.16: Quarterly real GDP growth rate, inflation and unemployment rate, 1992-2012 .....	43
Figure 4.17: Impulse response of 'LnNTn' to one unit shock to 'LnD' and 'LnNTn' .....	48
Figure 4.18: Impulse response of 'LnNTr' to one unit shock to 'LnD' and 'LnNTr' .....	48
Figure 4.19: Impulse response of 'LnD' to one unit shock to 'LnNTr' and 'LnNTn' .....	48
Figure 4.20: Impulse response of 'LnBn' to one unit shock to macroeconomic variables .....	51
Figure 4.21: Impulse response of 'LnBr' to one unit shock to macroeconomic variables .....	51
Figure 4.22: Impulse response of 'LnBn' and 'LnBr' to one unit shock to itself.....	51



Figure 4.23: Impulse response of macroeconomic variables to one unit shock to 'LnBn' .....	52
Figure 4.24: Impulse response of macroeconomic variables to one unit shock to 'LnBr' .....	53
Figure 4.25: Impulse response of 'LnFn' to one unit shocks in macroeconomic variables .....	55
Figure 4.26: Impulse response of 'LnFr' to one unit shock to macroeconomic variables .....	55
Figure 4.27: Impulse response of macroeconomic variables to one unit shock to 'LnFn' .....	56
Figure 4.28: Impulse response of macroeconomic variables to one unit shock to 'LnFr' .....	56
Figure 4.29: Impulse response of 'LnFn' to one unit shock to 'LnFn' and 'LnFr' .....	56
Figure 4.30: Impulse response of 'LnTn' to one unit shock to 'LnTn', 'LnW' and 'LnD' .....	59
Figure 4.31: Impulse response of 'LnTr' to one unit shock to 'LnTr', 'LnW' and 'LnD' .....	59
Figure 4.32: Impulse response of 'LnD' to one unit shock to 'LnTn', 'LnTr' and 'LnD' .....	59
Figure 4.33: Impulse response of 'LnW' to one unit shock to 'LnTn', 'LnTr' and 'LnW' .....	59
Figure 5.1: Out-of-sample forecasts for number of people charged with non-traffic offences.....	72
Figure 5.2: Out-of-sample forecasts using Equal-Weighted-Average for number of people charged with non-traffic offences.....	72
Figure 5.3: Out-of-sample forecasts for number of people charged with burglary .....	74
Figure 5.4: Out-of-sample forecasts for number of people charged with fraud .....	75
Figure 5.5: Out-of-sample forecasts for number of people charged with fraud: MAE, MSE and MPE for Equal-Weighted-Average.....	76
Figure 5.6: Out-of-sample forecasts for number of people charged with theft.....	78
Figure 5.7: Out-of-sample forecasts for number of people charged with theft: MAE, MSE and MPE for Equal-Weighted-Average.....	79
Figure 5.8: Number of people charged with non-traffic offences: 2015-2019 forecasts by ARIMA((4), (1,4), 0) and ARIMA(0, (1,4), (4)) .....	81
Figure 5.9: Final forecast for number of people charged with non-traffic offences, 2015-2019 .....	81
Figure 5.10: Number of people charged with burglary: 2015-2019 forecasts by VECM(5) and ARIMAX(1,1,0) .....	82
Figure 5.11: Final forecast for number of people charged with burglary, 2015-2019 .....	82
Figure 5.12: Number of people charged with fraud: 2015-2019 forecasts by ARIMA(0,1,1) (0,1,1) <sub>4</sub> and ARIMAX(2, 1, 0) .....	83
Figure 5.13: Final forecast for number of people charged with fraud, 2015-2019 .....	83
Figure 5.14: Final forecast for number of people charged with theft, 2015-2019 .....	84

# List of Tables

Table 1.1: List of variables used in this study with variable names, variable descriptions and variable codes .....	4
Table 3.1: Granger-Causality test hypothesis and restriction for VAR(K) .....	20
Table 4.1: Dickey-Fuller Unit Root Tests at a national level .....	32
Table 4.2: Cointegration test result for unemployment and crime model at a national level .....	33
Table 4.3: List of unemployment and crime models with good model diagnostics at a national level .....	33
Table 4.4: Granger-Causality test result for unemployment and crime model at a national level .....	34
Table 4.5: Regional level cointegration test results for unemployment and crime model .....	40
Table 4.6: List of regional level unemployment and crime models with good model diagnostics.....	40
Table 4.7: Granger-Causality test for unemployment and crime model at a regional level .....	41
Table 4.8: Granger-Causality test from VAR(4) for Non-traffic and the state of the economy model.....	45
Table 4.9: Correlation between principal components and original variables.....	46
Table 4.10: Eigenvectors for PCA for 'U', 'GDP' and 'I' .....	46
Table 4.11 Granger-Causality test for non-traffic and the state of the economy model .....	47
Table 4.12: Granger-Causality test for burglary and the state of the economy model (full model) ....	49
Table 4.13: Granger-Causality test for Model 4.3.....	51
Table 4.14: Granger-Causality test for fraud and the state of the economy model (full model) .....	53
Table 4.15: Granger-Causality test for Model 4.4.....	55
Table 4.16: Granger-Causality test for theft and the state of the economy model (full model).....	57
Table 4.17: Granger-Causality test for model 4.5 .....	58
Table 4.18: Cointegration test for crime rate and crime number model .....	61
Table 4.19: Granger-Causality test for Model 4.6.....	62
Table 4.20: Granger-Causality test for Model 4.7.....	62
Table 4.21: Summary table for ARIMA modelling .....	67
Table 5.1: Recursive window and out-of-sample forecasting period for multi-step-ahead-forecast ..	69
Table 5.2: In-sample forecasts comparison, 1992-2012 .....	70
Table 5.3: Out-of-sample comparison for number of people charged with non-traffic offence .....	71
Table 5.4: Out-of-sample forecasts for number of people charged with non-traffic offences: MAE, MSE and MPE for Equal-Weighted-Average .....	73
Table 5.5: Out-of-sample forecasts comparison for number of people charged with burglary.....	74
Table 5.6: Out-of-sample forecasts comparison for number of people charged with fraud .....	75

Table 5.7: Out-of-sample forecasts using Equal-Weighted-Average for number of people charged with fraud .....	76
Table 5.8: Out-of-sample comparison for number of people charged with theft.....	77
Table 5.9: Out-of-sample forecasts using Equal-Weighted-Average for number of people charged with theft.....	79
Table 5.10: Models used for final forecasts.....	80
Table 6.1: Average quarterly change and average annual change for each variable .....	86

# Chapter 1 Introduction

The crime downturn since 2011 and the economic recovery after the 2008 Global Financial Crisis have led many people to think about the relationship between crime and the state of the economy. The future trend of crime levels can be debatable as well. Based on information available, can crime levels keep dropping for another five years or even longer? The main objective of this study is to investigate the association between crime and the state of the economy in New Zealand. Forecasting crime is another objective of this study.

## 1.1 Background and purpose of this study

Both time series data and micro level data can be used to study the association between crime and macroeconomic indicators (Small & Lewis, 1996). Micro-level dataset has an advantage of capturing information about an individual offender. Research based on microdata can be used to study the impact on an individual offender when their personal and financial condition changes. More commonly, demographic variables such as age, ethnicity and education level can be modelled using micro-level dataset. For time series data, the focus in the past was to study the relationship between unemployment and crime. The unemployment rate has been treated as an index of total macroeconomic activity. The movement of the unemployment rate is coincident with business cycles, and a high unemployment rate indicates a lower level of production and consumer spending (Bober, 1968; Granger, 1980 as cited in Cantor & Land, 1985). This is the main reason why some researchers pick the unemployment rate to represent macroeconomic activity.

Small & Lewis' (1996) study used regional data from 1961 to 1987. The conclusion was that a long-term relationship between unemployment and crime exists in most of the regions in New Zealand, and test results support that unemployment is causing crime rather than crime is causing unemployment. A more recent study conducted in Canada in 2015 found a negative short-term relationship between property crimes and the unemployment rate. There was no long-term relationship found between the crime rate and the unemployment rate (Janko & Popli, 2015). It is not surprising to have a different result from various studies, but

how to explain the conclusion becomes important. Cantor & Land (1985) used criminal opportunity theory and criminal motivation theory to explain the association between unemployment and crime. Criminal opportunity theory explains how unemployment associated with a person who becomes a potential crime target victim. Criminal motivation theory explains that the offender's motivation to commit crime is because of unemployment. This study will focus on investigating the association between economically motivated crime and the state of the economy. Finding the cause of the crime is outside the scope of this study. In fact, the cause of crime is a complicated topic. It is hard to explain the recent fall in crime in New Zealand, and there is no agreed explanation so far (Ministry of Justice, 2012).

The aim of this study is to investigate whether there are associations between crime and the state of the New Zealand economy. Small and Lewis's (1996) unemployment and crime model will be reproduced at both regional and national levels. The purpose of this is to discover if the relationship still holds using data that are more recent. There are three key objectives in this study. The first objective is to rebuild Small and Lewis's (1996) model. The second objective is to discover the association between crime and the state of the New Zealand economy and understand which of the chosen macroeconomic variables relate to crime. The third objective is to forecast the number of people charged with burglary, fraud, theft and non-traffic offences, then compare the accuracy of the forecasts between different models. The performance of seasonal differencing and seasonal dummies will be compared too; those are the methods used to model seasonality.

Data used in this study came from Ministry of Justice<sup>1</sup>, Statistics New Zealand<sup>2</sup>, Reserve Bank of New Zealand<sup>3</sup> and The Treasury<sup>4</sup>. This study focuses on economically motivated crimes, which includes Theft and related offences (theft), Unlawful entry with intent/burglary,

---

<sup>1</sup> Number of people charged per quarter by offence type and region requested from Ministry of Justice at 2015.

<sup>2</sup> Population statistics used to calculate crime rate and regional unemployment rate was downloaded from Statistics NZ's 'infoshare'. Population statistics is under the subject 'Population Estimates' in 'infoshare'. Unemployment rate is classified under 'Work Income and Spending' then 'Household Labour Force Survey – HLF'

<sup>3</sup> Inflation, Real GDP growth rate, Unemployment rate and Household debt and disposable income ratio are available on Reserve Bank of New Zealand's website. URL: <http://www.rbnz.govt.nz/monetary-policy/inflation>

<sup>4</sup> Forecasting of Inflation, Real GDP growth rate, Unemployment rate are available at The Treasury website

break and enter (burglary) and Fraud, deception and related offences (fraud). The non-traffic offence is a good measurement for overall crime level. This will be studied as well. The quarterly number of people charged in the criminal justice system and crime rate per 1,000 populations are the two measurements used for crime.

$$\text{Crime rate} = \frac{\text{Quarterly number of people charged by offence type}}{\text{Quarterly population estimates}} \times 1000$$

Macroeconomic indicators indicate the state of the economy at any point in time. Those macroeconomic indicators can be Gross Domestic Product (GDP), unemployment rate, inflation, average income and private debt (Parry and Kemp, 2000, ch.1). Macroeconomics is a study of the economy as a whole, the aim of macroeconomics is to explain how economic change can affect households, individuals and markets. Study of macroeconomics may include inflation, economic growth and unemployment (Mankiw, 2001, ch.22). Real GDP growth rate, unemployment rate, inflation, personal average weekly income and household debt and disposable income ratio are the chosen macroeconomic indicators in this study to represent the state of the economy in New Zealand. Real GDP growth rate, unemployment rate, inflation are the three important macroeconomic indicators to measure economic growth, labour market condition and cost of living. Personal income is an important indicator to judge how a person is doing economically and household debt and disposable income ratio is a good measurement of private debt (Mankiw, 2001, ch.22 and ch.26). Table 1.1 lists all variables that will be used in this study.

Variable name	Variable Description	Code
unemployment_rate_all	Quarterly unemployment rate	U
unemployment_rate_alllog	Quarterly unemployment rate in logarithm form	LnU
real_gdp_growth	Annual real GDP growth rate in each quarter	GDP
inflation	Annual inflation rate in each quarter	I
avg_incomelog	Personal average weekly income in each quarter in logarithm form	LnW
Hld_debt_disp_income_ptlog	Quarterly household debt and disposable income ratio in logarithm form	LnD
non_traffic_qtrlog	Quarterly number of people charged with non-traffic offences in logarithm form	LnNTr
non_traffic_ratelog	Quarterly crime rate for non-traffic offences in logarithm form	LnNTn
Burglary_qtrlog	Quarterly number of people charged with burglary offence in logarithm form	LnBr
Burglary_ratelog	Quarterly crime rate for burglary offence in logarithm form	LnBn

Theft_qtrlog	Quarterly number of people charged with theft offence in logarithm form	LnTr
Theft_ratelog	Quarterly crime rate theft offence in logarithm form	LnTn
Fraud_qtrlog	Quarterly number of people charged with fraud offence in logarithm form	LnFr
Fraud_ratelog	Quarterly crime rate for fraud offence in logarithm form	LnFn
QTR1	Quarterly dummy variable, 1 means Quarter 1, 0 means Quarter 2-4	Q1
QTR2	Quarterly dummy variable, 1 means Quarter 2, 0 means Quarter 1, 3 4	Q2
QTR3	Quarterly dummy variable, 1 means Quarter 3, 0 means Quarter 1, 2, 4	Q3
QTR4	Quarterly dummy variable, 1 means Quarter 4, 0 means Quarter 1-3	Q4
EA	Dummy variable for 'Easter', 1 means the Easter falls in the quarter, 0 means Easter does not fall in the quarter.	EA

**Table 1.1: List of variables used in this study with variable names, variable descriptions and variable codes**

For regional analysis, Auckland, Nelson/Marlborough/West Coast, Gisborne/Hawke's Bay are the chosen regions. Regional data from Ministry of Justice are based on court region. Regional breakdown of the unemployment rate produced by Statistics New Zealand is done using data from the Household Labour Force Survey. The breakdown is based on regional council areas. Northland, Auckland, Nelson/Marlborough/West Coast, Gisborne/Hawke's Bay are the only four regions that can be matched between Ministry of Justice's crime statistics and Statistics New Zealand's regional unemployment rate. In fact, each one of the four chosen regions has their characters, and all of them have at least one factor much higher or lower than the national average. At 2013, Northland had the lowest average household income in New Zealand. Gisborne and Hawke's Bay had a large proportion of the population as Maori, and average household income and employment rate were lower than the national average. Auckland continued to be the biggest city and the economic powerhouse of New Zealand. Nelson, Marlborough and West Coast had a bigger proportion of the population aged over 65 than the New Zealand average (Ministry of Business, Innovation and Employment, 2014).

## 1.2 Structure of this study

The first part of this study is to remake Small & Lewis' (1996) study at both national and regional levels. Lag length of Vector Autoregressive (VAR) system will be chosen using information criteria such as Akaike's Information Criterion (AIC), Akaike's Information Criterion Correction (AICC), Bayesian Information Criterion (BIC) and Hannan-Quinn Criterion (HQC). Cointegration test using Johansen method tests cointegration between crime and unemployment rate based on selected lag length. Vector Error Correction Model (VECM) will replace VAR if cointegration is found between the unemployment rate and crime. Model diagnostics will be monitored for all chosen models. In this instance, model diagnostics checking includes test of autocorrelation and heteroscedasticity in error term, checking whether univariate error terms are white noise, checking cross-correlations between univariate error terms. Information criteria are used again to pick the best model if there is more than one model with good model diagnostics. Granger-Causality test and impulse response will be produced from the best model.

The second model will have crime variables and all chosen macroeconomic variables modelled using VAR or VECM. The methodology employed for the second model is the same as the first model. For the second model, each offence category has its own model. The number of people charged and crime rate by offence categories are included in the same model. Macroeconomic variables included in the second model are Real GDP growth rate, weekly average personal income, inflation, unemployment rate and household debt and disposable income ratio. The model includes all macroeconomic variables, the number of people charged and the crime rate for each offence category is called full model. Model with good diagnostics and the lowest information criterion will be chosen to perform causality test. Some of the macroeconomic variables can be treated as exogenous variables based on the result of Granger-Causality test and parameter estimates. The economic variables will be modelled as exogenous variable if their movement causes the movement of crime variables and not vice versa. Principal Component Analysis (PCA) will be used if there are more than one exogenous variable. The reason to use PCA is to remove correlation between macroeconomic variables. The model with reduced number of endogenous variables is called reduced model. Finally, Granger-Causality test and the impulse response from the best reduced model will be used to explain the relationship between crime and macroeconomic variables. The same as choosing the best full model, reduced models with good model



diagnostics will be compared by information criterion. The model with the lowest information criterion is the best reduced model.

The third model aims to explain the relationship between crime variables in different offence categories. Number of people charged with burglary, theft and fraud are endogenous variables. There are no macroeconomic variables involved in this model. A further model will be built using crime rate. Model three uses the same methodology as model one and model two. The purpose of this model is to discover if there are associations between crime variables.

The fourth model uses ARIMA model to forecast the number of people charged with burglary, theft, fraud and non-traffic offences. Macroeconomic variables will not be included in ARIMA models. ACF and PACF will be checked to determine the order of AR and MA terms. SAS uses BIC for its automatic ARIMA order selection procedure, this is another method to choose the order of ARIMA model. Same as VECM and VAR model, Dicky Fuller test will be used to test stationary for ARIMA model. All models with good diagnostics from model two to model four will be used to produce forecasts for the number of people charged with burglary, theft, fraud and non-traffic offences. MSE, MPE and MAE will be used for both in-sample and out-of-sample comparisons. The model with the best out-of-sample MSE, MAE and MPE will be used to produce a final forecast. An equal-weighted average will be used to combine forecast if MSE, MAE and MPE are similar between models.

Data used in this study are quarterly data with strong seasonal patterns showing in some variables. Based on criminal opportunity theory, public holiday could create more crime targets for offenders, especially for property crime. This is because there are many people travelling during the holiday period. Hence there are more unoccupied houses, which is a potential crime target for theft and burglary. Seasonal dummy variables and seasonal differencing are the two methods used in this study to model seasonality. Model one to model four described above will have both approaches applied, and the performance of the two seasonality modelling methods will be compared. 'Easter' effect will be modelled using 'Easter' dummy. Easter is the only public holiday that moves between quarters. Therefore, the seasonal dummies will not be able to model 'Easter' Effect. For this reason, an 'Easter' dummy is created and will be included in all models mentioned above.

## Chapter 2 Literature review

Small and Lewis's (1996) research studied the association between economically motivated crime and unemployment by using regional time series data between 1961 and 1987 from 15 regions in New Zealand. Burglary, car conversion, theft, receiving and fraud were four of the offence categories classified as economically motivated crime. The study focused on causality and cointegration relation between crime and unemployment. Cointegration test and Granger-Causality test are both performed using Vector Error Correction Model (VECM). Cointegrating Regression ADF test was used to test cointegration. The study suggested that crime and unemployment were linked in some way. Furthermore, the evidence was stronger that unemployment causes crime rather than crime causes unemployment. Cointegration test concluded that strong cointegration relationships between economically motivated crime and unemployment were found in most of the regions. It means that long-term relations between economically motivated crime and unemployment were common at the regional level.

There was a similar study carried out by Janko and Popli (2015) in Canada using annual national and regional data from 1979 to 2006. VECM was used to study both long-run and short-run relations between crime and unemployment rate. Cointegration was used to test the long-run relationship. Total crime, violence crime and property crime were modelled separately. Property Crime includes Breaking and entering, Frauds, Theft and Robbery. Crime rate per 100,000 population are used to measure crime, and all crime variables are in natural logarithms. Janko and Popli (2015) concluded that there is no long-run relationship between crime rate and unemployment rate. However, negative short-run relationships were found between some types of property crimes (Fraud and Robbery) and unemployment rate.

Cantor and Land (1985) studied the association between unemployment and crime in the US using annual data between 1946 and 1982. Homicide, rape, aggravated assault, robbery, burglary, larceny-theft and motor vehicle theft were included in the study. The purpose of the research was to test two hypotheses developed from criminal opportunity theory and criminal motivation theory. The first hypothesis was that the unemployment rate had a negative partial contemporaneous effect on de-trended variations in crime. Cantor and Land used the Cochrane-Orcutt Procedure and a distributed lag model in their study. Results were different

from different models. The total effect of unemployment on crime was small but statistically significant to five crime categories (Homicide, robbery, burglary, larceny-theft and motor vehicle theft). Property-related crime (robbery, burglary and larceny-theft) mainly have the positive partial effect to unemployment (Cantor & Land, 1985). The second hypothesis was that the movements of unemployment rate between years had a positive partial contemporaneous effect on the movement of crime rate one year later. In conclusion, the lagged effect of the unemployment rate and crime was weak.

In Cantor & Land's (1985) research, criminal opportunity theory and criminal motivation theory were used to explain the association between unemployment and crime. Criminal opportunity theory explains how unemployment is associated with a victim becoming a potential crime target. A crime target could be an individual or properties owned by an individual. Criminal motivation theory describes unemployment as a motivation to commit a crime for a potential offender. Criminal motivation theory suggests that unemployment could be one of the factors to drive crime. In addition, criminal opportunity and criminal motivation are prerequisites for a potential offender to commit a crime. Therefore, both criminal opportunity theory and criminal motivation theory are needed to explain the association between crime and unemployment (Cohen & Felson, 1979).

Ministry of Justice of New Zealand produces Justice Sector Forecast using a mixture of time series and simulation modelling. Justice Sector Forecast is based on assumptions which represent each stage an offender has to go through in the criminal justice system. Assumptions include numbers entering the court system (number of charging events), number of court disposals, proportion of people remanded in custody, average time spent on custodial remand, the proportion of individuals convicted, the proportion of convicted offenders sentenced in each sentence type, the length of sentence imposed and proportion of prison sentence served. Those assumptions are made and agreed by Justice Sector agencies such as New Zealand Police, Department of Corrections, Ministry of Justice and The Crown Law Office. The number of court disposals is a good measurement to reflect the number of charging events. It is the most similar forecast which can be used to compare with a forecast of the number of people charged produced by this study. Based on assumptions from 2012, a number of charging events would decrease 1.7% in 2012-2013, 1.3% reduction was predicted in 2013-2014 and 2014-2015. A bigger drop of 1.9% per year was projected from 2015-2016 onwards. For the number of court disposals forecasting, the downwards trend was forecasted

between 2013-2014 and 2021-2022. This long-term falling trend was different compared with the historical trend of court disposals. There was a flat trend from 1991 to 2002 and number of court disposals per year was around 140,000. A smaller increase was recorded between 2009 and 2010 and the number of court disposals peaked at 188,000 at 2010. Then, decreasing started from 2011. There is no agreed overarching explanation of fall in crime so far. However, some policy changes were responsible for some part of the crime reduction. Policing Excellence was one of the policies to reduce some of charging events, it intended to use alternatives to prosecution. This would lead to a lower number of court disposals and number of people charged by police (Ministry of Justice, 2012).

# Chapter 3 Statistical Methodology

It is common that only one variable is modelled as the dependent variable in Autoregressive Integrated Moving Average (ARIMA) models or Linear Regression Models. To explain the impact of macroeconomic variables on crime, ARIMA model or Linear Regression Model with single dependent variable need to treat crime variables as the dependent variable and macroeconomic variables as independent variables. A separate model is required to explain the effects of crime on macroeconomic variables. For this instance, each macroeconomic variable needs to be modelled as the dependent variable, and crime variables should be treated as independent variables. Vector Autoregressive (VAR) model is an extended version of ARIMA model. It is a multivariate model, which models the dynamic relationships between time series variables (Davidson & MacKinnon, 2004, ch. 13). All variables can be treated as endogenous variables in VAR models and model results can explain two-way causality relationships between crime and macroeconomic variables.

Macroeconomic variables are very important to this study, and it is not surprising if a long-term equilibrium exists between them. Cointegration is a term when long-term equilibrium is found between variables. The Johansen test is used to find cointegration between variables. First of all, lag length of VAR model will be chosen using information criteria such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Hannan–Quinn Information Criterion (HQC). Secondly, cointegration test will be performed based on chosen lag length of VAR. VAR model will convert to Vector Error Correction Model (VECM) if cointegration is found among variables. Otherwise, VAR will be used to fit the data.

Granger-Causality test can discover whether the movement of macroeconomic variables causes the movement of crime and vice versa. The test result can be used to identify whether a list of variables should be modelled as exogenous variables in VAR or VECM. Detailed information about VAR model, Cointegration, Granger-Causality test and VECM are available in later sections.

### 3.1 Stationary and Test for unit root

A variable  $Y_t$  is weakly stationary if all conditions listed below are met.

$$E(Y_t) = \mu < \infty \quad \forall t$$

$$Var(Y_t) = \sigma^2 < \infty \quad \forall t$$

$$Cov(Y_t, Y_{t-k}) = \sigma_k \quad \forall t \text{ where } k = 1, 2, 3, \dots$$

Those three conditions specify that mean, variance and autocovariance of the variable do not change over the time. Strict stationarity requires the joint probability distribution of  $Y_t$  and  $Y_{t-k}$  remains the same for all  $t$  and  $k$ . It is stronger than weak stationarity (Verbeek & Leuven, 2000, ch. 8). Both ARIMA and VAR models only require variables to be weakly stationary.

First differencing is usually enough to transform a variable from nonstationary to stationary but some variables may need to be differenced multiple times to become stationary. The process of differencing  $d$  times to achieve stationarity is called integrated of order  $d$  ( $I(d)$ ). In fact,  $I(d)$  is a result of having the number of unit roots equal to  $d$ . As a result, an  $I(0)$  variable is stationary. It means that there is no unit root in the variable (Davidson & MacKinnon, 2004, ch. 14). Dickey-Fuller tests are commonly used to test unit root in time series data. The hypotheses of the test are that data contains a unit root versus the series is stationary. The test includes models to test AR(1) model for random walk in three different cases. A random walk is a nonstationary time series process, which is defined as when  $\phi$  equals to one in all three cases of Dickey-Fuller test. Null hypothesis ( $H_0: \phi = 1$ ) and alternative hypothesis ( $H_1: \phi < 1$ ) are same for all three cases of Dickey-Fuller test. Three cases of Dickey-Fuller test are listed below:

Case 1: Test of random walk against AR(1):  $y_t = \phi y_{t-1} + u_t$

When  $\phi = 1$ ,  $y_t - y_{t-1} = \Delta y_t = u_t$

Case 2: Test of random walk against AR(1) with drift:  $y_t = \phi y_{t-1} + \delta + u_t$

When  $\phi = 1$ ,  $y_t - y_{t-1} = \Delta y_t = \delta + u_t$

Case 3: Test of random walk against AR(1) with drift and deterministic time trend:

$$y_t = \phi y_{t-1} + \delta + \lambda t + u_t$$

When  $\phi = 1$ ,  $y_t - y_{t-1} = \Delta y_t = \delta + \lambda t + u_t$

In all three cases  $u_t \sim IID(0, \sigma^2)$ ,  $y_0 = 0$ ,  $\delta$  is a drift term and  $t$  is a deterministic time trend. Both drift and deterministic time trend are not part of the Dickey-Fuller test. There are parameters allowed in the model (Brooks, 2002, ch. 7). The value of  $\phi$  will be different if the model includes drift or deterministic time trend, this can sometimes improve the estimation of  $\phi$  and therefore change the test result.  $\Delta y_t$  will not be stationary if drift or deterministic time trend are not constant in different periods. The difference of drift and deterministic time trend in various periods could include in  $u_t$ , which will make  $u_t$  fail to meet the three conditions of stationarity. In SAS, zero mean, single mean and trend are referred to as a test for random walk, random walk with drift and random walk with drift and deterministic time trend (SAS Institute Inc, 2014, ch.35).

## 3.2 ARIMA and linear regression model

ARIMAX or linear regression model can be another option if VECM or VAR model conclude that movement of macroeconomic variables causes movement of crime and not vice versa. Crime variable is the dependent variable, and all macroeconomic variables are independent variables if the ARIMAX or linear regression model is used to model the impact of macroeconomic variables on crime. This section will discuss and compare ARIMA model and different forms of linear regression models.

### Linear regression model

Equation 3.1 is a linear regression model written in matrix form,

$$Y_t = X_t \beta + U_t \quad (3.1)$$

Ordinary Least Square (OLS) can be used to estimate  $\beta$  where  $\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y$ .  $\hat{\beta}_{OLS}$  is a “Best Linear Unbiased Estimator” (BLUE) of  $\beta_{OLS}$  if all five assumptions of Gauss-Markov Theorem are met.

Those five assumptions are:

1.  $E(U_t) = 0 \quad \forall t$

This can be interpreted as the error term of the model has zero mean.

2.  $var(U_t) = \sigma^2 < \infty$

This means that the error term of the model is homoscedastic (i.e. the error term has constant variance).

3.  $cov(U_i, U_j) = 0, i \neq j$

This means that the error term of the model is not autocorrelated.

4.  $cov(U_t, X_t) = 0$

This means that the error term of the model is independent of the independent variables.

5.  $U_t \sim N(0, \sigma^2)$

Error term is normally distributed with mean equal to zero and variance equal to  $\sigma^2$ .

Linear regression model only allows contemporaneous relationships. It means that the changes of the dependent variable at current time  $t$  only modelled by movement of independent variables at current time  $t$ . Distributed lag model is an extension of the linear regression model, which allows lags of independent variables included in the model. It is a good model to try when the impact on independent variables do not affect the dependent variable immediately. However, coefficient estimates can be biased if the sample size is small (Brooks, 2002, ch.4).

$$Y_t = \beta_0 + X_t\beta_1 + X_{t-1}\beta_2 + \cdots + X_{t-j}\beta_{j+1} + U_t \quad (3.2)$$

Equation 3.2 is Distributed lag model with lags of  $X$  modelled until  $t-j$ . The  $\beta$  parameters can be estimated by OLS.

### **Cochrane-Orcutt procedure**

Consider linear regression model (Equation 3.1),  $\beta$  can be estimated using OLS.  $\hat{\beta}$  is not 'BLUE' if there is autocorrelation in the error term. Cochrane-Orcutt procedure can be used when autocorrelation has been discovered in the error term of a linear regression model. The purpose of Cochrane-Orcutt procedure is to transform both dependent and independent



variables in order to make them suitable to be estimated by OLS and make  $\hat{\beta}$  'BLUE'. There are three steps to estimate  $\beta$  using Cochrane-Orcutt procedure.

1. Use OLS to estimate  $\beta$  in Equation 3.1.
2. Model error term  $U_t$  using AR(1) model:  $\hat{U}_t = \hat{\rho}\hat{U}_{t-1} + \hat{V}_t$  (3.3)
3. Using  $\hat{\rho}$  to transform  $X_t$  and  $Y_t$ . Based on Equation 3.1 and Equation 3.3,

$$\hat{\rho}Y_{t-1} = \hat{\rho}X_{t-1}\beta + \hat{\rho}U_{t-1} \quad (3.4)$$

$$\text{Equation 3.1} - \text{Equation 3.4} = Y_t - \hat{\rho}Y_{t-1} = (X_t - \hat{\rho}X_{t-1})\beta + (U_t - \hat{\rho}U_{t-1})$$

Regression can be modelled using OLS based on new variable  $Y_t^*$  and  $X_t^*$

$$\text{where } Y_t^* = Y_t - \hat{\rho}Y_{t-1}, X_t^* = X_t - \hat{\rho}X_{t-1}, V_t = U_t - \hat{\rho}U_{t-1}$$

(Brooks, 2002, ch.4)

### ARIMA

Some time series variables have a strong correlation between the current value and its past value. It means that lags of the variable itself are needed to model its current value. This model is called Autoregressive (AR). ARIMA is a classic time series model, which combines autoregressive and moving average (MA) into the same model. MA indicates regression error correlated to its past value. 'Integrated' part of the ARIMA is a differencing process to transform variable from nonstationary to stationary (Davidson & MacKinnon, 1999, ch.3). Equation 3.5 transforms nonstationary variable  $Y'_t$  to a stationary variable  $Y_t$  by differencing of order  $d$ .

$$Y_t = (1 - B^d)Y'_t = Y'_t - Y'_{t-d} \quad (d > 0) \quad (3.5)$$

where  $B$  is the backshift or lag operator

Time series model such as ARIMA is usually a-theoretical. It is not based on any underlying theoretical model, because it forecasts the variable using their past values and possible current and the previous values of the error term (Brooks, 2002, ch.5). ARIMAX model is ARIMA with exogenous variables and can be seen as an extended version of the linear regression model. AR and MA are the components not included in linear regression model. As a result, it is fair to conclude that ARIMAX model is a combination of ARIMA model and linear regression model.

In this case, ARIMA (p, d, q) can be written as Equation 3.6 in ARMA form using already differenced variable  $Y_t$  (Equation 3.5). ARIMAX model becomes ARMAX model if the already differenced variable is used in the model. ARMAX is given by Equation 3.7.

$$Y_t = \alpha + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

$$\text{or } (1 - \phi_1 B - \dots - \phi_p B^p) \times (1 - B)^d Y'_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \quad (3.6)$$

$$Y_t = \alpha + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \beta_0 X_t + \beta_1 X_{t-1} + \dots + \beta_r X_{t-r} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (3.7)$$

ARMAX reduce to ARMA when all  $\beta$  equals to zero in Equation 3.7 (Newbold & Bos 1994, ch. 7 and ch. 9). In some cases, ARIMA model can have only certain lags of AR or MA significant. For example, ARIMA ((p), d, (q)) only have pth AR term and qth MA term significant. Therefore, ARIMA ((p), d, (q)) only has one AR term and one MA term. For ARIMA (p, d, q), it has a total of p+q items with the number of AR terms equal to p and q is the number of MA terms.

Some variables require both seasonal and first differencing to become stationary. At the same time, both regular and seasonal AR and MA terms are needed as well. Multiplicative seasonal ARIMA (p, d, q)(P, D, Q)<sub>s</sub> (Equation 3.8) is the model can have all terms described above, it is given by

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps}) \times (1 - B)^d (1 - B^s)^D Y'_t$$

$$= (1 - \theta_1 B - \dots - \theta_q B^q)(1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs}) \varepsilon_t \quad (3.8)$$

where  $Y'_t$  is the original variable before the first and seasonal differencing. Simple ARIMA (p, d, q)(P, D, Q)<sub>s</sub> can fit the data well in most of the cases because of multiple AR and MA terms involved. ARIMA (0, 1, 1)(0, 1, 1)<sub>s</sub> referred as Airline Model (Equation 3.9) is a good example,

$$Y_t = (\varepsilon_t - \theta_1 \varepsilon_{t-1}) - (\Theta_1 \varepsilon_{t-s} - \theta_1 \Theta_1 \varepsilon_{t-s-1}) \quad (3.9)$$

where  $Y_t = (1 - B)(1 - B^s)Y'_t = (Y'_t - Y'_{t-1}) - (Y'_{t-s} - Y'_{t-s-1})$  (Newbold & Bos 1994, ch. 7).

Maximum likelihood estimation can be used to estimate parameter from ARIMA model. Use Equation 3.6 as an example, let  $\beta = (\alpha \ \phi_1 \ \dots \ \phi_p \ \theta_1 \ \dots \ \theta_p)'$  be the  $p+q+1$  dimensional vector of the model parameters for ARMA (p, q). The likelihood function  $\mathcal{L}(\beta, \sigma_\varepsilon^2)$  and log-likelihood function  $\text{Ln}(\mathcal{L}(\beta, \sigma_\varepsilon^2))$  are as follows

$$\mathcal{L}(\beta, \sigma_\varepsilon^2) = \prod_{t=1}^n f(y_t | y_{t-1}, \dots, y_1)$$

$$\text{Ln}(\mathcal{L}(\beta, \sigma_\varepsilon^2)) = \sum_{i=1}^n \text{Ln}(f(y_t | y_{t-1}, \dots, y_1))$$

In this case,  $f(y_t | y_{t-1}, \dots, y_1)$  is the joint density function for all observations. Maximum likelihood estimation of  $\beta$  is the value that maximizes the log-likelihood function  $\text{Ln}(\mathcal{L}(\beta, \sigma_\varepsilon^2))$  (Shumway & Stoffer, 2011, ch.5; Davidson & MacKinnon, 2004, ch.10 & ch.13).

In conclusion, linear regression models data based on different probability theory. It focuses on the conditional probability distribution of dependent variable given independent variables. A multivariate model such as VECM or VAR model focuses on the joint probability distribution of endogenous variables which means the impact on multicollinearity between variables has been modelled in VECM and VAR model (Brooks, 2002, ch. 6). Multicollinearity occurs when correlations are found between independent variables. It is a big issue in Linear Regression Model and ARIMAX model, because  $\hat{\beta}$  is biased when multicollinearity occurs. Principal component analysis (PCA) is good method to remove correlation between variables. PCA uses an orthogonal transformation to convert a set of correlated variables into a set of linearly uncorrelated variables called principal components (Brooks, 2002, ch.4). Principal components will replace original variables if variables are modelled as independent or exogenous variables. Detailed information about PCA is available in section 3.5.

### 3.3 VAR and VECM Modelling

VECM or VAR model contains a number of univariate models and allows more than one endogenous variable. The number of univariate models in a VECM or VAR model is equal to the number of endogenous variables. In fact, VAR or VECM is a system regression model.

#### 3.3.1 VAR

The current value of each endogenous variable is modelled based on the previous value of itself and other endogenous variables (Brooks, 2002, ch.6). For example, VAR(k) contains endogenous variable  $y_{1t}$  and  $y_{2t}$  can be written as  $Y_t = B_0 + B_1 Y_{t-1} + B_2 Y_{t-2} + \dots + B_k Y_{t-k} + U_t$  where  $U_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$ ,  $E(U_t) = 0$ ,  $E(U_t U_t') = \Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$ ,  $E(U_t U_{t-k}') = 0$  ( $k \neq 0$ ). The matrix form of VAR(k) is

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \beta_{21} & \alpha_{21} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \beta_{12} & \alpha_{12} \\ \beta_{22} & \alpha_{22} \end{pmatrix} \begin{pmatrix} y_{1t-2} \\ y_{2t-2} \end{pmatrix} + \dots + \begin{pmatrix} \beta_{1k} & \alpha_{1k} \\ \beta_{2k} & \alpha_{2k} \end{pmatrix} \begin{pmatrix} y_{1t-k} \\ y_{2t-k} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \quad (3.10)$$

and the univariate form is

$$y_{1t} = \beta_{10} + \beta_{11}y_{1t-1} + \dots + \beta_{1k}y_{1t-k} + \alpha_{11}y_{2t-1} + \dots + \alpha_{1k}y_{2t-k} + u_{1t}$$

$$y_{2t} = \beta_{20} + \beta_{21}y_{1t-1} + \dots + \beta_{2k}y_{1t-k} + \alpha_{21}y_{2t-1} + \dots + \alpha_{2k}y_{2t-k} + u_{2t}$$

Let  $B = (B_0 \ B_1 \ \dots \ B_k)$ ,  $Y' = (y_{1t} \ y_{2t})$ ,  $Z_t = (1 \ Y_{t-1}' \ Y_{t-2}' \ \dots \ Y_{t-k+1}')$  and  $Z' = (Z_0 \ Z_1 \ \dots \ Z_{t-1})$ . B can be estimated as  $\hat{B} = Y'Z(Z'Z)^{-1}$  and  $\hat{B}$  is a conditional maximum likelihood estimator with estimated covariance matrix  $\hat{\Sigma}$ . For VAR(k) with n as the number of observations,  $\hat{\Sigma} = \frac{1}{n-k} \sum_{t=1}^n \hat{U}_t \hat{U}_t'$  (Shumway & Stoffer, 2011, ch.5).

One of the advantages of VAR model is that all variables can be endogenous variables. The model is built based on AR lag of itself and other endogenous variables. This can produce a better estimate in some case. However, VAR model with a higher number of endogenous variables may affect the accuracy of the parameter estimates due to low degrees of freedom. Degrees of freedom is calculated as follow,

Degrees of freedom = Sample size – Number of parameters to be estimated  
(Brooks, 2002, ch.6).

In this study, there are seven variables to be modelled in VAR model. For VAR(4), this means that there are 28 parameters that need to be estimated.

### 3.3.2 Cointegration and VECM

Two nonstationary variables are cointegrated if the linear combination of them is stationary. Long-term or equilibrium phenomenon exists if variables are cointegrated. The cointegration may become insignificant in the short-run, but it will return in the long-run (Brooks, 2002, ch.7). It is important to consider cointegration in time series modelling because it models long-term relationship. VECM models both long-term and short-term relationships. The Johansen test can be used to determine all possible linearly independent cointegrating relationships when there are more than two variables in a VAR model (Brooks, 2002, ch.7). Assuming there is a set of  $I(1)$  variables, VAR(k) can be written as

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_k y_{t-k} + u_t$$

where  $\beta_1, \beta_2 \dots \beta_k$  are  $g \times g$  matrices and  $y_t, y_{t-1} \dots y_{t-k}$  and  $u_t$  are  $g \times 1$  vectors.

VAR(k) needs to transform to VECM(k) to perform Johansen test (Brooks, 2002, ch. 7). VECM(k) can be written as

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \cdots + \Gamma_{k-1} \Delta y_{t-(k-1)} + u_t$$

where  $\Pi = (\sum_{j=1}^k \beta_j) - I_g$  and  $\Gamma_i = (\sum_{j=1}^i \beta_j) - I_g$ .

Rank of cointegration test is the rank of matrix  $\Pi$ , and it is the number of cointegrating relations among all endogenous variables. In other words, the rank of the cointegration test is the number of linear combinations which are stationary. The hypotheses of Johansen test are:

$$H_0: r = 0 \text{ vs. } H_1: 0 < r \leq g$$

$$H_0: r = 1 \text{ vs. } H_1: 1 < r \leq g$$

$$H_0: r = 2 \text{ vs. } H_1: 2 < r \leq g$$

.....

$$H_0: r = g-1 \text{ vs. } H_1: r = g.$$

$\lambda_{trace}$  is the test statistic for the Johansen test, which is given by  $\lambda_{trace}(r) = -T \sum_{l=r+1}^g \log(1 - \hat{\lambda}_l)$ . 'r' is the number of cointegrating vectors under ( $H_0$ ),  $\hat{\lambda}_l$  is the estimated value for the  $l$ th ordered eigenvalue from matrix  $\Pi$ .  $r$  is the rank of Cointegration test and  $g$  is the number of endogenous variables. The  $r$  value of the first non-rejection of the null hypothesis is the rank of matrix  $\Pi$ . For the Johansen test, null hypothesis will be rejected if  $\lambda_{trace}$  is larger than its critical value. SAS produces  $\lambda_{trace}$  and P value for Cointegration test using Johansen method. Therefore, null hypothesis is rejected at 95% confidence level if P value is less than 5%. Cointegration test using Johansen's Method is important before VAR or VECM modeling. Long-run coefficient matrix  $\Pi$  with rank equal to zero will make  $\Pi y_{t-k}$  equal to zero. Long-run coefficient matrix  $\Pi$  will be a full rank matrix if  $r$  equals to  $g$ . It means that  $y_t$  is stationary (Brooks, 2002, ch. 7). Both full ranked and zero ranked  $\Pi$  will convert VECM back to VAR model. VECM should be used if cointegration is found and rank is between one and  $g-1$ . VECM is an extended version of VAR model for a set of variables with cointegration relations. Cointegration test can produce a different result if a different order of VAR model is used. The lag length of the VAR model can be chosen using information criteria such as AIC, AICC, BIC or HQC. The lag length with the lowest information criterion will be the chosen lag length for VAR. Cointegration may need to be tested multiple times if more than one VAR lag length is selected.

### 3.3.3 Granger-Causality test

Parameter estimates can be used to check whether two variables are significant to each other. However, this can be difficult given a large number of parameters in VECM or VAR model. The conclusion from Granger-Causality test is whether the movement of variable  $y_{1t}$  causes the movement of variable  $y_{2t}$ , hence it is an easy way to summarise the relationship between two variables. Granger-Causality test is performed under the restriction of setting all lags of one variable to zero. Assuming there are two variables ( $y_{1t}$  and  $y_{2t}$ ) in VAR( $k$ ) (Equation 3.11), Table 3.1 lists all possible hypotheses for Granger-Causality test.

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \beta_{21} & \alpha_{21} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \beta_{12} & \alpha_{12} \\ \beta_{22} & \alpha_{22} \end{pmatrix} \begin{pmatrix} y_{1t-2} \\ y_{2t-2} \end{pmatrix} + \dots + \begin{pmatrix} \beta_{1k} & \alpha_{1k} \\ \beta_{2k} & \alpha_{2k} \end{pmatrix} \begin{pmatrix} y_{1t-k} \\ y_{2t-k} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \quad (3.11)$$

Hypothesis	Restriction
1. Lags of $y_{1t}$ does not cause the movement of $y_{2t}$	$\beta_{21} = \beta_{22} = \dots = \beta_{2k} = 0$
2. Lags of $y_{1t}$ does not cause the movement of $y_{1t}$	$\beta_{11} = \beta_{12} = \dots = \beta_{1k} = 0$
3. Lags of $y_{2t}$ does not cause the movement of $y_{1t}$	$\alpha_{11} = \alpha_{12} = \dots = \alpha_{1k} = 0$
4. Lags of $y_{2t}$ does not cause the movement of $y_{2t}$	$\alpha_{21} = \alpha_{22} = \dots = \alpha_{2k} = 0$

**Table 3.1: Granger-Causality test hypothesis and restriction for VAR(K)**

An F-test is used to test the hypotheses of Granger-Causality test, and the test statistic is

$$\text{F-Test Statistics} = \frac{\text{RRSS} - \text{URSS}}{\text{URSS}} \times \frac{T-k}{m}$$

where URSS is residual sum of squares from the unrestricted model, RRSS is residual sum of squares from the restricted model, T is the number of observations, m is the number of restrictions, k is the number of regressors in the unrestricted model. Granger-Causality test estimates each univariate model separately using OLS. First of all, Equation 3.11 will be estimated, and then URSS is calculated. Secondly, each univariate model will be estimated again with restrictions applied, and RRSS will be estimated (Brooks, 2002, ch.6). In this instance, Equation 3.11 transforms to Equation 3.12 once restriction for hypothesis 1 is imposed.

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{11} & \alpha_{11} \\ 0 & \alpha_{21} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \beta_{12} & \alpha_{12} \\ 0 & \alpha_{22} \end{pmatrix} \begin{pmatrix} y_{1t-2} \\ y_{2t-2} \end{pmatrix} + \dots + \begin{pmatrix} \beta_{1k} & \alpha_{1k} \\ 0 & \alpha_{2k} \end{pmatrix} \begin{pmatrix} y_{1t-k} \\ y_{2t-k} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \quad (3.12)$$

In fact, Granger-Causality test is the test to compare the unrestricted model against the restricted model using an F-test. The form of the restricted model is based on restrictions listed in Table 3.1.

### **Identify exogenous variables in VECM or VAR model**

VECM or VAR model can have more than one endogenous variable. Initially, all variables can be treated as endogenous variables, and this is a learning experience for researchers. Association between endogenous variables can be tested using Granger-Causality test and estimated parameters. Then Granger-Causality testing is used to identify a list of endogenous variables, which can be modelled as exogenous variables.  $y_{1t}$  will be strongly exogenous if Granger-Causality test shows that the movement of  $y_{1t}$  causes the movement of  $y_{2t}$  and not vice versa. In this case, lags of  $y_{1t}$  are significant in the univariate model of  $y_{2t}$  and lags of  $y_{2t}$  may not be significant in the univariate model of  $y_{1t}$ . As a result,  $y_{1t}$  is an exogenous variable

because lags of  $y_{1t}$  are needed to explain the current value of  $y_{2t}$  and not vice versa.  $y_{1t}$  and  $y_{2t}$  have bi-directional causality relationship if Granger-Causality test result shows that the movement of  $y_{1t}$  causes the movement of  $y_{2t}$  and vice versa. This indicates that lags of  $y_{1t}$  and  $y_{2t}$  are significant in each other's univariate model (Brooks, 2002, ch. 6). In this case, lags of  $y_{1t}$  are needed to explain the current value of  $y_{2t}$  and vice versa. For this reason, both  $y_{1t}$  and  $y_{2t}$  should be endogenous variables in VECM or VAR model. Granger-Causality test can be performed using 'proc varmax' procedure in SAS. The null hypothesis of the Granger-Causality test is that  $y_{1t}$  is influenced only by its lags. Lags of  $y_{2t}$  do not influence  $y_{1t}$  (SAS Institute Inc., 2014, ch.35).

### 3.3.4 Impulse Response

Impulse Response is a good method to test how variables respond to shocks to each of the variables in VECM or VAR models. It can test whether shocks to a given variable have a positive or negative effect on other variables, and how long it will take the effect to die away. A unit shock is applied to the error term of each univariate model, and this needs to express VAR model as Vector Moving Average (VMA) model. The shock will die away if the system is stable.

For VAR(k) model,  $Y_t = B_1 Y_{t-1} + B_2 Y_{t-2} + \dots + B_k Y_{t-k} + U_t$ . Impulse response of a unit shock to  $y_{1t}$  at time  $t=0$  is  $Y_0 = \begin{pmatrix} u_{10} \\ u_{20} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ . The effect to a unit shock to  $y_{1t}$  at  $t=1, 2, \dots$  are as follows:

$$Y_1 = B_1 Y_0 = \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \beta_{21} & \alpha_{21} \end{pmatrix} Y_0 = \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \beta_{21} & \alpha_{21} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \beta_{11} \\ \beta_{21} \end{pmatrix}$$

$$Y_2 = B_1 Y_1 + B_2 Y_0 = \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \beta_{21} & \alpha_{21} \end{pmatrix} \begin{pmatrix} \beta_{11} \\ \beta_{21} \end{pmatrix} + \begin{pmatrix} \beta_{12} & \alpha_{12} \\ \beta_{22} & \alpha_{22} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \beta_{21} & \alpha_{21} \end{pmatrix} \begin{pmatrix} \beta_{11} \\ \beta_{21} \end{pmatrix} + \begin{pmatrix} \beta_{12} \\ \beta_{22} \end{pmatrix}$$

$$\vdots$$

Using the same method, impulse response of a unit shock to  $y_{2t}$  at time  $t=0$  is  $Y_0 = \begin{pmatrix} u_{10} \\ u_{20} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ . The effect to a unit shock to  $y_{1t}$  at  $t=1, 2, \dots$  are as follows:

$$Y_1 = B_1 Y_0 = \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \beta_{21} & \alpha_{21} \end{pmatrix} Y_0 = \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \beta_{21} & \alpha_{21} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \end{pmatrix}$$



$$Y_2 = B_1 Y_1 + B_2 Y_0 = \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \beta_{21} & \alpha_{21} \end{pmatrix} \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \end{pmatrix} + \begin{pmatrix} \beta_{12} & \alpha_{12} \\ \beta_{22} & \alpha_{22} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \beta_{21} & \alpha_{21} \end{pmatrix} \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \end{pmatrix} + \begin{pmatrix} \alpha_{12} \\ \alpha_{22} \end{pmatrix}$$

$$\vdots$$

(Brooks, 2002, ch.6).

In this case, impulse response of  $y_{1t}$  to one unit shock to  $y_{1t}$  itself at time  $t = 1$  equals to  $\beta_{11}$ , and at time  $t = 2$  equals to  $\beta_{11}^2 + \alpha_{11}\beta_{21} + \beta_{12}$ . Impulse response of  $y_{2t}$  to one unit shock to  $y_{1t}$  at time  $t = 1$  equals to  $\beta_{21}$ , and at time  $t = 2$  equals to  $\beta_{21}\beta_{11} + \alpha_{21}\beta_{21} + \beta_{22}$ . Impulse response of  $y_{1t}$  to one unit shock to  $y_{2t}$  at time  $t = 1$  is  $\alpha_{11}$ , and equals to  $\beta_{11}\alpha_{11} + \alpha_{11}\alpha_{21} + \alpha_{12}$  when  $t = 2$ . Impulse response of  $y_{2t}$  to one unit shock to  $y_{2t}$  at time  $t = 1$  is  $\alpha_{21}$ , and at time  $t = 2$  is  $\beta_{21}\alpha_{11} + \alpha_{22} + \alpha_{21}^2$ .

### 3.4 Seasonality Modelling

Seasonal dummy and seasonal differencing are used in this study to deal with seasonality in data. In fact, the seasonal dummy will not remove seasonality in data, it will use seasonal dummies to model the impact of seasonality. However, seasonal differencing intended to remove seasonal pattern in data. In this case, seasonal differencing transform the data from a quarterly number of people charged to the annual changes in the number of people charged.

#### Seasonal Dummy variable

Using seasonal dummy variable is one way to model seasonality. This allows variables with seasonal patterns to modelled without seasonal differencing applied to the data.

In fact, adding dummy variables is a useful method to model all types of calendar effects (Brooks, 2002, ch. 9). ‘Easter’ effect is a good example. It will be modelled in this study by using ‘Easter’ dummy variable.

$$Y_t = \beta_0 + \beta_1 Q_1 + \beta_2 Q_2 + \beta_3 Q_3 + \beta_4 X_t + \varepsilon_t \quad (3.13)$$

Seasonal dummy variables will add seasonal impact into the model (Equation 3.13). For example, the impact on quarter one ( $Q_1$ ) will be  $\beta_0 + \beta_1$  and the impact on quarter four ( $Q_4$ )

will be  $\beta_0$ . VAR model can be adjusted by using restriction function in SAS. For variables that do not have seasonal patterns,  $\beta_1, \beta_2$  or  $\beta_3$  can be set to zero manually.

### **Seasonal differencing**

The seasonal differencing for quarterly data is the differencing between the value in the current period and the value four periods before the current period. It is a good way to remove seasonality from data.

$$(1 - B^4)Y_t = Y_t - Y_{t-4} \quad (3.14)$$

$$(1 - B^1)(1 - B^4)Y_t = (Y_t - Y_{t-4}) - (Y_{t-1} - Y_{t-5}) \quad (3.15)$$

where  $B$  is the backshift or lag operator

In fact, seasonal differencing transforms the data from quarterly data to annual differencing. Some variables are still nonstationary after seasonal differencing. In this instance, one way to achieve stationary is to apply both seasonal differencing and first differencing to the variable (Equation 3.15).

## **3.5 Principal Components Analysis**

Some macroeconomic variables are highly correlated, and a good example is an association between inflation and unemployment rate. The Phillips Curve explains the relationship between inflation and unemployment rate (Mankiw, 2000, ch. 33). The main reason to use Principal Components Analysis (PCA) in this study is to remove correlation between exogenous variables. The aim of PCA is to transform the original variables to a set of linearly uncorrelated variables called principal components. The number of principal components is equal to the number of original variables. Newly created principal components will replace original variables in the model. However, not all principal components are needed to retain most of the information measured by the original variables. Assume there are  $k$  variables. PCA can be calculated as follows,

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} \text{ and } Var(X) = \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_k^2 \end{pmatrix}$$

Principal components  $P_1, P_2$  up to  $P_k$  are the linear combinations of  $x_1, x_2$  up to  $x_k$ , given by

$$P_1 = e_{11}x_1 + e_{12}x_2 + \cdots + e_{1k}x_k$$

$$P_2 = e_{21}x_1 + e_{22}x_2 + \cdots + e_{2k}x_k$$

... ..

$$P_k = e_{k1}x_1 + e_{k2}x_2 + \cdots + e_{kk}x_k$$

where  $\sum_{j=1}^k e_{ij}^2 = 1 \forall i = 1, 2, 3 \dots k$ , and  $e_i = (e_{i1} \ e_{i2} \ \cdots \ e_{ik})'$  is an eigenvector of variance and covariance matrix  $\Sigma$ . Variance and covariance of principal components are

$$VAR(P_i) = VAR(e_{i1}x_1 + e_{i2}x_2 + \cdots + e_{ik}x_k) = \lambda_i = e_i' \Sigma e_i \quad \forall i = 1, 2, 3 \dots k$$

$$COV(P_i, P_j) = 0 \quad \forall i, j = 1, 2, 3 \dots k \text{ and } i \neq j$$

where  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$ .  $\lambda_i$  defined as an eigenvalue of variance and covariance matrix  $\Sigma$ , and it is the variance of  $P_i$ . The most significant principal component is the one with the largest eigenvalue. PCA is a purely mathematical exercise, it means that calculating principal components do not require any assumptions be made to the structure, distribution or other properties of the original variables.  $\phi_i$  is given by

$$\phi_i = \frac{\lambda_i}{\sum_{i=1}^k \lambda_i}$$

It measures the proportion of variation explained by each principal component. It is an important figure when deciding the number of principal components required for modelling. Assume there are  $m$  principal components chosen to be included in the model. Then,  $\phi_1 + \phi_2 + \cdots + \phi_m$  is the total variation explained by those  $m$  principal components. Information lost by not including all principal components will be minimum if  $\phi_1 + \phi_2 + \cdots + \phi_m$  is big enough or close to one. In this study, little information should be lost because of PCA hence the aim is to have all original variables to be well explained by chosen  $m$

principal components. Therefore, PCA is also a good way to reduce dimension for VAR or VECM given there are already large numbers of parameters that need to be estimated (Brooks, 2002, ch.4).

### 3.6 Model diagnostics

Based on Equation 3.10, univariate error terms in VECM or VAR model need to be white noise with homoscedasticity. The other important model diagnostic for VAR or VECM is that cross-correlation should not be significant between univariate error terms. For ARIMA model, error terms need to be white noise and homoscedastic too. A white noise error term will have zero mean and constant variance. It is an independently and identically distributed (IID) random variable (Verbeek & Leuven, 2000, ch. 8).

#### Durbin-Watson test

Durbin-Watson test (DW Test) only tests autocorrelation of each specific order with its previous order. For example, Durbin-Watson test can test autocorrelation between  $u_t$  and  $u_{t-1}$  or  $u_{t-1}$  and  $u_{t-2}$ , autocorrelation between  $u_t$  and  $u_{t-2}$  cannot be tested using Durbin-Watson test. Let  $\hat{u}_t = \rho \hat{u}_{t-1} + v_t$ , the DW test statistic that tests autocorrelation between  $u_t$  and  $u_{t-1}$  is  $DW = \frac{\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=2}^T (\hat{u}_t)^2}$  (Brooks, 2002, ch.4).

#### Breusch-Godfrey test

Breusch-Godfrey Test is used by SAS (2014, ch.35) to test AR disturbance in VAR or VECM's univariate error term. Consider univariate model of VAR(k) from Equation 3.10 (reproduced here):

$$y_{1t} = \beta_{10} + \beta_{11}y_{1t-1} + \cdots + \beta_{1k}y_{1t-k} + \alpha_{11}y_{2t-1} + \cdots + \alpha_{1k}y_{2t-k} + u_{1t} \quad (3.10)$$

$$u_{1t} = \rho_1 u_{1t-1} + \rho_2 u_{1t-2} + \cdots + \rho_p u_{1t-p} + v_{1t}, v_{1t} \sim N(0, \sigma_v^2) \quad (3.16)$$

Equation 3.17 is given by substituting Equation 3.16 into Equation 3.10

$$y_{1t} = \beta_{10} + \beta_{11}y_{1t-1} + \cdots + \beta_{1k}y_{1t-k} + \alpha_{11}y_{2t-1} + \cdots + \alpha_{1k}y_{2t-k} + \rho_1 u_{1t-1} + \cdots + \rho_p u_{1t-p} + v_{1t} \quad (3.17)$$

Auxiliary regression models  $\hat{u}_{1t}$  as dependent variable and all regressors from Equation 3.17 as independent variables. It is given by

$$\hat{u}_{1t} = \beta_{10} + \beta_{11}y_{1t-1} + \cdots + \beta_{1k}y_{1t-k} + \alpha_{11}y_{2t-1} + \cdots + \alpha_{1k}y_{2t-k} + \rho_1 \hat{u}_{1t-1} + \cdots + \rho_p \hat{u}_{1t-p} + v_{1t} \quad (3.18)$$

where  $v_{1t} \sim N(0, \sigma_v^2)$ .

The Breusch-Godfrey Test is testing autocorrelation of any order up to  $p$  using auxiliary regression. The null hypothesis is  $\rho_1 = \rho_2 = \cdots = \rho_p = 0$ , which means there is no autocorrelation in the univariate error term. The alternative hypothesis is that at least one of  $\rho_1, \rho_2, \dots, \rho_p$  is not zero and therefore there is serial correlation in the univariate error term. Lagrange Multiplier (LM) test statistic is equal to  $(T-p)R^2$  ( $T$  is number of observations) and it is asymptotically distributed as chi square with  $p$  degrees of freedom (Breusch, 1978; Brooks, 2002, ch.4; Godfrey, 1978). SAS reports test statistics and their p-values for Breusch-Godfrey Test for  $p$  equals to 1, 2, 3, and 4. The null hypothesis is that univariate error term has no autocorrelation (SAS Institute Inc, 2014, ch.35).

### **Ljung-Box Test**

SAS uses Ljung–Box test as the default to test whether the error term of an ARIMA model is white noise. The null hypothesis is that the ARIMA model's error term is white noise, with no autocorrelation (SAS Institute Inc, 2014, ch.7). The test statistic is

$$Q = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k} \quad \text{where } \hat{\rho}_k = \frac{\sum_{t=k+1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum_{t=1}^n \hat{\varepsilon}_t^2}.$$

$\hat{\rho}_k$  is the sample autocorrelation at lag  $k$ . 'm' is the number of lags being tested. 'n' is the number of observations.  $\hat{\varepsilon}_1$  to  $\hat{\varepsilon}_n$  are the estimated error terms from ARIMA model. In this case test statistic  $Q$  follows chi-square distribution with degrees of freedom equal to  $m$  (Ljung & Box, 1978).

### **Portmanteau Test**

For VAR or VECM, the portmanteau test is used to test cross-correlation between univariate error terms. The null hypothesis is that the error terms are uncorrelated (SAS Institute Inc, 2014, ch. 35). The portmanteau test is a joint test to test the first m cross-correlations are zero.  $Q_m$  is the test statistic for a portmanteau test up to lag m, it is calculated as follows:

$$Q_m = T^2 \sum_{l=1}^m (T-l)^{-1} \text{tr} \{ \widehat{\rho}_\varepsilon(l) \widehat{\rho}_\varepsilon(0)^{-1} \widehat{\rho}_\varepsilon(-l) \widehat{\rho}_\varepsilon(0)^{-1} \}$$

where  $\widehat{\rho}_\varepsilon(l) = \widehat{V}_\varepsilon^{-\frac{1}{2}} C_\varepsilon(l) \widehat{V}_\varepsilon^{-\frac{1}{2}}$ ,  $C_\varepsilon(l) = T^{-1} \sum_{t=l+1}^T \varepsilon_t \varepsilon_{t-l}'$ ,  $\widehat{V}_\varepsilon = \text{Diag}(\hat{\sigma}_{11}^2, \hat{\sigma}_{22}^2, \dots, \hat{\sigma}_{kk}^2)$  and

$\widehat{\rho}_\varepsilon(-l) = \widehat{\rho}_\varepsilon(l)'$ .  $C_\varepsilon(l)$  is residual term cross-covariance matrices.  $\widehat{\rho}_\varepsilon(l)$  is residual cross-correlation matrices. 'k' is the number of endogenous variables in VECM or VAR model. 'T' is the number of observations and 'p' is the order of VECM or VAR model.  $Q_m$  follows a chi-square distribution with degrees of freedom equal to  $k^2(m-p)$  (Hosking, 1980). The maximum value of 'm' is set to 20 in this study, and the minimum of 'm' equals p+1.

### **Test for Heteroscedasticity**

For VECM or VAR model, SAS uses Autoregressive Conditional Heteroscedastic (ARCH) disturbances test for testing heteroscedasticity in the error term of univariate model (SAS Institute Inc, 2014, ch. 35). Testing for ARCH disturbances is a LM Test based on ARCH(q) model. ARCH(q) can be written as

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_q \varepsilon_{t-q}^2$$

where  $h_t$  is equal to  $\sigma_t^2$ , it is the conditional variance at time t, and  $\varepsilon_t$  is the univariate error term at time t (Brooks, 2002, ch.8).

An ARCH(q) model is built on squared residuals from the univariate model in a VECM or VAR model. The null hypothesis of the LM test is  $\alpha_1 = \alpha_2 = \dots = \alpha_q = 0$ . The alternative hypothesis is that at least one of  $\alpha_1, \alpha_2, \dots, \alpha_q$  is not zero. The test statistic is equal to  $TR^2$ , it has asymptotically a chi square distribution with q degrees of freedom. T is number of observations and  $R^2$  is the coefficient of multiple correlation of the ARCH(q) model (Engle,

1982). SAS's ARCH disturbance test is based on the ARCH(1) model. The null hypothesis is that the univariate error term has equal covariance (SAS Institute Inc, 2014, ch. 35). In this case,  $h_t = \sigma_t^2 = \alpha_0$  if the null hypothesis is accepted. It means that the conditional variance at time  $t$  is a constant and hence there is no heteroscedasticity.

### **Information Criteria**

The lag length of VAR can be selected using information criteria such as AIC, AICC, BIC or HQC. Those are not indicators to judge the fitness of the model. An information criterion is an indicator for model selection for a given set of data. However, only models of the same type can be compared. Models that use a different set of variables cannot be compared by information criteria. The principle of selecting a model using information criteria is to choose the number of parameters (or lag length) which minimise an information criterion (Brooks, 2002, ch.5). Information criteria such as AIC, AICC, BIC or HQC are given by

$$AIC = -2\log(\mathcal{L}(\hat{\theta})) + 2k$$

$$AICC = -2\log(\mathcal{L}(\hat{\theta})) + 2k + \frac{2k(k+1)}{n-k-1} = AIC + \frac{2k(k+1)}{n-k-1}$$

$$BIC = -2\log(\mathcal{L}(\hat{\theta})) + k\log(n)$$

$$HQC = -2\log(\mathcal{L}(\hat{\theta})) + 2k\log(\log(n))$$

where  $\mathcal{L}(\hat{\theta})$  is the maximised value of a likelihood function, 'n' is number of observations, 'k' is number of parameters.  $-2\log(\mathcal{L}(\hat{\theta}))$  in AIC measures lack of model fit,  $2k$  in AIC measures the 'penalty' of increasing the number of parameters (or lag length) in the model. AICC converges to AIC when the number of observations is large. BIC was developed in order to find a consistent 'penalty' term. 'klog(n)' is the 'penalty' term for BIC, it is asymptotically unbiased and has variance tending to zero. HQC is more like BIC, but with a different 'penalty' term (Burnham & Anderson, 2002, ch.2 and ch. 6; Burnham & Anderson, 2004; Hannan & Quinn, 1979). There are arguments about which information criterion to use, this study will use all information criterion listed in this section to select VAR and ARIMA model lag length.

# Chapter 4 Results

This chapter describes the results from each model. Modelling diagnostics will be discussed in this chapter, together with results of Granger-Causality tests and impulse responses.

## 4.1 Model 1: Unemployment and crime

Remaking Small & Lewis's (1996) unemployment and crime model is a good starting point for this study. Small & Lewis's (1996) study used data from 1961 to 1987, and the conclusion was that the unemployment causes crime and not vice versa. Another more recent Canadian study conducted by Janko & Popli (2005) used data between 1979 and 2006. They concluded that the crime rate and the unemployment rate had a short-run relationship only. Both Small & Lewis and Janko & Popli did not cover the period of the recent global financial crises that started from 2008. The social and economic environment changed dramatically throughout the 1960s to early 2000s, so it is not surprising that these studies find different results.

### 4.1.1 National Level Analysis

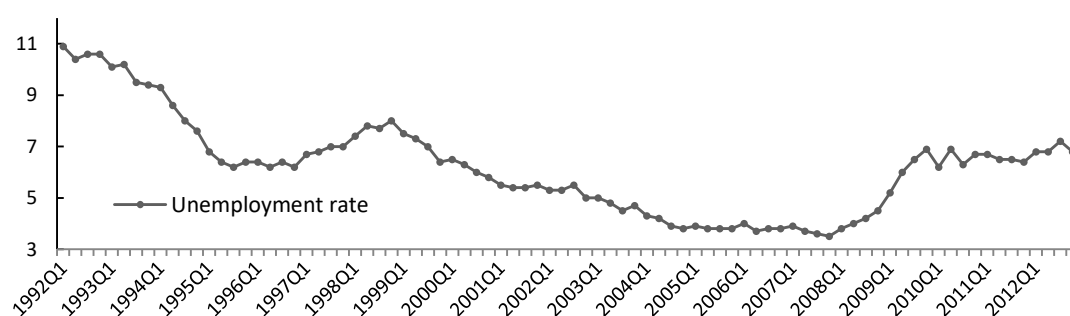
Crime variables and unemployment rate are in logarithmic form in this study. Logarithm can narrow the range of the variable, and it does not change the pattern of the original variable. It is a good way to deal with heteroscedasticity and outliers. Normally, logarithm applies to variables measured in positive dollar amount or variables measuring numbers of people. Variables in percentage can transform to logarithmic form too (Wooldridge, 2009, ch.6). It made sense for Small & Lewis (1996) to use logarithmic variables because they used the registered number of people unemployed as the measurement of the unemployment and number of reported offences as the measurement of crime.

According to Figure 4.1, the highest level of the unemployment rate was 10.9% in the first quarter of 1992, and the lowest was 3.5% in the fourth quarter of 2007. The unemployment rate was constantly at a low level of 4% and under between the third quarter of 2004 and the second quarter of 2008. From the second quarter of 2009 to the fourth quarter of 2012, the unemployment rate was above 6%. The unemployment rate was above 9% for the first

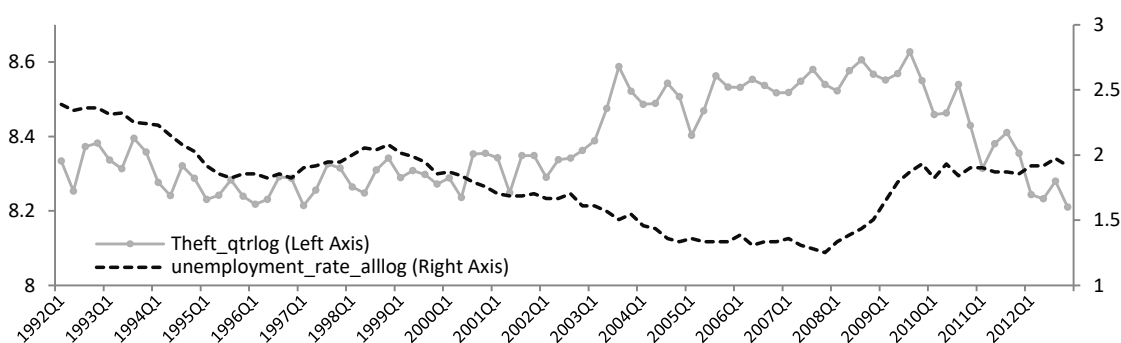


quarter of 1992 to the first quarter of 1994. In summary, the level of the unemployment rate between 1992 and 2012 is in a wide range. Therefore, it is sensible to apply natural logarithm to the unemployment rate to prevent heteroscedasticity. 'LnU' is unemployment rate in natural logarithmic form, which can be defined as the growth rate of the unemployment rate. First differencing of 'LnU' is the quarterly change of growth rate of the unemployment rate. It means that change of growth rate of the unemployment rate will be zero when the unemployment rate rises or drop at a constant rate. Unemployment rate rises or falls faster if the growth rate of the unemployment rate is increasing.

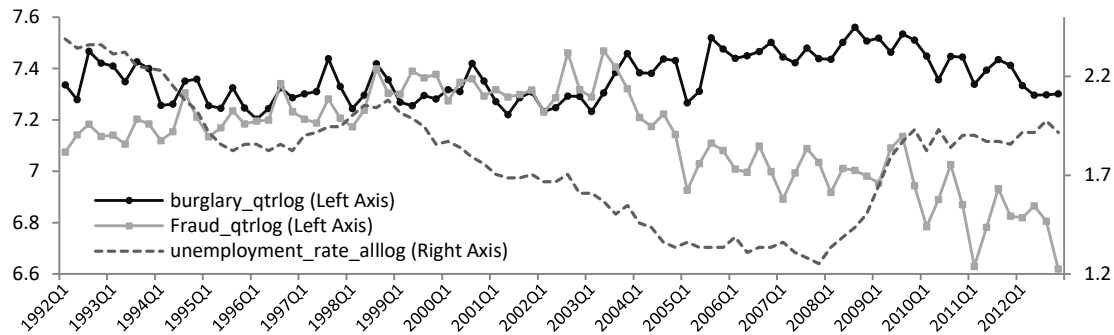
Based on Figure 4.2 and Figure 4.3, the number of people charged with Theft, Burglary and Non-traffic offences reached their peak between late 2008 and early 2009 and then dropped until 2012. However, the number of people charged with Fraud peaked at a much earlier date of quarter two in 2003 and then began to drop.



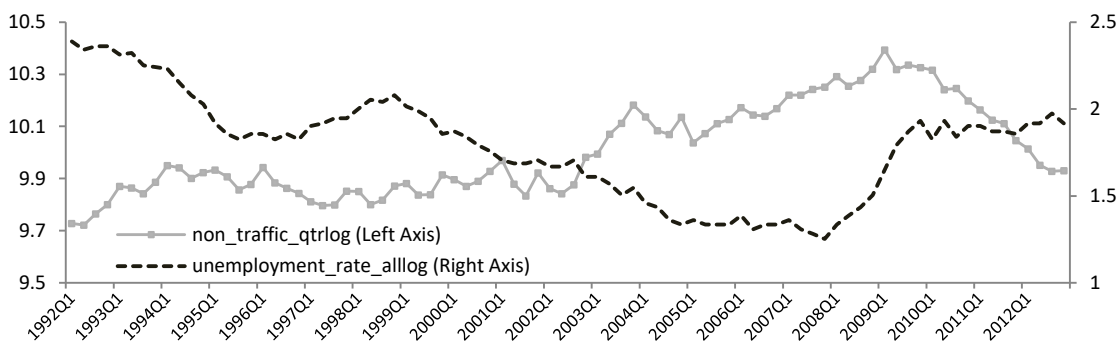
**Figure 4.1: Quarterly unemployment rate in New Zealand, 1992-2012**



**Figure 4.2: Quarterly unemployment rate and number of people charged with theft in logarithm form, 1992-2012**



**Figure 4.3: Quarterly unemployment rate and number of people charged with burglary and fraud in logarithm form, 1992-2012**



**Figure 4.4: Quarterly unemployment rate and number of people charged with Non-traffic offences in logarithm form, 1992-2012**

According to Figure 4.2 – Figure 4.4, ‘LnBn’, ‘LnTn’, ‘LnFn’ and ‘LnNTn’ had a mixture of increasing, decreasing or flat trends between 1992 and 2012. ‘LnTn’ was flat between 1992 and 2002, a minor increasing trend recorded from 2003 to 2008. Other than a different trend, ‘LnTn’ reached to a much higher level since 2004. Then, the decreasing trend started from 2009. ‘LnBn’ and ‘LnNTn’ had a similar trend as ‘LnTn’, they increased from 1992 to 2008 and then continued to fall since 2009.

‘LnU’ reached its lowest level between 1992 and 2012 in the fourth quarter of 2007. The lower level of ‘LnU’ between 2003 and 2008 are parallel with a higher level of ‘LnTn’, ‘LnBn’ and ‘LnNTn’ over the same period. High level of ‘LnU’ coexists with a lower level of ‘LnTn’, ‘LnBn’ and ‘LnNTn’ between 1992 and 2001. Historical trends recommend that ‘LnU’ and ‘LnTn’ moves in the opposite direction, hence their relationship seems to be negative. Figure 4.2 to Figure 4.4 suggest that ‘LnTn’ and ‘LnNTn’ have much stronger correlation to ‘LnU’ compared with ‘LnBn’ and ‘LnFn’.

	$\Delta(0)$	$\Delta(1)$	$\Delta(4)$	$\Delta(1,4)$
<b>LnTn</b>	Non stationary	Stationary	Non stationary	Stationary
<b>LnFn</b>	Non stationary	Stationary	Stationary	Stationary
<b>LnBn</b>	Non stationary	Stationary	Stationary	Stationary
<b>LnNTn</b>	Non stationary	Stationary	Non stationary	Stationary
<b>LnU</b>	Non stationary	Stationary	Non stationary	Stationary

**Table 4.1: Dickey-Fuller Unit Root Tests at a national level**

‘LnU’, ‘LnNTn’, ‘LnTn’, ‘LnBn’ and ‘LnFn’ look like nonstationary variables based on the different trends in different periods. Dickey-Fuller Unit Root Test (Table 4.1) shows that all four variables are stationary after first differencing. Seasonal patterns in ‘LnTn’, ‘LnFn’, ‘LnBn’ and ‘LnNTn’ can be identified from the graph. For ‘LnTn’, ‘LnFn’ and ‘LnBn’, it is higher in the third quarter and lower in the first and fourth quarters. The seasonal patterns in ‘LnNTn’ are different compared with the other crime variables, the first and fourth quarters are higher and the second and third quarters are lower. In fact, seasonal differencing cannot make ‘LnTn’ and ‘LnNTn’ stationary. Both first differencing and seasonal differencing are required for stationarity.

#### 4.1.1.1 Summary Result

Cointegration is important in VAR or VECM modelling, it tests whether a linear combination of a set of nonstationary variables is stationary. There are many factors that can change the result of cointegration test. For example, lag length, different variable combinations. Seasonal dummy variables and seasonal differencing are used to treat seasonality in data, and different methods could have different cointegration test results.

In this study, lag length of VAR model is chosen by information criteria such as AIC, AICC, BIC and HQC. In some cases, there is more than one lag length selected because of different results from different information criteria. For this reason, cointegration test needs to be performed multiple times by using different lag lengths, hence a set of variables can have different conclusions from different cointegration tests.

	Seasonal Dummy Model	Seasonal Differencing Model
<b>LnBn</b>	No	No
<b>LnFn</b>	No	No
<b>LnTn</b>	Yes, with P=4 (19.06%) <sup>5</sup>	Yes, with P=4 (6.4%)
<b>LnNTn</b>	No	Yes, with P=4 (6.45%), P=6 (5.21%)

**Table 4.2: Cointegration test result for unemployment and crime model at a national level**

Table 4.2 shows that cointegration relation is found between ‘LnTn’ and ‘LnU’ in models using both seasonal dummies and seasonal differencing. It means that both quarterly growth rate and annual growth rate of a number of people charged with theft have a long-term relationship with the growth rate of the unemployment rate. ‘LnNTn’ only cointegrates with ‘LnU’ in seasonal differencing models. It indicates that growth rate of the unemployment rate only has a long-term relationship with annual growth rate of number of people charged with non-traffic offences. As a result, VECM is needed for unemployment and non-traffic model, and unemployment and theft model because of cointegration relations between variables.

	Seasonal Differencing Model	Seasonal Dummy Model
<b>LnNTn</b>	VECM(4), VECM(6)** <sup>6</sup>	No good model
<b>LnTn</b>	VECM(4)	VECM(4)**
<b>LnFn</b>	VAR(4), VAR(5)	VAR(2)**
<b>LnBn</b>	No good model	VAR(4)**

**Table 4.3: List of unemployment and crime models with good model diagnostics at a national level**

Low lag VECM or VAR models are not good due to autocorrelation in univariate error terms in most cases. For crime and unemployment model at a national level, Table 4.3 shows that seasonal dummies performed better than seasonal differencing in terms of seasonality modeling. This conclusion is made because of lower information criterion recorded for models using seasonal dummies.

Seasonal differencing converts quarterly data to annual changes, then transformed variables are used in the model. Models using seasonal dummies are including three dummy variables as exogenous variables, it models the seasonality by adding same seasonal adjustments based on parameter coefficient estimates from seasonal dummy variables. On one hand, seasonal

<sup>5</sup> Percentage value in brackets is the P-value for cointegration test. All significant cointegration tests have rank equal to one because there are only two endogenous variables in unemployment and crime model.

<sup>6</sup> Model with \*\* is the best model picked by Information Criterion.

dummy could be an extra seasonal adjustment when an AR(4) term is included in the model. On the other hand, seasonal dummy could introduce seasonal patterns in error terms when endogenous variables do not have regular seasonal patterns. In this case, seasonal differencing could be a better way to model seasonality.

	Seasonal Differencing Model	Seasonal Dummy Model
<b>LnBn</b>	VAR(4): LnU -> LnBn (10.93%) <sup>7</sup>	-
<b>LnFn</b>	- <sup>8</sup>	-
<b>LnTn</b>	VECM(4): LnU-> LnTn (15.37%), LnTn -> LnU (6.77%)	VECM(4): LnU-> LnTn (1.3%)
<b>LnNTn</b>	VECM(6): LnNTn->LnU (3.16%), LnU-> LnNTn (6.69%)	-

**Table 4.4: Granger-Causality test result for unemployment and crime model at a national level**

Based on Table 4.4, results of Granger-Causality test are similar to the cointegration test result. Cointegration test implies that 'LnNTn' and 'LnTn' both have cointegration relationship with 'LnU'. Granger-Causality test suggests that movement of 'LnU' causes the movement of 'LnNTn', 'LnTn' and 'LnBn'. Movements of 'LnNTn' and 'LnTn' are causing the movements of 'LnU'. Fraud is the only offence category that has no relationship with the unemployment rate in both long-run and short-run. From parameter estimation (Appendix Table 3), none of the 'LnU' lags is significant to 'LnFn', and lags of 'LnFn' are not significant to 'LnU' either. A Granger-Causality test finds no causality relationship between 'LnU' and 'LnFn'. As a result, it is safe to conclude that growth rate of the unemployment rate and growth rate of number of people charged with fraud are not related.

Parameters estimate (Appendix Table 1 and Appendix Table 4) confirms that long-term relationship of 'LnNTn' and 'LnTn' with 'LnU' are negative. It means that more people charged with theft or non-traffic offences will reduce the growth rate of the unemployment rate. Short-term relationships are mainly insignificant, and the sign of coefficient estimates moves between positive and negative. In conclusion, not all crime variables have significant explanatory power to the unemployment rate in the short-run at the national level. There are many parameters estimated by VAR or VECM models, and finding a pattern using parameter

<sup>7</sup> Percentage value in brackets is the P-value for Granger-Causality test

<sup>8</sup> '-' indicates that there is no significant causality relations found, all P-value from Granger-Causality tests are greater than 20%

estimate can be difficult. Impulse response (in section 4.1.1.2) analysis is a good method to summarise the impact of variables on other variables in the same model.

#### 4.1.1.2 Impulse response

Impulse response summarises response to each endogenous variable from the shocks to itself and to the other endogenous variables in VECM or VAR models. Impulse responses are produced from the best model (listed in Table 4.3) selected by information criterion. Each variable requires a different length of time to digest the shocks. For VECM, impulse response of shocks will reach to an equilibrium level in the long-run because of the cointegration relationship modeled in VECM. For VAR, the model focuses on short-term relationship. In this case, cointegrations are found between 'LnTn' and 'LnU', 'LnNTn' and 'LnU'. Based on Figure 4.5, Figure 4.6 and Figure 4.7, impulse response of 'LnFn' to shocks to itself and 'LnU' reduced to zero in the long-term partly because no cointegration is found between 'LnFn' and 'LnU' and impact of shocks cannot stay in the system forever.



Figure 4.5: Impulse response of crime variables to one unit shock to 'LnU' from unemployment and crime models

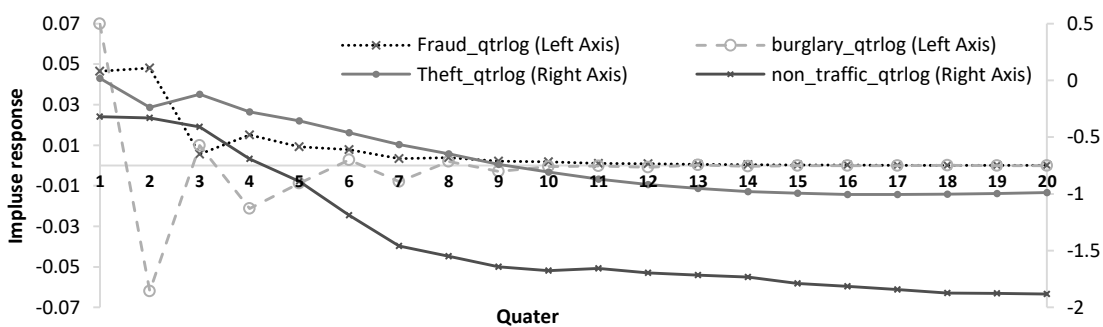
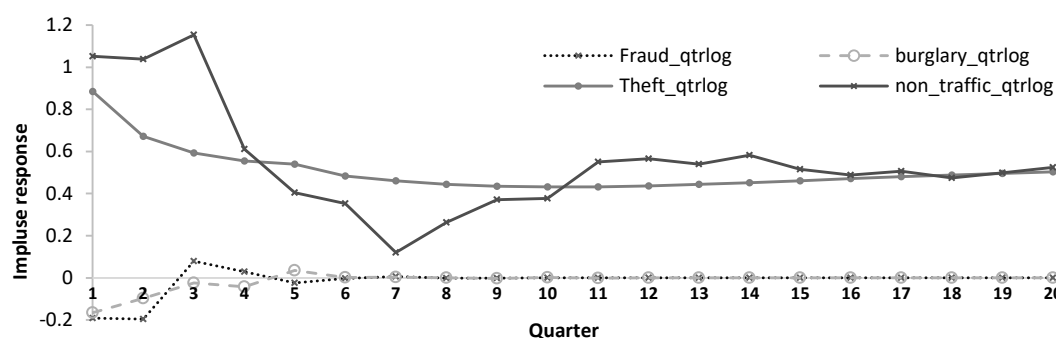


Figure 4.6: Impulse response of 'LnU' to one unit shock to crime variables from unemployment and crime models



**Figure 4.7: Impulse response of crime variables to one unit shock to itself from unemployment and crime models**

‘LnFn’ has negative impulse response of shocks to ‘LnU’ within five years since the shock. The highest level of impulse response of ‘LnFn’ to shocks to ‘LnU’ and ‘LnFn’ are both in the first two quarters. However, the size of impulse response to one unit shock is only about 0.35% of the average value of ‘LnFn’. Furthermore, ‘LnU’ has two years of positive impulse response caused by shocks to ‘LnFn’, then reached to almost zero nine quarters after the shocks. The largest impulse response (about 0.7% of average value of ‘LnFn’) to ‘LnFn’ is in the first two quarters after one unit shocks to itself.

At nine quarters after one unit shock to ‘LnU’, impulse response of ‘LnBn’ reduces to a negligible level. The highest level of impulse response is in the first two quarters after the shocks. ‘LnBn’ increases in the first two-quarters since the one-unit shock to ‘LnU’. Impulse response of ‘LnBn’ moves between negative and positive from the third to the fifth quarter after the shocks. However, impulse response of ‘LnU’ to a unit shock to ‘LnBn’ is a short life event. Compare with average value of ‘LnBn’, the only noticeable response comes from the first two quarters after the shock.

Impulse response of ‘LnTn’ is small and positive in the first two quarters to a unit shock to ‘LnU’. The size of impulse response to one unit shock is only about 0.1% of the average value for ‘LnTn’. After two quarters of increasing, impulse response of ‘LnTn’ starts to drop for ten consecutive quarters before it reaches a long-run persistent level of around -0.6. Impulse response of ‘LnNTn’ has a similar pattern to ‘LnTn’. There is a very small increase in the first quarter since the shock. For a unit shock to ‘LnU’, ‘LnNTn’ reaches to a constant level in negative at 13 quarters after the shock. Impulse response of ‘LnU’ to shocks to ‘LnTn’ and ‘LnNTn’ reaches a long-run equilibrium level of -1 and -1.8 at the 14th quarter.

Each crime variable has different pattern in the impulse response of shocks to itself. All crime variables have greater impulse response in the first four quarters since the shocks to itself. This reflects the autoregressive structure in the first four lags of all of four variables. Impulse responses of 'LnFn' and 'LnBn' have positive values between the third and the fifth quarter after the shock. To compare with average value of 'LnFn' and 'LnBn', size of impulse responses are too small to be considered significant. 'LnFn' and 'LnBn' have the highest level of impulse response of its own shock at the first quarter. For 'LnTn', shocks to itself have impulse response drop to a lower level after three-quarters. Then, impulse response reaches to persistent decreasing patterns in long-run from the eighth quarter after the shock. 'LnNTn' responded differently to shocks from itself. Impulse response peaks at the third quarters and then reaches to the lowest level at the seventh quarters after the shock. Impulse response is back to increasing trend from the eighth quarter to the eleventh, then long-run persistent decreasing starts from the twelfth quarter.

In conclusion, impulse responses of 'LnNTn' and 'LnTn' are much bigger and take longer to die away than 'LnBn' and 'LnFn'. This is consistent with the parameter estimates<sup>9</sup>, association between the unemployment rate and a number of people charged with fraud is the weakest compared with the association between unemployment rate and other crime variables. For crime variables, the size of impulse responses are much bigger to the shocks to itself rather than shocks to the unemployment rate.

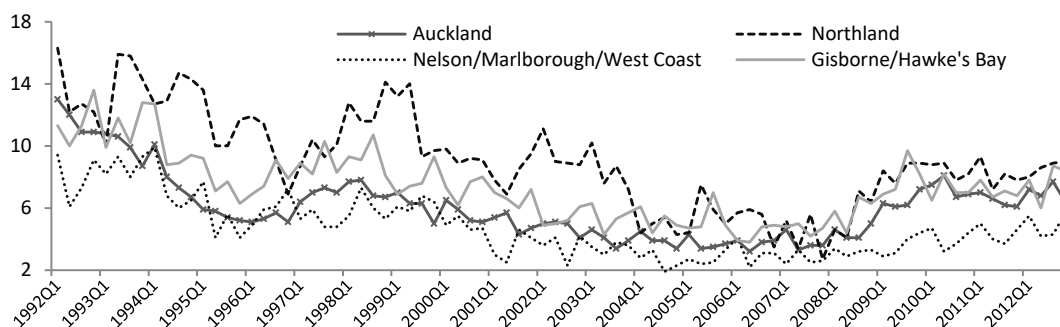
### 4.1.2 Regional Analysis

For all three offence categories, seasonal patterns in regional data are not as strong as at the national level. Auckland has stronger seasonal patterns than other regions. In most of the years, number of people charged is lower in the first and fourth quarters and higher in the second and third quarters for all offence categories. Other than Auckland, the number of people charged with fraud, theft or burglary did not have big fluctuations compared with the unemployment rate. Unemployment rate at regional level had a similar trend as at the national level. Gisborne/Hawke's Bay and Northland had much higher unemployment rate than Auckland and Nelson/Marlborough/West Coast between 1992 and 2012. It is not clear how unemployment rate and economically motivated crime related by just exploring graphs.

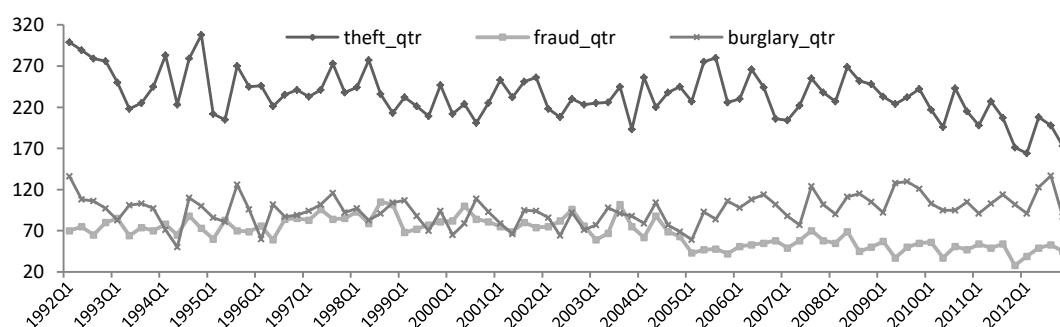
---

<sup>9</sup> Parameter estimates are available in Appendix A

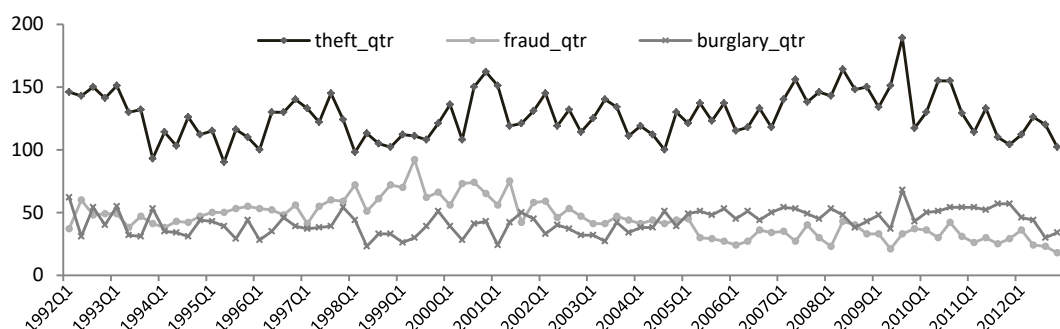




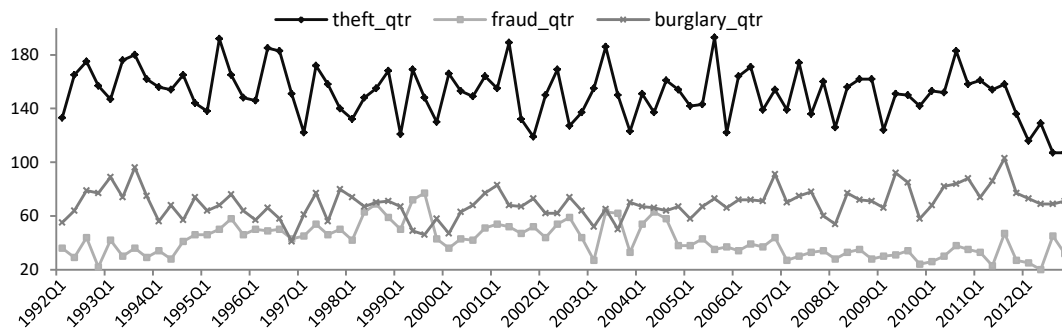
**Figure 4.8: Quarterly unemployment rate by region, 1992-2012**



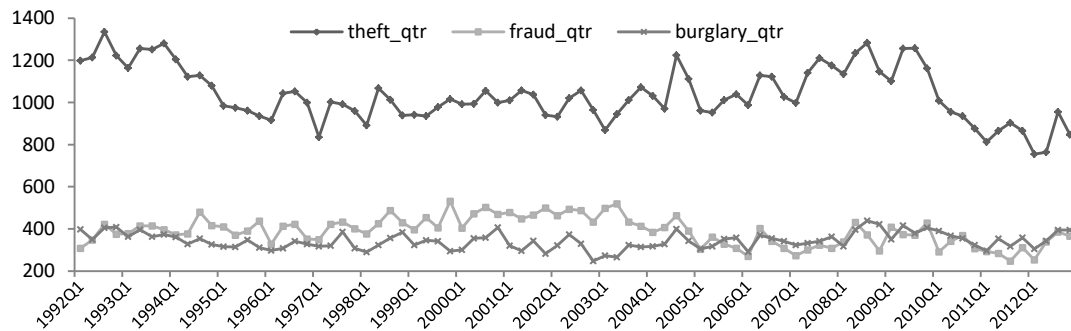
**Figure 4.9: Quarterly number of people charged by offence categories in Gisborne/Hawke's Bay, 1992-2012**



**Figure 4.10: Quarterly number of people charged by offence categories in Nelson/Marlborough/West Coast, 1992-2012**



**Figure 4.11: Quarterly number of people charged by offence categories in Northland, 1992-2012**



**Figure 4.12: Quarterly number of people charged by offence categories in Auckland, 1992-2012**

#### 4.1.2.1 Summary Result

At a regional level, the methodology used to test cointegration is the same as at the national level. The lag length of each model is picked using information criteria such as AIC, AICC, BIC and HQC, then the cointegration test is based on the selected lag length.

At national level (Table 4.2), cointegration relation exists between ‘LnTn’ and ‘LnU’. This is still the case in regional analysis (Table 4.5). ‘LnTn’ and ‘LnFN’ cointegrate with ‘LnU’ in Auckland and Northland. Cointegration relations are found between ‘LnBn’ and ‘LnU’ in all of the four regions. Auckland and Northland have much stronger cointegration between crime and unemployment compared with other regions. Number of people charged with burglary cointegrate with the unemployment rate in all regions. Therefore, it is fair to conclude that regional cointegration relationships between economically motivated crime and unemployment rate are stronger than at the national level.

		Burglary	Fraud	Theft
<b>Seasonal Dummy Model</b>	AKL <sup>10</sup>	p=1 (18.61%) <sup>11</sup>	-	p=1 (5.46%)
	NHL	p=4 (5.4%)	-	p=2 (5.04%)
	GH	p=1 (8.79%)	-	-
	NMW	- <sup>12</sup>	-	-
<b>Seasonal Differencing Model</b>	AKL	-	p=4 (6.18%)	p=4 (5.19%)
	NHL	p=4 (5.5%)	p=2 (5.64%) p=4 (5.49%)	p=4 (6.7%)
	GH	-	-	-
	NMW	p=7 (7.46%)	-	-

Table 4.5: Regional level cointegration test results for unemployment and crime model

		Theft	Fraud	Burglary
<b>Seasonal Dummy Model</b>	AKL	VAR(4)* <sup>13</sup>	VAR(2)*, VAR(4)	VAR(4)*
	NHL	VAR(4)*	No good model	VECM(4)*
	GH	VAR(4)*	VAR(2)*, VAR(4)	VAR(4)*
	NMW	VAR(4)*	VAR(2)*, VAR(4)	VAR(4)*
<b>Seasonal Differencing Model</b>	AKL	VECM (5)	VAR(5)	VAR(5)
	NHL	No good model	VECM(4)*	VAR(4)
	GH	VAR(4)	VAR(4)	VAR(4)
	NMW	VAR(4)	No good model	VAR(4), VECM(7)

Table 4.6: List of regional level unemployment and crime models with good model diagnostics

Lower lag order for VAR or VECM models is not good in some cases, often due to autocorrelation in univariate error terms. At national level analysis, seasonal dummy performed better than seasonal differencing, with models using seasonal dummies having lower information criteria values. The regional analysis has the same pattern (Table 4.6). Seasonal dummy variable is a better way to model seasonality in this case.

<sup>10</sup> AKL: Auckland, NHL: Northland, GH: Gisborne/Hawke's Bay, NMW: Nelson/Marlborough/West Coast

<sup>11</sup> Percentage value in brackets is the P-value for cointegration test. All significant cointegration tests have rank equal to one because there are only two endogenous variables in unemployment and crime model at a regional level

<sup>12</sup> '-' indicates that there is no significant cointegration relations found, all P-values from cointegration test are smaller than 5% for rank=1

<sup>13</sup> Models with \* is the best model picked by Information Criteria.

	Seasonal Differencing Model			Seasonal Dummy Model		
	Burglary	Fraud	Theft	Burglary	Fraud	Theft
<b>AKL</b>	-	-	VECM(5) LnU->LnTn (7.2%)	-	-	-
<b>NHL</b>	_ <sup>14</sup>	VECM(4) LnU->LnFn (0.51%)	-	-	-	VAR(4) LnU->LnTn (8.23%)
<b>GH</b>	-	-	-	-	-	-
<b>NMW</b>	VECM(7) LnU -> LnBn (1.53%) <sup>15</sup> LnBn-> LnU (8.06%)	-	-	-	-	-

**Table 4.7: Granger-Causality test for unemployment and crime model at a regional level**

Based on Table 4.7, the Granger-Causality test has similar results as cointegration test. At a national level, movements of ‘LnTn’ and ‘LnBn’ are caused by movement of ‘LnU’. At a regional level, movements of ‘LnU’ cause the movements of ‘LnTn’ in Auckland and Northland. These results are the same as the conclusion from Small and Lewis (1996). Bi-directional causality is found between ‘LnBn’ and ‘LnU’ in Nelson/Marlborough/West Coast. In Northland, movement of ‘LnU’ causes movement of ‘LnFn’. Conclusions from this study are the same as Small & Lewis’s conclusion. In summary, the regional causality relationship between unemployment rate and crime is that unemployment rate causes economically motivated crimes and not vice versa. In Gisborne/Hawke’s Bay, no significant causality between unemployment rate and economically motivated crime has been found. Easter is not significant to most of the models at the national level analysis. The insignificant impact of ‘Easter’ is still the case at the regional level analysis. ‘Easter’ is significant to ‘LnTn’ in Gisborne/Hawke’s Bay only.

<sup>14</sup> ‘-’ indicates that there is no significant causality relations found, all P-value from Granger-Causality test are greater than 20%

<sup>15</sup> Percentage value in brackets is the P-value for Granger-Causality test

## 4.2 Model 2: Crime and the state of the economy

Unemployment rate is an important macroeconomic indicator. However, other variables such as average income, household debt and disposable income ratio, inflation, real GDP growth rate are also needed to give a big picture about the state of the New Zealand economy. Therefore, it is important to include all major macroeconomic variables in VECM or VAR model in order to capture the movement of New Zealand economy. All macroeconomic variables move systemically. It means that the movement of one variable is caused by changes in other variables. For example, higher unemployment rate could cause average income to drop and then household debt and disposable income ratio may go up. As a result, financial condition may come under pressure for some households. The relationship between unemployment rate and inflation is another well-known macroeconomic theory, explained by the Philipps curve. The Philipps curve concludes that higher unemployment rate is correlated with lower inflation, and lower unemployment rate is correlated with higher inflation (Mankiw, 2000, ch. 33).

Burglary, theft and fraud are analysed separately in this case. There are seven endogenous variables in each model. Number of people charged and the crime rate are modelled in the same model. Seasonal dummy variables and seasonal differencing are the two different methods used to model seasonality. Logarithms are applied to both quarterly number of people charged and crime rate.

Personal weekly average income (average income) household debt and disposable income ratio, real GDP growth rate, inflation, unemployment rate are the macroeconomic variables included in the model. Real GDP growth rate, unemployment rate and inflation are measured as percentages. In most cases, the values of real GDP growth rate, unemployment rate and inflation are small and real GDP growth rate has some negative values. Therefore, they will not have logarithms applied. Average income and Household debt and disposable income ratio have a long-term increasing trend. Household debt and disposable income ratio is measured as a percentage, and the value is above 100% in recent years. Average income is the personal weekly income averaged each quarter. Therefore, both variables are transformed to logarithmic form, the first or seasonal differenced logarithmic variables can be explained as quarterly or annual growth rates.

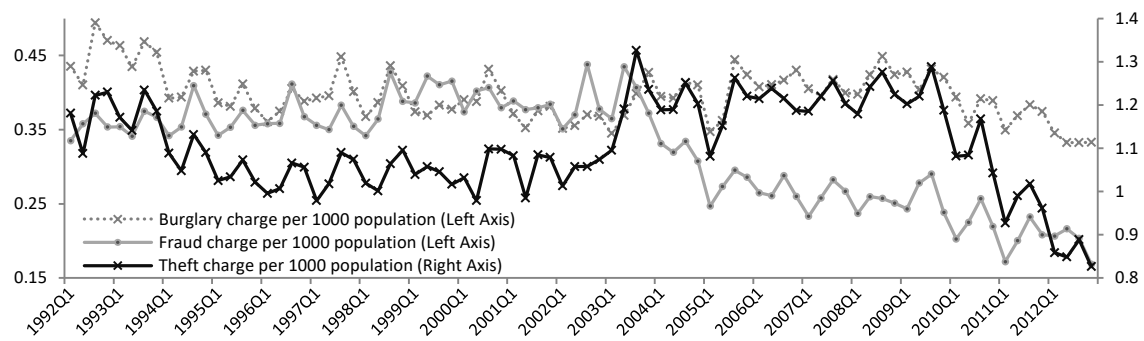


Figure 4.13: Quarterly crime rate by offence type, 1992-2012

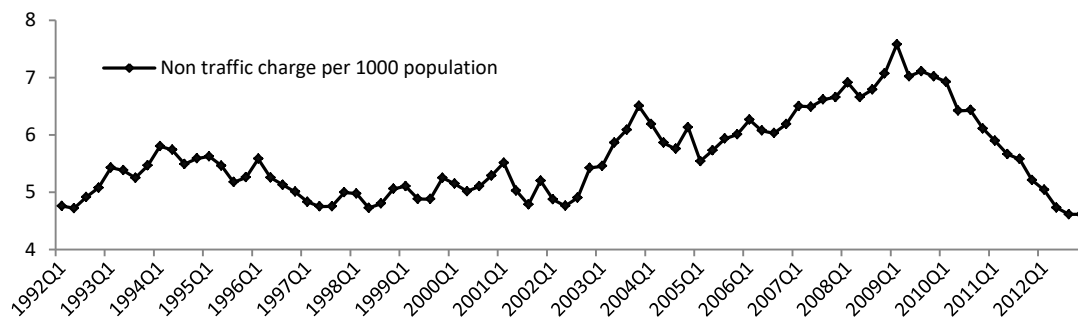


Figure 4.14: Quarterly crime rate for non-traffic offences, 1992-2012

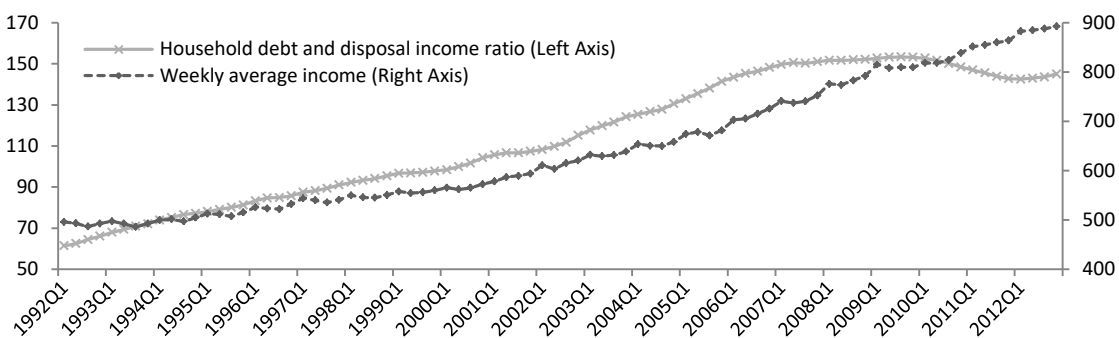


Figure 4.15 Weekly personal average income and Household debt and disposable income ratio, 1992-2012

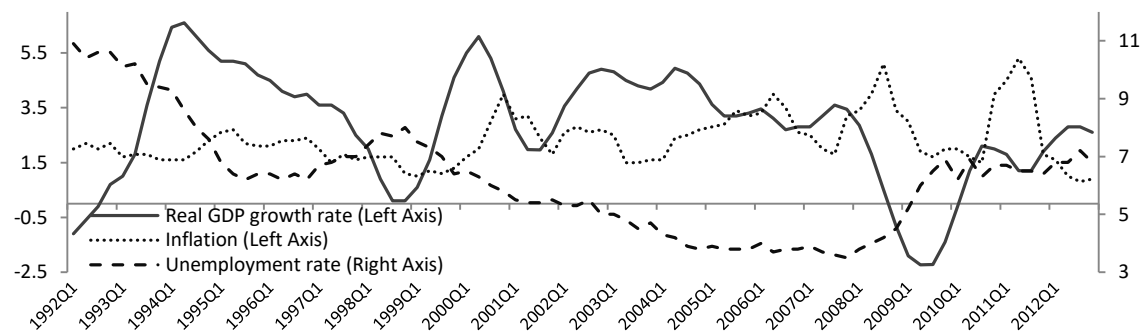


Figure 4.16: Quarterly real GDP growth rate, inflation and unemployment rate, 1992-2012

It is difficult to see the relationship between macroeconomic variables and crime by just exploring graphs. Household debt and disposable income ratio changed from increasing trend to a flat trend from 2007 and then started to drop from 2010. This is when the crime rate and quarterly number of people charged started to drop. At the same period, Real GDP growth rate, inflation and unemployment rate all started to recover from the 2008 global financial crisis. Therefore, there is a moderate level of association between crime rates (or the number of people charged) and the state of the economy in New Zealand by just analysing trends from graphs of macroeconomic variables and crime variables.

Crime and the state of the economy model uses the same methodology as crime and unemployment model. First of all lag lengths of VAR models are chosen based on information criteria. In this case, all macroeconomic variables, the number of people charged per quarter and quarterly crime rate are modelled in the same model. Secondly, Johansen cointegration test will be performed using selected lags from the step one. VECM is used if significant cointegration relations are found. The final step is to pick models with good model diagnostics. Granger-Causality test and parameter estimates from those good models are used to determine a list of variables that could be treated as exogenous. Weekly average income has a regular seasonal pattern. The highest average weekly income is recorded in the first quarter and the lowest is in quarter three. The same trend and seasonal patterns are found between crime rate and number of people charged by offence type. For this reason, seasonality modelling is necessary and both seasonal differencing and seasonal dummies will be used.

It is difficult to produce a good model when there are many endogenous variables. The common issues are autocorrelation and heteroscedasticity in univariate error terms. There are correlations between univariate error terms in every model. It means that VAR or VECM is not a good multivariate model. However, each univariate model in VAR or VECM can still be a good model. Therefore, correlations between univariate error terms should not be a reason to disregard Granger-Causality test results and parameter estimates.

#### **4.2.1 Non-traffic and the state of the economy model**

Non-traffic offences include number of people charged with all other offence categories except traffic offences. It includes some of the serious offence categories such as Homicide, Drug offences and sexual offences. It is a good variable to measure the overall crime level in

New Zealand. There are cointegrating relations among crime and macroeconomic variables for all chosen VAR lag lengths. For unemployment and crime model, only two endogenous variables are included, and the rank of the cointegration test can only be one. It means that there is only one cointegration relation between endogenous variables. However, seven variables have been modelled in this section. Ranks of the cointegration test are different when different lag lengths are used. None of the VECM models is a good model because of autoregressive structure or heteroscedasticity in univariate error terms. VAR(4) using seasonal differencing is the only good model produced in this section. Each univariate model is good from VAR(4), but cross-correlation between univariate errors is the issue. ‘Easter’ is removed from the model because it is not significant in all endogenous variables.

	LnNTn	LnNTr
<b>LnW</b>	-. <sup>16</sup>	-
<b>LnD</b>	LnD->LnNTn (4.58%) <sup>17</sup> LnNTn->LnD (12.43%)	LnD->LnNTr (4.15%) LnNTr->LnD (11.83%)
<b>I</b>	-	-
<b>GDP</b>	-	-
<b>U</b>	-	-

**Table 4.8: Granger-Causality test from VAR(4)<sup>18</sup> for Non-traffic and the state of the economy model**

Based on Granger-Causality test result in Table 4.8, all macroeconomic variables should be exogenous variables except ‘LnD’ (natural logarithm of household debt and disposable income ratio). This is because the only variable that has causality relations with crime variables is ‘LnD’. The New Zealand Treasury forecast inflation, unemployment rate and real GDP growth rate. Those forecasts are used in this study when inflation, unemployment rate and real GDP growth rate are modelled as exogenous variables. Weekly average income and household debt and disposable income ratio do not have forecasts available to use. As a result, ‘LnNTn’, ‘LnNTr’, ‘LnW’ and ‘LnD’ are endogenous variables in reduced model and ‘U’, ‘GDP’ and ‘I’ are exogenous variables.

<sup>16</sup> ‘-’ indicates that there is no significant causality relations found, all P-value from Granger-Causality test are greater than 20%

<sup>17</sup> Percentage value in brackets is the P-value for Granger-Causality test

<sup>18</sup> VAR (4) is non-traffic and the state of the economy model using seasonal differencing



### **Reduced Model**

Principal Component Analysis is used in order to remove correlation between exogenous variables. Based on Table 4.9, unemployment rate and inflation are highly correlated with the first principal component (prin1). Eigenvectors show that unemployment rate and inflation moves in an opposite direction. This is the same as the relationship explained by the Phillips Curve. High unemployment rate is associated with low inflation and low unemployment rate is associated with high inflation. Thus, the 'prin1' explains the combined effects of unemployment rate and inflation. Real GDP growth rate, which measures economic growth, is the main variable in the second principal component (prin2). In total, the first two principal components explained 81.85% of the variation. Therefore, prin1 and prin2 are enough to explain 'U', 'GDP' and 'I'. 'prin1' and 'prin2' are replacing 'U', 'GDP' and 'I' to be modelled as exogenous variable. In this case, principal component analysis removed correlation between exogenous variables, and reduced number of exogenous variables from three to two.

<b>Pearson Correlation Coefficients</b>			
	<b>Prin1</b>	<b>Prin2</b>	<b>Prin3</b>
<b>U</b>	-0.859	-0.082	0.506
<b>GDP</b>	0.305	0.911	0.277
<b>I</b>	0.761	-0.458	0.460

**Table 4.9: Correlation between principal components and original variables**

<b>Eigenvectors</b>			
	<b>Prin1</b>	<b>Prin2</b>	<b>Prin3</b>
<b>U</b>	-0.723	-0.080	0.686
<b>GDP</b>	0.257	0.891	0.375
<b>I</b>	0.641	-0.448	0.624

**Table 4.10: Eigenvectors for PCA for 'U', 'GDP' and 'I'**

In reduced model, cointegration test is significant in VECM(1). However, VECM(1) is not a good model due to autoregressive structure in univariate error term. Models with extra lag length such as VAR(4) remove autoregressive structure in univariate error terms even though cointegration becomes insignificant. Therefore, cointegration test result using  $P=1$  may not be accurate. As for the full model, all reduced models still have problems with cross-correlation between univariate error terms. Reduced model VAR(4) is given by

$$\begin{pmatrix} LnNTr_{1,t} \\ LnNTn_{2,t} \\ LnD_{3,t} \\ LnW_{4,t} \end{pmatrix} = A_0 + A_1 \begin{pmatrix} LnNTr_{1,t-1} \\ LnNTn_{2,t-1} \\ LnD_{3,t-1} \\ LnW_{4,t-1} \end{pmatrix} + \dots + A_4 \begin{pmatrix} LnNTr_{1,t-4} \\ LnNTn_{2,t-4} \\ LnD_{3,t-4} \\ LnW_{4,t-4} \end{pmatrix} + (C)(EA_t) + (D) \begin{pmatrix} prin1_t \\ prin2_t \end{pmatrix} + (E_t) \quad (4.1)$$

where  $A_0$  and  $C$  are  $4 \times 1$  vectors,  $A_1$  to  $A_4$  are  $4 \times 4$  matrices,  $D$  is a  $4 \times 2$  matrix.

Model 4.1 is the only model with good univariate models. Based on Granger-Causality test results, ‘LnD’ has bi-directional causality relation with ‘LnNTr’ and ‘LnNTn’. Movements of ‘LnW’ are not causing the movements of ‘LnNTr’ or ‘LnNTn’ and vice versa. All of the significant parameters to ‘LnW’ are lags of itself. Lags of ‘LnW’ are not significant to other endogenous variables. As a result, ‘LnW’ is removed from the model. VAR(1, 2, 4) is the second reduced model for economic and non-traffic model, ‘LnNTr’, ‘LnNTn’ and ‘LnD’ are endogenous variables. ‘Easter’, ‘prin1’ and ‘prin2’ are exogenous variables. VAR(1, 2, 4) is given by

$$\begin{pmatrix} LnNTr_{1,t} \\ LnNTn_{2,t} \\ LnD_{3,t} \end{pmatrix} = A_0 + A_1 \begin{pmatrix} LnNTr_{1,t-1} \\ LnNTn_{2,t-1} \\ LnD_{3,t-1} \end{pmatrix} + \dots + A_4 \begin{pmatrix} LnNTr_{1,t-4} \\ LnNTn_{2,t-4} \\ LnD_{3,t-4} \end{pmatrix} + (C)(EA) + (D) \begin{pmatrix} prin1_t \\ prin2_t \end{pmatrix} + (E_t) \quad (4.2)$$

where  $A_0$  and  $C$  are  $3 \times 1$  vectors,  $A_1$  to  $A_4$  are  $3 \times 3$  matrices,  $D$  is a  $3 \times 2$  matrix.

No cointegration relation is found in Model 4.2 for all chosen lag lengths. For this reason, VECM is not necessary. There is no cross-correlation between univariate error terms in Model 4.2 and all model diagnostics are good. Therefore, Model 4.2 is a good model.

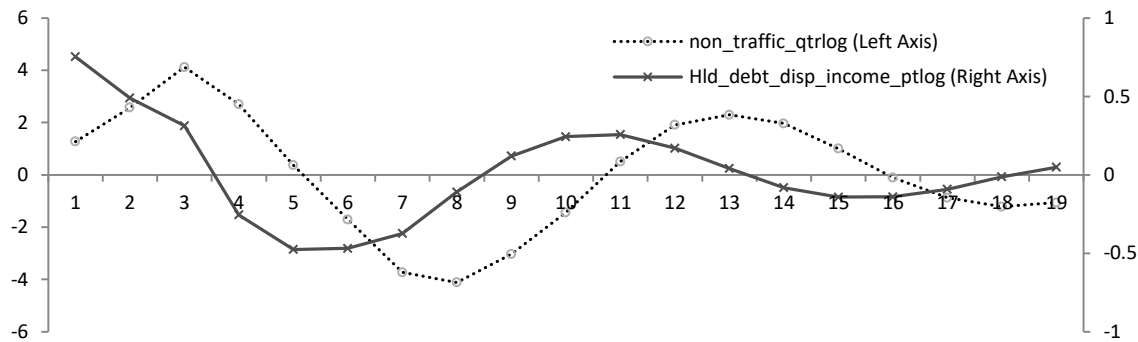
	<b>LnNTn</b>	<b>LnNTr</b>
<b>LnD</b>	LnD->LnNTn (4.58%) <sup>19</sup> LnNTn->LnD (12.43%)	LnD->LnNTr (4.15%) LnNTr->LnD (11.83%)
<b>EA</b>	EA -> LnNTn (6.65%)	EA -> LnNTr (7.15%)
<b>Prin1</b>	- <sup>20</sup>	-
<b>Prin2</b>	-	-

**Table 4.11 Granger-Causality test for non-traffic and the state of the economy model**

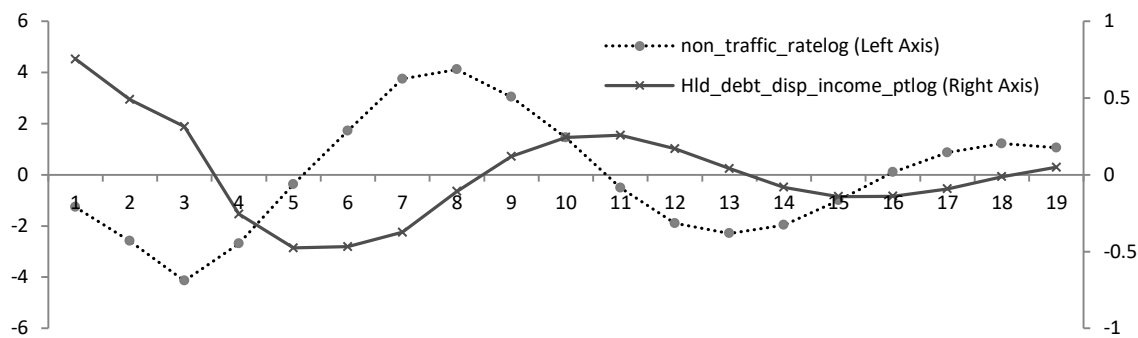
<sup>19</sup> Percentage value in brackets is the P-value for Granger-Causality test

<sup>20</sup> ‘-’ indicates that there is no significant causality relations found, all P-value from Granger-Causality test are greater than 20%

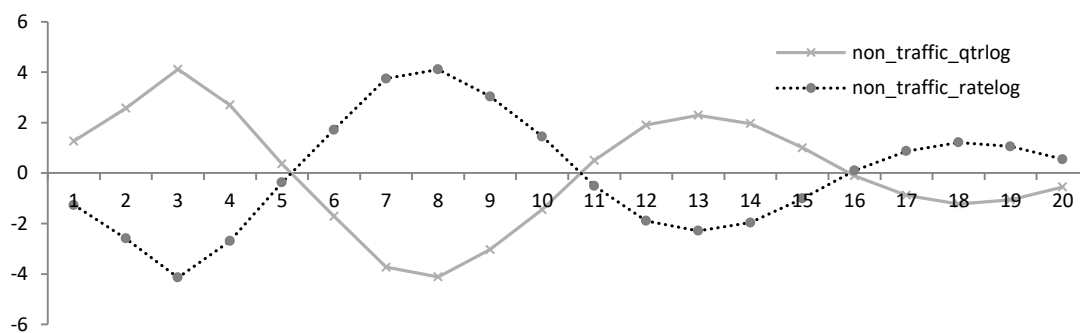
Granger-Causality test results in Table 4.11 show that the change to the growth rate of household debt and disposable income ratio causes the change to the growth rate of number of people charged with non-traffic offences (and also growth rate of total number of people charged with non-traffic offences per 1,000 population) and vice versa. ‘Easter’ is significant to both crime variables. These results are the same as the Granger-Causality test results from the full model (Table 4.8).



**Figure 4.17: Impulse response of 'LnNTn' to one unit shock to 'LnD' and 'LnNTn'**



**Figure 4.18: Impulse response of 'LnNTr' to one unit shock to 'LnD' and 'LnNTr'**



**Figure 4.19: Impulse response of 'LnD' to one unit shock to 'LnNTr' and 'LnNTn'**

Impulse response of 'LnNTn' and 'LnNTr' behaved almost the same to shocks to 'LnD'. The peak impulse response of 'LnNTn' and 'LnNTr' is at the first quarter after the shock and

almost dies away from the eighth quarter. Shocks to 'LnD' cause 'LnNTn' or 'LnNTr' to increase in short-term (three-quarters after the shock) and decrease in medium term (between fourth and seventh quarter after the shock). Impulse response of 'LnNTn' and 'LnNTr' to shocks to 'LnD' reduced to a low level from the eighth quarter after the shocks. Impulse response of 'LnD' to shocks to 'LnNTr' and 'LnNTn' is getting smaller and smaller after the shocks and it moves between positive and negative in every five quarters. In addition, impulse response of 'LnNTn' and 'LnNTr' are bigger to shocks to itself.

### 4.2.2 Burglary and the state of the economy model

For full model of burglary and the state of the economy model, there are two VAR(4) models produced by using different seasonality methods. Both VAR(4) models have the problem of cross-correlations between univariate error terms. However, each univariate model in VAR(4) is good.

	LnBr		LnBn	
	Seasonal dummy	Seasonal differencing	Seasonal dummy	Seasonal differencing
<b>GDP</b>	LnBn -> GDP (9.88%) <sup>21</sup>	LnBn -> GDP (16.45%)	LnBn -> GDP (9.77%)	LnBn -> GDP (16.16%)
<b>LnW</b>	LnW -> LnBn (16.58%)	LnBn -> LnW (7.99%)	LnW -> LnBn (20%)	LnBn -> LnW (7.58%)
<b>LnD</b>	- <sup>22</sup>	-	-	-
<b>U</b>	-	LnU -> LnBn (16.73%)	-	LnU -> LnBn (17.26%)
<b>I</b>	-	-	-	-

**Table 4.12: Granger-Causality test for burglary and the state of the economy model (full model)**

Based on Granger-Causality test results in Table 4.12, movements of 'LnBr' and 'LnBn' cause movement of 'GDP'. Likewise, movement of 'LnW' causes the movement of 'LnBr' and 'LnBn' and vice versa. Therefore, 'LnW', 'GDP', 'LnBr' and 'LnBn' are still to be modelled as endogenous variables. 'U' causes movement of 'LnBn' and 'LnBr' and not vice versa. 'LnD' and 'I' have no significant causality relationship with 'LnBn' and 'LnBr'. As a result, 'LnD', 'I' and 'U' should be exogenous variables. Forecasts of 'U', 'GDP', 'I' came

<sup>21</sup> Percentage value in brackets is the P-value for Granger-Causality test

<sup>22</sup> '-' indicates that there is no significant causality relations found, all P-value from Granger-Causality test are greater than 20%

from New Zealand Treasury. However, there is no predicted value of ‘LnD’ available from other sources, hence ‘LnD’ is modelled as an endogenous variable. This is because VECM or VAR model will produce forecasts for all endogenous variables.

### **Reduced model**

Cointegration test results are different due to different methods that are used to model seasonality. Cointegration relations are found among endogenous variables when seasonal dummies are used<sup>23</sup>. Maximum of three cointegration relations are found when seasonal differencing is used<sup>24</sup>. Autocorrelation in univariate error term is the problem for low lag model such as VECM(1) and VAR(2). This is the common problem for most of the low lag models in this study. Most of the higher lag models remove autocorrelation in error term. However, the univariate error terms are still not white noise for some endogenous variables. VAR(4) is the the only model with good model diagnostics for burglary and the state of the economy model, it is given by

$$\begin{pmatrix} \text{LnBr}_{1,t} \\ \text{LnBn}_{2,t} \\ \text{LnW}_{3,t} \\ \text{GDP}_{4,t} \\ \text{LnD}_{5,t} \end{pmatrix} = A_0 + A_1 \begin{pmatrix} \text{LnBr}_{1,t-1} \\ \text{LnBn}_{2,t-1} \\ \text{LnW}_{3,t-1} \\ \text{GDP}_{4,t-1} \\ \text{LnD}_{5,t-1} \end{pmatrix} + \dots + A_4 \begin{pmatrix} \text{LnBr}_{1,t-4} \\ \text{LnBn}_{2,t-4} \\ \text{LnW}_{3,t-4} \\ \text{GDP}_{4,t-4} \\ \text{LnD}_{5,t-4} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} Q1 \\ Q2 \\ Q3 \end{pmatrix} + (D)(\text{Prin1}_t) + (E_t) \quad (4.3)$$

where  $A_0$  and  $D$  are  $5 \times 1$  vectors,  $A_1$  to  $A_4$  are  $5 \times 5$  matrices. ‘EA’ (Easter dummy variable) is removed because it is insignificant to all endogenous variables. Coefficient estimates of seasonal dummy variables in ‘LnD’ and ‘GDP’ univariate models are set to zero because seasonal dummies are not significant to ‘LnD’ and ‘GDP’. ‘prin1’ is the first principal component made from unemployment rate (‘U’) and inflation (‘I’). It explained 69.14% of the variation. ‘U’ and ‘I’ are both highly correlated with ‘prin1’. Therefore, ‘prin1’ is enough to represent unemployment rate and inflation in Model 4.3.

<sup>23</sup> Cointegration test is significant only when  $P=5$  is used, the rank equals to 1 with P-value of 18.73%.

<sup>24</sup> The maximum number of cointegration relation was found when  $P=4$  is used, the rank equals to 3 with P-value of 16.64%.

	LnBr	LnBn
<b>GDP</b>	LnBr -> GDP (9.88%) <sup>25</sup>	LnBn -> GDP (9.77%)
<b>LnW</b>	LnW -> LnBr (17.29%)	LnW -> LnBn (16.58%)
<b>LnD</b>	<sub>26</sub>	-
<b>U</b>	-	-
<b>I</b>	-	-
<b>Prin1</b>	-	-

Table 4.13: Granger-Causality test for Model 4.3

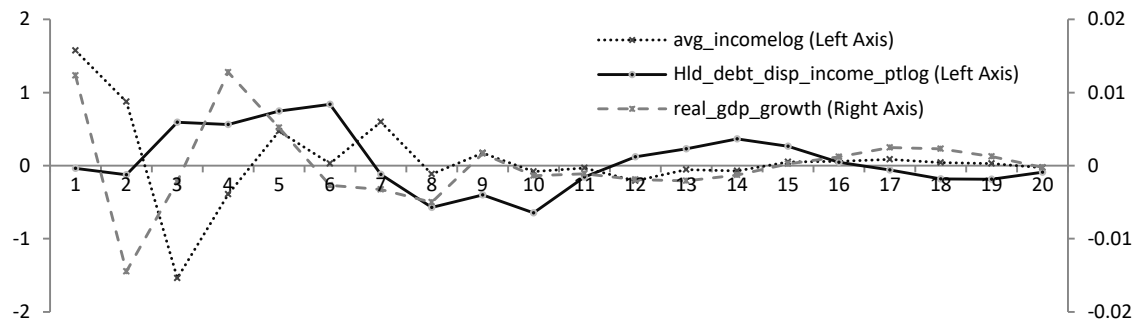


Figure 4.20: Impulse response of 'LnBn' to one unit shock to macroeconomic variables

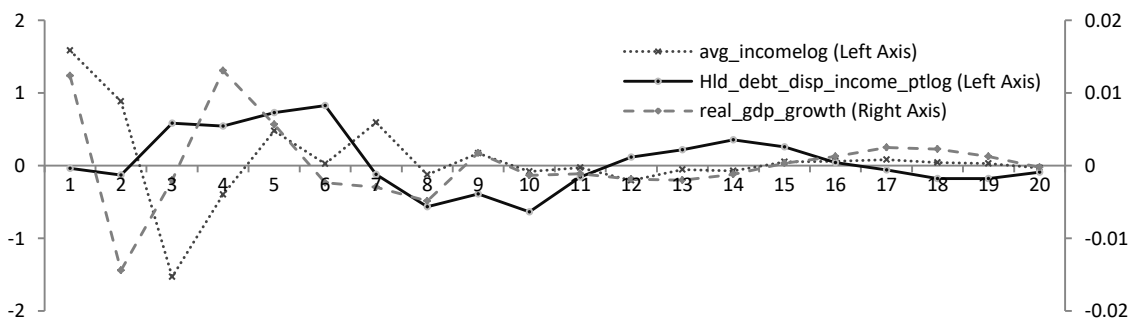


Figure 4.21: Impulse response of 'LnBr' to one unit shock to macroeconomic variables

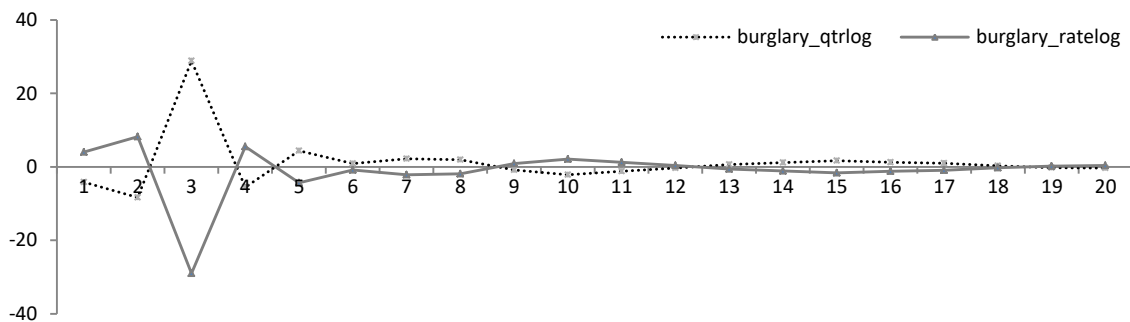


Figure 4.22: Impulse response of 'LnBn' and 'LnBr' to one unit shock to itself

<sup>25</sup> Percentage value in brackets is the P-value for Granger-Causality test<sup>26</sup> '-' indicates that there is no significant causality relations found, all P-value from Granger-Causality test are greater than 20%

Based on parameter estimates, 'prin1' is not significant to 'LnBr' and 'LnBn'. Granger-Causality test have same conclusions, movements of 'prin1' do not cause movements of 'LnBr' and 'LnBn'. For the relationship between 'GDP' and burglary charges, Granger-Causality test concludes that movements of 'LnBr' and 'LnBn' cause movements of 'GDP' at 10% significance level, not vice versa. This causality relation is shown by the impulse response. Impulse response is big to both 'LnBr' and 'LnBn' in the first four quarters after the shocks to 'GDP', then it started to drop to almost zero from the fifth quarter onwards.

For personal weekly average income, Granger-Causality test concluded that changes to 'LnW' cause changes to 'LnBr' and 'LnBn' at 20% significance level and not vice versa. Impulse response of 'LnBr' and 'LnBn' to shocks to 'LnW' dropped to almost zero at two years after the shocks. 'LnW' increased in the first two-quarters, and decreased in the third and fourth quarter after the shocks. In fact, the impulse response of 'LnBr' and 'LnBn' to the shocks to 'LnW' already started to get smaller one year after the shocks. As a result, impact to 'LnBr' and 'LnBn' are small from movements of 'LnW'. This is because Granger-Causality test is significant only at 20% significance level and the size of impulse response is small as well.

Impulse responses of 'LnBr' and 'LnBn' to shocks to growth rate of household debt and disposable income ratio (LnD) have different shape. They moved between positive and negative at different periods. Impulse responses of 'LnBr' and 'LnBn' are positive from the third quarter to the sixth quarter after the shocks, then the peak is in the sixth quarter after the shock. Impulse responses of 'LnBr' and 'LnBn' to shocks to 'LnD' are very small since the sixth quarter after the shock. In fact, impact on 'LnBr' and 'LnBn' to shocks to 'LnD' stayed in system for a very short period.

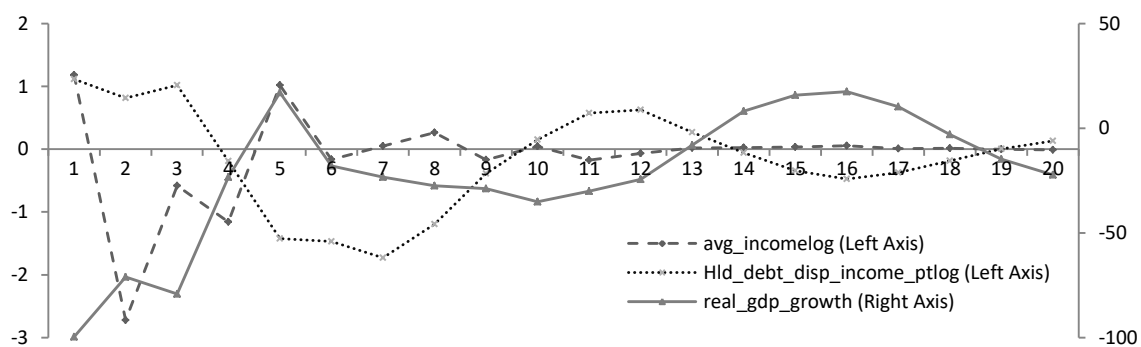
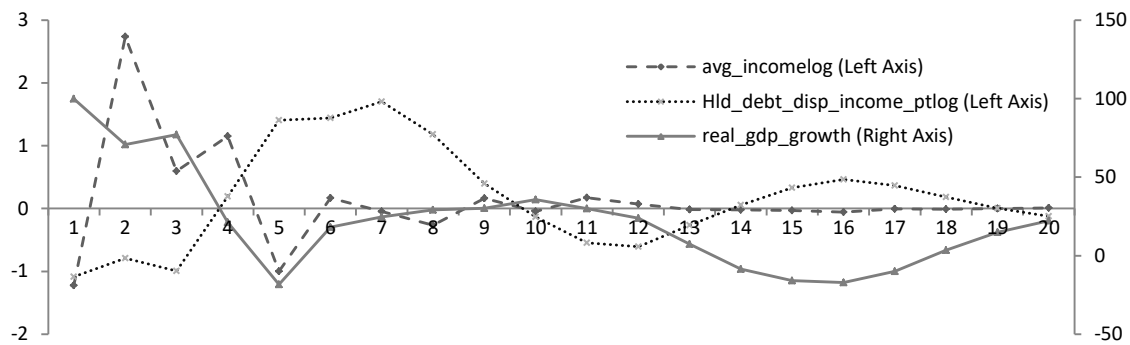


Figure 4.23: Impulse response of macroeconomic variables to one unit shock to 'LnBn'



**Figure 4.24: Impulse response of macroeconomic variables to one unit shock to 'LnBr'**

According to Figure 4.23 and Figure 4.24, patterns of impulse responses of macroeconomic variables to shocks to 'LnBr' and 'LnBn' are the same, but in an opposite direction. This is because the average value of 'LnBr' is negative (-0.93), and the average value of 'LnBn' is positive (7.36). Impulse response of 'GDP' reduces to a low level after four quarters since the shocks to 'LnBr' and 'LnBn'. At the six quarters after the shocks to 'LnBr' and 'LnBn', impulse response of 'LnW' reduces to almost zero. 'LnD' needs nine quarters to reduce impulse response to a low level. To compare with 'LnD' and 'LnW', impulse response of 'GDP' is much bigger to the shocks to 'LnBr' and 'LnBn'. This result matches with Granger-Causality test, 'GDP' is the only variable that has a causal relation with 'LnBr' and 'LnBn'.

### 4.2.3 Fraud and the state of the economy model

VAR(5) using seasonal differencing is the best model for full model of fraud and the state of the economy model. Each univariate model is good in VAR(5). Portmanteau test for cross-correlations of residuals suggested that there are cross-correlations between univariate errors and this is the only issue in model diagnostics.

	LnFr	LnFn
<b>GDP</b>	-	-
<b>LnW</b>	LnW -> LnFr (11.54%)	LnW -> LnFn (10.66%)
<b>LnD</b>	-	-
<b>U</b>	-	-
<b>I</b>	I -> LnFr (0.22%), LnFr -> I (0.03%)	I -> LnFn (0.23%), LnFn -> I (0.03%)

**Table 4.14<sup>27</sup>: Granger-Causality test for fraud and the state of the economy model (full model)**

<sup>27</sup> '-' indicates that there is no significant causality relations found, all P-value from Granger-Causality test are greater than 20% and percentage value in brackets is the P-value for Granger-Causality test



Granger-Causality test result in Table 4.14 shows that ‘I’ has bi-directional causality relationship with ‘LnFr’ and ‘LnFn’. It means that lags of ‘I’ are significant to lags of ‘LnFr’ and ‘LnFn’ and vice versa. Movements of ‘LnW’ cause the movements of ‘LnFr’ and ‘LnFn’ at 12% significance level, and not vice versa. According to parameter estimates, lags of ‘LnFr’ and ‘LnFn’ are significant to ‘LnW’, lags of ‘LnD’ are significant to ‘LnFr’ and ‘LnFn’. Therefore, ‘LnD’, ‘LnW’ and ‘I’ should be modelled as endogenous variables. ‘U’ and ‘GDP’ can be exogenous variables.

### **Reduced model**

For reduced models, low lag models such as VECM(1) have issues with autocorrelation in univariate error terms. A higher lag model such as VECM(4) still cannot remove autocorrelation from univariate error terms completely. Cross-correlation between univariate error terms is still an issue for the reduced model.

Cointegration test results are the same when different methods are used to model seasonality. There is one cointegration relationship among endogenous variables when lag length is one. An extra cointegration relationship is found when  $p=4$  is used. The extra lag length may bring error term from nonstationary to stationary. Therefore, cointegration relationship may become significant in higher order model. No VECM models are good for the reduced fraud and the state of the economy model. For this reason, coefficient estimates for long-term relationships can be inaccurate. Model 4.4 is the best fraud and the state of the economy model in reduced form. It is a VAR(4) using seasonal dummy variables to model seasonality. ‘Easter’ and ‘GDP’ are deleted because they are not significant in all univariate models. Therefore, Model 4.4 can be written as

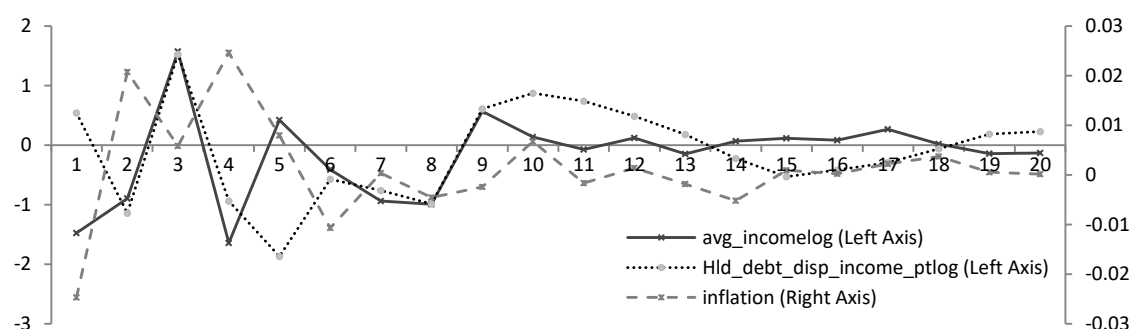
$$\begin{pmatrix} LnFr_{1,t} \\ LnFn_{2,t} \\ LnW_{3,t} \\ I_{4,t} \\ LnD_{5,t} \end{pmatrix} = A_0 + A_1 \begin{pmatrix} LnFr_{1,t-1} \\ LnFn_{2,t-1} \\ LnW_{3,t-1} \\ I_{4,t-1} \\ LnD_{5,t-1} \end{pmatrix} + \dots + A_4 \begin{pmatrix} LnFr_{1,t-4} \\ LnFn_{2,t-4} \\ LnW_{3,t-4} \\ I_{4,t-4} \\ LnD_{5,t-4} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} Q1 \\ Q2 \\ Q3 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ g_3 \\ g_4 \\ 0 \end{pmatrix} (U_t) + (E_t) \quad (4.4)$$

where  $A_0$  is a  $5 \times 1$  vector,  $A_1$  to  $A_4$  are  $5 \times 5$  matrices. Based on parameter estimates, seasonal dummies are not significant to ‘I’ and ‘LnD’, ‘U’ is not significant to ‘LnFn’, ‘LnFr’ and ‘LnD’. Therefore, coefficient estimates of seasonal dummies in ‘I’ and ‘LnD’ univariate models are set to zero. Coefficient estimate of ‘U’ are set to zero in ‘LnFn’, ‘LnFr’ and ‘LnD’

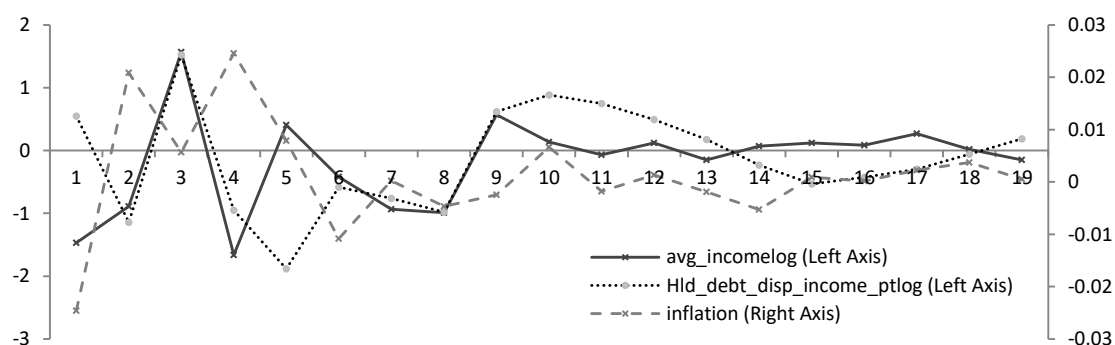
univariate models for the same reason. For Model 4.4, 'LnFn', 'LnFr', 'LnW' and 'I' all have the first differencing applied. Both first and seasonal differencing are needed for 'LnD' to achieve stationarity.

	LnFr	LnFn
LnW	<sup>28</sup>	-
LnD	-	-
I	I -> LnFr (3.34%) <sup>29</sup>	I -> LnFn (3.5%)

**Table 4.15: Granger-Causality test for Model 4.4**



**Figure 4.25: Impulse response of 'LnFn' to one unit shocks in macroeconomic variables**



**Figure 4.26: Impulse response of 'LnFr' to one unit shock to macroeconomic variables**

Granger-Causality tests from Table 4.15 suggest that movement of inflation causes the movement of 'LnFn' and 'LnFr' and not vice versa. Inflation is the only variable that has significant Granger-Causality relationships with 'LnFn' and 'LnFr'. Impulse responses of 'LnFn' and 'LnFr' are bigger in the first four quarters after the shocks to 'I' and 'LnW'. Impulse response reaches its highest level at four quarters after the shocks, then it becomes

<sup>28</sup> '-' indicates that there is no significant causality relations found, all P-value from Granger-Causality test are greater than 20%

<sup>29</sup> Percentage value in brackets is the P-value for Granger-Causality test

smaller and smaller until dropping to a level close to zero in the long-term. For inflation, impulse responses of 'LnFn' and 'LnFr' are small enough to be treated as zero at seven quarters after the shocks. To the shocks to 'LnW' and 'LnD', they take longer for impulse responses of 'LnFn' and 'LnFr' to die away. Impulse responses of 'LnFn' and 'LnFr' to shocks to 'LnD' are small in the first quarter, then getting bigger from the second to the fifth quarter after the shock. Impulse responses are too small to notice from the 13<sup>th</sup> quarter after the shocks to 'LnD'. Shocks to 'LnW' need to wait 11 quarters until impulse response of 'LnFn' and 'LnFr' reach the level close to zero.

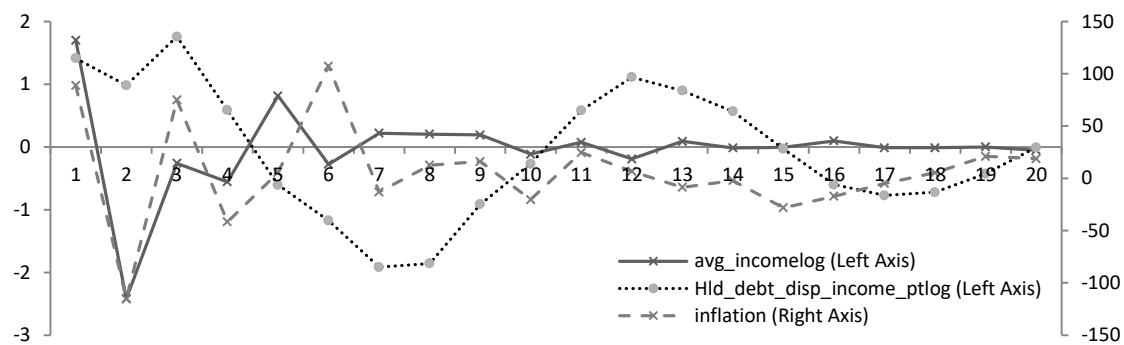


Figure 4.27: Impulse response of macroeconomic variables to one unit shock to 'LnFn'

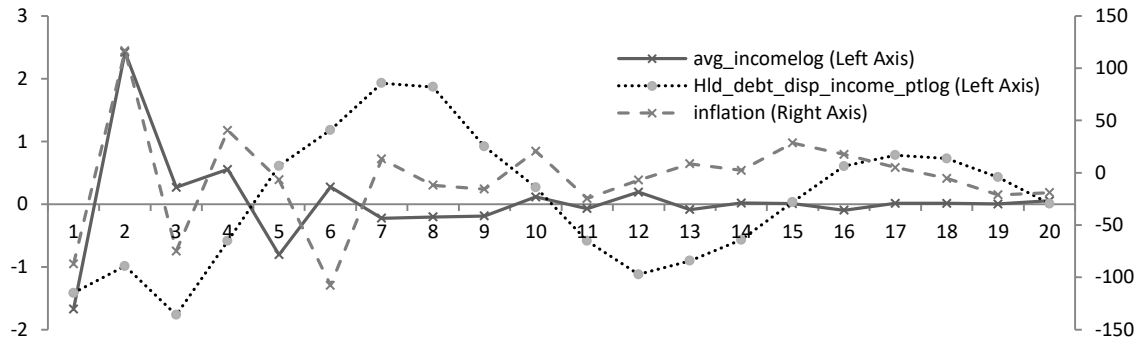


Figure 4.28: Impulse response of macroeconomic variables to one unit shock to 'LnFr'

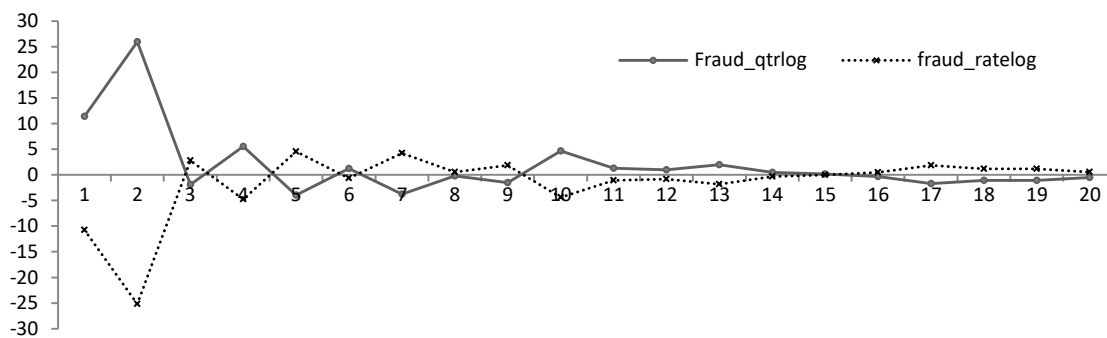


Figure 4.29: Impulse response of 'LnFn' to one unit shock to 'LnFn' and 'LnFr'

The same as for the burglary and the state of the economy model, impulse response of macroeconomic variables to shocks to 'LnFr' and 'LnFn' have the same movement, but in an opposite direction. Impulse response of 'LnW' and 'I' to shocks to 'LnFr' and 'LnFn' peak at two quarters after the shock and then stay at a high level until six quarters after the shocks. Impulse responses become very small from the seventh quarter after the shocks. Impulse responses of 'LnD' to shocks to 'LnFr' and 'LnFn' peak at seventh quarters after the shocks. Then, it dropped to a low level from the ninth quarter after the shock. Impulse responses of 'LnFr' and 'LnFn' to shocks to itself are big in the first two-quarters after the shock. Impulse responses reduce to a smaller scale until tenth quarter, and then reduce further from the 11<sup>th</sup> quarter after the shock. Impulse responses of 'LnFr' and 'LnFn' to one unit shock to itself are much bigger than shocks from macroeconomic variables. Therefore, it is reasonable to conclude that the movements of 'LnFr' and 'LnFn' are mainly caused by its own historical movement. Movements of macroeconomic variables have a much smaller contribution to movement of 'LnFr' and 'LnFn'.

#### 4.2.4 Theft and the state of the economy model

VAR(4) using seasonal differencing is the only good full model for theft and the state of the economy model. All of the univariate models' error terms are white noise, hence there is no autoregressive structure or heteroscedasticity in their error terms. Cross-correlations between univariate error terms is the only issue in VAR(4).

	LnTr	LnTn
<b>GDP</b>	-	-
<b>LnW</b>	LnW-> LnTr (12.2%), LnTr -> LnW (5.96%)	LnW-> LnTn (11.26%), LnTn -> LnW (5.9%)
<b>LnD</b>	LnD-> LnTr (12.66%), LnTr -> LnD (8.35%)	LnD-> LnTn (13.04%), LnTn -> LnD (8.13%)
<b>U</b>	-	-
<b>I</b>	LnTr -> I (17.33%)	LnTn -> I (17.08%)

**Table 4.16: Granger-Causality test for theft and the state of the economy model (full model)**

Granger-Causality test (Table 4.16) suggests that 'LnD' and 'LnW' have bi-directional causality relationship with both 'LnTr' and 'LnTn' at 15% significance level. Other variables are not showing significant causality relationships. As a result, 'LnTr', 'LnTn', 'LnD' and 'LnW' are endogenous variables in the reduced model. 'I', 'U' and 'GDP' are exogenous variables.

### Reduced model

Cointegration test results are different when different lag lengths are used. In fact, all cointegration relationships have rank equal to one. The same as models from earlier sections, low lag reduced models have heteroscedasticity or autoregressive structure in their univariate error terms. Model 4.5 is the only good model for theft and the state of the economy model in reduced form. It can be written as

$$\begin{pmatrix} LnTr_{1,t} \\ LnTn_{2,t} \\ LnD_{3,t} \\ LnW_{4,t} \end{pmatrix} = A_0 + A_1 \begin{pmatrix} LnTr_{1,t-1} \\ LnTn_{2,t-1} \\ LnD_{3,t-1} \\ LnW_{4,t-1} \end{pmatrix} + \dots + A_4 \begin{pmatrix} LnTr_{1,t-4} \\ LnTn_{2,t-4} \\ LnD_{3,t-4} \\ LnW_{4,t-4} \end{pmatrix} + (C)(EA_t) + (D) \begin{pmatrix} prin1_t \\ prin2_t \end{pmatrix} + (E_t) \quad (4.5)$$

where  $A_0$  and  $C$  are  $4 \times 1$  vectors,  $A_1$  to  $A_4$  are  $4 \times 4$  matrices,  $D$  is a  $4 \times 2$  matrix.

	<b>LnTr</b>	<b>LnTn</b>
<b>LnW</b>	LnW-> LnTr (12.2%), LnTr -> LnW (5.96%)	LnW-> LnTn (11.26%), LnTn -> LnW (5.9%)
<b>LnD</b>	LnD-> LnTr (12.66%), LnTr -> LnD (8.35%)	LnD-> LnTn (13.04%), LnTn -> LnD (8.13%)
<b>Prin1</b>	-	-
<b>Prin2</b>	-	-

**Table 4.17: Granger-Causality test for model 4.5**

Granger-Causality test for the reduced model has similar results as the full model. Both ‘LnW’ and ‘LnD’ have bi-directional causality relations with both ‘LnTr’ and ‘LnTn’. Movements of ‘LnTr’ and ‘LnTn’ cause the movements of ‘LnW’ and ‘LnD’ at 10% significance level. Movements of ‘LnW’ and ‘LnD’ cause the movements of ‘LnTr’ and ‘LnTn’ at 15% significance level. The movement of ‘prin1’ or ‘prin2’ does not move ‘LnTr’ or ‘LnTn’. It means that changes of the unemployment rate and inflation or real GDP growth rate do not have impact on the growth rate of number of people charged with theft or crime rate for theft offence. Based on parameter estimates, effects of ‘LnW’ and ‘LnD’ to ‘LnTn’ and ‘LnTr’ move between positive and negative at different lags. Impulse response is a good method to check how ‘LnTr’ or ‘LnTn’ respond to the shocks to ‘LnW’ or ‘LnD’.

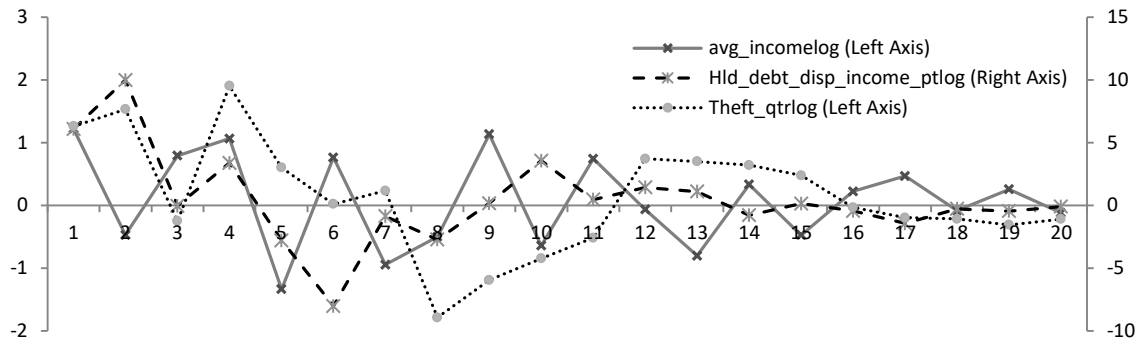


Figure 4.30: Impulse response of 'LnTn' to one unit shock to 'LnTn', 'LnW' and 'LnD'

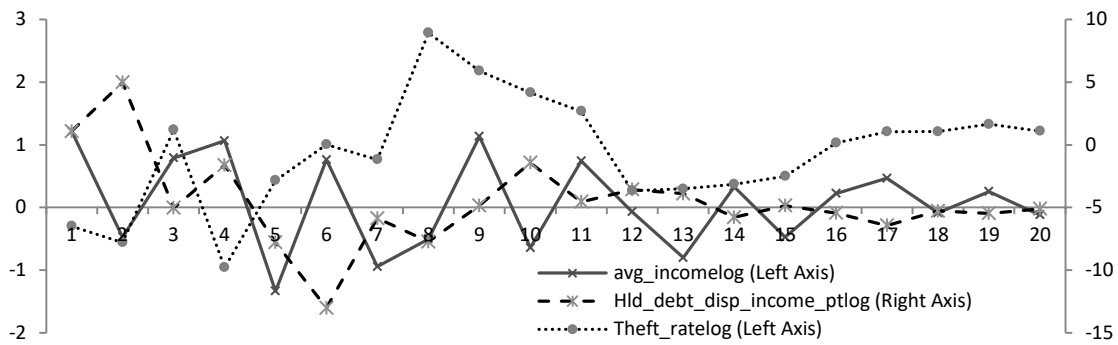


Figure 4.31: Impulse response of 'LnTr' to one unit shock to 'LnTr', 'LnW' and 'LnD'

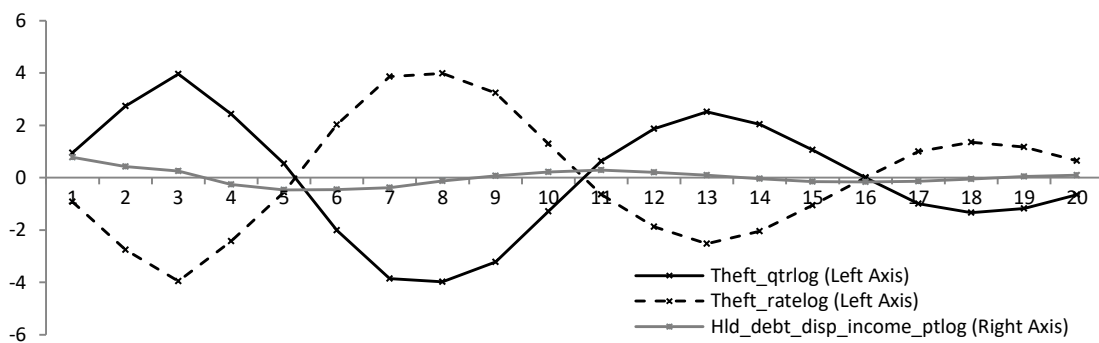


Figure 4.32: Impulse response of 'LnD' to one unit shock to 'LnTn', 'LnTr' and 'LnD'

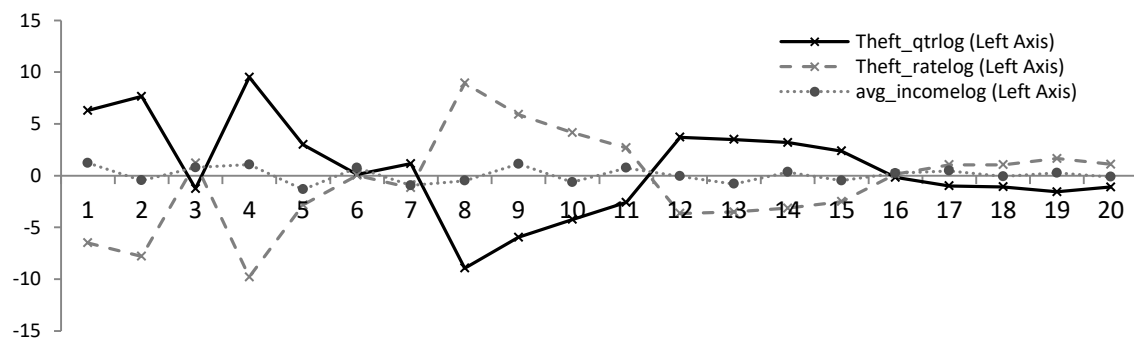


Figure 4.33: Impulse response of 'LnW' to one unit shock to 'LnTn', 'LnTr' and 'LnW'

Granger-Causality test shows that both 'LnD' and 'LnW' have bi-directional causality to 'LnTn' and 'LnTr'. However, the size of the impulse response of 'LnTn' and 'LnTr' to shocks to 'LnD' and 'LnW' are small compared with the average value of 'LnTn' and 'LnTr'. The highest level of impulse response of 'LnTn' and 'LnTr' comes from shocks to itself.

Impulse response of 'LnTn' and 'LnTr' to shocks to 'LnD' are positive in the short-term, they are negative in the medium-term. In this case, short-term is defined as the first year after the shocks, medium-term is the second year after the shocks. After shocks to 'LnD', the impulse response of 'LnTn' and 'LnTr' peaks at the second quarter, then drops to very low levels from the ninth quarter. Impulse responses of 'LnTn' and 'LnTr' to shocks to 'LnW' are both reasonably flat, and the peak is at nine quarters after the shocks. Actually, impulse responses of 'LnTn' and 'LnTr' have similar patterns to shocks to 'LnW' at the first eleven quarters after the shocks. Compared with average values of 'LnTn' and 'LnTr', size of impulse response become very small from twelve quarters after the shocks.

The main effect of the movement in 'LnTn' and 'LnTr' comes from itself. This is because impulse responses of 'LnTn' and 'LnTr' to the shocks to itself are much bigger than shocks to 'LnD' or 'LnW'. Impulse response of 'LnD' moves between positive and negative in five quarters pattern and gets smaller and smaller over time. Shocks to 'LnTr' and 'LnTn' resulted in the impulse response of 'LnW' to move between positive and negative. Based on Figure 4.32 and Figure 4.33, impulse responses of 'LnW' and 'LnD' to shocks to itself are much smaller than shocks to 'LnTn' and 'LnTr'.

### **4.3 Model 3: Association between Theft, Burglary and Fraud**

Some offenders committed different types of crime through their life. It is common for an offender to start with the less serious types of crime such as shoplifting, then move to more serious crime type such as burglary or robbery. Therefore, crime variables used in this study may have some degrees of association. The purpose of this section is to test the relationship between crimes in different offence categories. Seasonality is modelled in this case using seasonal differencing and seasonal dummies. There are two models in this section. First the crime number model, which models the number of people charged each quarter with burglary

(LnBn), fraud (LnFn) and theft (LnFr). The crime rate model uses the same methodology as the crime number model. 'LnBr', 'LnTr' and 'LnFr' are endogenous variables for crime rate model. Both seasonal dummies and seasonal differencing are used for seasonality.

Unit root test results (Table 4.1) suggests that variable such as 'LnBn' and 'LnFn' are not cointegrated after seasonal differencing because they are already stationary. Therefore, the maximum number of cointegration relationships is equal to two when seasonal differencing is used. Based on cointegration test results in Table 4.18, there are some degrees of cointegration between crime variables. Most of the models found cointegration relationships when seasonal dummy is used to model seasonality. There is a lower number of cointegration relationships found when seasonal differencing is used.

	Seasonal Dummy Model	Seasonal Differencing Model
<b>Crime number model</b>	P=1 (Rank=2, 13.6%) <sup>30</sup> P=4 (Rank=1, 20.26%) P=5 (Rank=1, 15.86%)	P=5 (Rank=1, 32.06%)
<b>Crime rate model</b>	P=1 (Rank=2, 14.49%)	P=8 (Rank=1, 22.89%)

**Table 4.18: Cointegration test for crime rate and crime number model**

VECM(5) using seasonal dummies is the only good crime number model in this section. It is given by

$$\begin{pmatrix} \Delta \text{LnTn}_{1,t} \\ \Delta \text{LnBn}_{2,t} \\ \Delta \text{LnFn}_{3,t} \end{pmatrix} = A_0 + A_1 \begin{pmatrix} \text{LnTr}_{1,t-1} \\ \text{LnBn}_{2,t-1} \\ \text{LnFn}_{3,t-1} \end{pmatrix} + A_2 \begin{pmatrix} \Delta \text{LnTr}_{1,t-1} \\ \Delta \text{LnBn}_{2,t-1} \\ \Delta \text{LnFn}_{3,t-1} \end{pmatrix} + \dots + A_5 \begin{pmatrix} \Delta \text{LnTn}_{1,t-4} \\ \Delta \text{LnBn}_{2,t-4} \\ \Delta \text{LnFn}_{3,t-4} \end{pmatrix} + (C)(EA_t) + (B) \begin{pmatrix} Q1_t \\ Q2_t \\ Q3_t \end{pmatrix} + (E_t) \quad (4.6)$$

where  $A_0$  and  $C$  are  $3 \times 1$  matrices,  $A_1$  to  $A_5$ ,  $B$  are  $3 \times 3$  matrices. Most of the models have heteroscedasticity in their univariate error terms. Autocorrelation in error terms is another common issue for low lag models.

---

<sup>30</sup> Percentage value in brackets is the P-value for cointegration test. P is the lag length used to perform cointegration test



VAR(1,2,4) using seasonal dummies is the only good model for the crime rate, and it can be written as

$$\begin{pmatrix} LnTr_{1,t} \\ LnBr_{2,t} \\ LnFr_{3,t} \end{pmatrix} = A_0 + A_1 \begin{pmatrix} LnTr_{1,t-1} \\ LnBr_{2,t-1} \\ LnFr_{3,t-1} \end{pmatrix} + A_2 \begin{pmatrix} LnTr_{1,t-2} \\ LnBr_{2,t-2} \\ LnFr_{3,t-2} \end{pmatrix} + A_3 \begin{pmatrix} LnTr_{1,t-4} \\ LnBr_{2,t-4} \\ LnFr_{3,t-4} \end{pmatrix} + (C)(EA_t) + (B) \begin{pmatrix} Q1_t \\ Q2_t \\ Q3_t \end{pmatrix} + (E_t) \quad (4.7)$$

where  $A_0$  and  $C$  are  $3 \times 1$  vectors.  $A_1$  to  $A_3$ ,  $B$  are  $3 \times 3$  matrices.

	LnTn	LnBn	LnFn
LnTn	NA	LnTn->LnBn (0.02%) <sup>31</sup> LnBn->LnTn (0.07%)	LnTn->LnFn (0.92%) LnFn->LnTn (4.62%)
LnBn		NA	No causality
LnFn			NA

**Table 4.19: Granger-Causality test for Model 4.6**

	LnTr	LnBr	LnFr
LnTr	NA	LnBn->LnTn (2.17%)	LnTn->LnFn (6.98%)
LnBr		NA	No causality
LnFr			NA

**Table 4.20: Granger-Causality test for Model 4.7**

Based on Table 4.19 and Table 4.20, causality relationships between crime variables are much stronger in the crime number model. For the crime number model, growth rate of number of people charged with theft has bi-directional causality with growth rate of number of people charged with burglary or Fraud. In crime rate model, the causality relationships are still significant between growth rate of crime rate in theft, burglary and fraud. However, bi-directional causality is lost in crime rate model. Movement of 'LnTr' causes the movement of 'LnFr' and not vice versa. Movement of 'LnBr' causes the movement of 'LnTr' and not vice versa. In conclusion, there are causality relations between theft, fraud and burglary crimes. It means that lags of crime variable are needed to explain each crime variable's current value. Therefore, Model 4.6 and Model 4.7 can be used to forecast crime variables.

---

<sup>31</sup> Percentage value in brackets in Table 4.19 and Table 4.20 is the P-value for Granger-Causality test

## 4.4 Model 4: Crime forecasting using ARIMA model

Based on the results from VAR and VECM models in Sections 4.1 to 4.3, impacts of macroeconomic variables are too small to make meaningful changes to crime variables. Moreover, impacts on shocks to crime variables themselves are much bigger than impact on shocks to macroeconomic variables. Therefore, macroeconomic variables may not improve the accuracy of crime variable forecasts. Univariate model ARIMA can be a good model to forecast crime variables. As with VAR or VECM models in earlier sections, seasonal dummy variables and seasonal differencing are used to model seasonality. Easter dummy is used to model the impact of Easter on the number of people charged. For this instance, ARIMA model becomes ARIMAX model because seasonal dummy and Easter dummy are included as independent variables.

Based on Unit Root test in Table 4.1, ‘LnTn’ and ‘LnNTn’ are not stationary after seasonal differencing. Both first and seasonal differencing are needed to achieve stationarity. ‘LnFn’ and ‘LnBn’ are stationary after seasonal differencing. All crime variables are stationary after first differencing. Based on Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) graphs, seasonal patterns are still strong after first differencing. Therefore, seasonal dummy variables are used to model seasonality for first differenced models. Another way to remove seasonal patterns is to apply seasonal and first differencing.

ARIMA is a univariate model, which models the current value of the dependent variable from its own historical movement. Model 4.8 is ARIMAX (p, (1,4), q) which includes ‘Easter’ dummy as an independent variable because seasonal differencing is the method to model seasonality. It is given by

$$Y_t = a_0 + a_1 Y_{t-1} + \dots + a_p Y_{t-p} + c_0 EA_t + \varepsilon_t + e_1 \varepsilon_{t-1} + \dots + e_q \varepsilon_{t-q} \quad (4.8)$$

where  $Y_t = (1 - B^1)(1 - B^4)Y'_t = (Y'_t - Y'_{t-1}) - (Y'_{t-4} - Y'_{t-5})$ . ARIMAX (p, 1, q) includes both ‘Easter’ and seasonal dummies as independent variable, it can be written as

$$Y_t = a_0 + a_1 Y_{t-1} + \dots + a_p Y_{t-p} + b_1 Q_{1t} + b_2 Q_{2t} + b_3 Q_{3t} + c_0 EA_t + \varepsilon_t + e_1 \varepsilon_{t-1} + \dots + e_q \varepsilon_{t-q} \quad (4.9)$$

where  $Y_t = (1 - B^1)Y'_t = Y'_t - Y'_{t-1}$ .

It is important to determine the value of  $p$  and  $q$  for ARIMA models. As with VAR or VECM, information criteria can be used to select values of  $p$  and  $q$ . SAS has Tentative Order Selection Test to pick values of  $p$  and  $q$  based on BIC. The model with the lowest value of BIC is the chosen model. Alternatively, ACF and PACF graph are also useful to decide the  $p$  and  $q$  values of ARIMA models. Data has AR structure when ACF decays linearly, exponentially or alternating exponentially. Significant lags in PACF reflect the lag of AR. Similarly, data has MA structure when PACF decays exponentially or alternating exponentially. Significant lags in ACF reflect the lag of MA. Based on parameter estimation and correlations of parameter estimates, AR or MA terms with the highest  $p$ -value and value of correlation to another parameter will be deleted until all AR or MA terms are significant and not highly correlated to other parameters.

### **Non-Traffic Model**

Tentative Order Selection Test suggests that ARIMAX(2, 1, 4) and ARIMAX(8, 1, 0) have lower BIC than other models. Based on ACF and PACF graphs, ARIMAX((4,8), 1, 0) is a good model. ACF declines linearly in seasonal lag with only lag four and eight positively significant. PACF is positively significant at lag four and eight. Both ARIMAX((4,8), 1, 0) and ARIMAX(8, 1, 0) reduce to ARIMAX((4), 1, 0), and ARIMAX(2, 1, 4) reduces to ARIMAX(0, 1, (4)). ‘Easter’ is an independent variable and seasonal dummies are insignificant.

Both ACF and PACF graphs only have single significant negative peak at lag four after the first and seasonal differencing. This recommends that ARIMAX(0, (1, 4), (4)) is a good model to fit the data. The best model from Tentative Order Selection are ARIMAX(4, (1, 4), 0) and ARIMAX(0, (1, 4), 4). ‘Easter’ is the independent variable in this case. Based on parameter estimates, ARIMAX(4, (1, 4), 0) reduces to ARIMA((3,4),(1,4),0) and ARIMAX(0, (1, 4), 4) reduces to ARIMA(0, (1, 4), (4)).

### **Burglary Model**

After first differencing, ACF for ‘LnBn’ has alternating exponential decline with a negative peak at lags two and six. ACF also has positive peaks at lags four and eight. There is a single significant negative peak at lag two for PACF. ACF graph suggests that there is seasonal AR structure after the first differencing. Therefore, ARIMAX(2, 1, 0) is a good model to try,

seasonal dummy variables are used to model seasonality. Tentative Order Selection choose ARIMAX(3, 1, 3) as the best model. After deleting insignificant lags, both ARIMAX(2, 1, 0) and ARIMAX(3, 1, 3) reduce to ARIMAX(1, 1, 0) with seasonal dummy as independent variables. In this case, 'Easter' is not significant to 'LnBn' and all model diagnostics for ARIMAX(1, 1, 0) are good.

'LnBn' is stationary after seasonal differencing. PACF and ACF graphs recommend that 'LnBn' has weak moving average structure and seasonality pattern after seasonal differencing. PACF graph shows the weak alternating exponential decline with positive peaks at lag five and lag nine. ACF graph shows the single significant positive peak at lag one. As a result, ARIMAX(0, 0, 1)(0, 0, 1)<sub>4</sub> can be a good model for 'LnBn' based on ACF and PACF graphs. ARIMA(5, 4, 1) is the best model recommended by Tentative Order Selection Test because it has the lowest BIC. ARIMAX(0, 0, 1)(0, 0, 1)<sub>4</sub> reduces to ARIMA(0, 0, 1)(0, 0, 1)<sub>4</sub> because 'Easter' is insignificant. However, model diagnostics conclude that ARIMA(0, 0, 1)(0, 0, 1)<sub>4</sub> is not a good model due to non-white-noise error terms in ARIMA(0, 0, 1)(0, 0, 1)<sub>4</sub>. ARIMAX(5, 4, 1) reduces to ARIMA(1, 0, 0)(1, 1, 0)<sub>4</sub> after deleting insignificant lags and 'Easter'. All model diagnostics for ARIMA(1, 0, 0)(1, 1, 0)<sub>4</sub> are good, hence it is the only good seasonal differencing model for 'LnBn' in this study.

### **Fraud Model**

'ΔLnFn' has alternating exponential decline ACF graph with a negative peak at lag two. PACF has a single significant negative peak at lag two. Thus, ARIMAX((2), 1, 0) may possibly be a good model in this case. Based on Tentative Order Selection, ARIMAX(2, 1, 2) is the best model. ARIMAX(2, 1, 2) reduces to ARIMAX(2, 1, 0) after excluding 'Easter' and insignificant lags. Seasonal dummies are significant, and still to be modelled as independent variables.

'LnFn' is stationary after seasonal differencing. However, there is a significant drop at 2004 after seasonal differencing. This is due to the big drop in number of people charged with fraud since 2004. Thus, 'LnFn' has both first differencing and seasonal differencing applied in order to remove the impact of the big drop from 2003 to 2004. After first and seasonal differencing, PACF graph declines exponentially in negative. Lags 4, 8 and 12 are significant. ACF graph has a single negative peak at lag four. This means that 'Δ<sup>1,4</sup>LnFn' has seasonal

moving average structure. Therefore,  $ARIMAX(0, 1, 1)(0, 1, 1)_4$  can be a good model for 'LnFn'.  $ARIMAX(0, 1, 1)(0, 1, 1)_4$  reduces to  $ARIMA(0, 1, 1)(0, 1, 1)_4$  because 'Easter' is not significant. All moving average lags are significant. According to Tentative Order Selection Test,  $ARIMAX(2, (1, 4), 4)$  is the best model.  $ARIMA(0, (1, 4), (4))$  is the final model once insignificant lags been removed from  $ARIMAX(2, (1, 4), 4)$ . However,  $ARIMA(0, (1, 4), (4))$  is not a good model because error term is not white noise in earlier lag. As a result,  $ARIMA(0, 1, 1)(0, 1, 1)_4$  is the only good model for 'LnFn'.

### **Theft Model**

After first differencing, 'LnTn' has a negative peak at lag two with the alternating exponential decline in ACF graph. PACF graph only has a significant peak in negative at lag two and a positive significant peak at lag four. Those graphs reflect that there is  $AR(2)$  structure in 'LnTn' after first differencing and  $ARIMAX((2), 1, 0)$  might be a good model.  $ARIMAX(2, 1, 2)$  has the lowest BIC from Tentative Order Selection.  $ARIMAX(2, 1, 2)$  reduces to  $ARIMAX((2), 1, 0)$  after deleting insignificant lags. Seasonal dummies and 'Easter' are significant to 'LnTn', hence they remain as independent variables in the model.

'LnTn' needs both seasonal and first differencing to be stationary. PACF graph shows that lag four and eight are negatively significant. ACF graph has lag four as the only significant lag and it is negative. This clearly indicates that there is moving average structure in ' $\Delta^{1,4}LnTn$ '.  $ARIMAX(0, (1,4), (4))$  is a good model for 'LnTn' with 'Easter' as an independent variable. Tentative Order Selection Test has a similar outcome, the model with the lowest BIC is  $ARIMAX(0, (1,4), 4)$ . The final result is that  $ARIMAX(0, (1,4), (4))$  reduces to  $ARIMA(0, (1,4), (4))$  because 'Easter' is not significant.  $ARIMAX(0, (1,4), 4)$  also reduces to  $ARIMA(0, (1,4), (4))$  after removing insignificant lags.

### **Conclusion**

From Table 4.21, ARIMA models using seasonal differencing reduce to either MA or AR model. 'Easter' is not significant in all models when seasonal differencing is used. Some models using seasonal differencing still have significant AR or MA terms at lag four. This gives the further adjustment to seasonality. According to ACF and PACF graphs, seasonal patterns are removed successfully when both first and seasonal differencing are applied to crime variables. Seasonal patterns are still in the data after first differencing. Consequently,

seasonality is modelled by seasonal dummy variables in first differencing ARIMA models. ARIMA reduces to AR in most of the first differencing models. Based on parameter estimates, 'LnNTn' is the only crime variable that has non-significant seasonal dummies in ARIMA model. The same conclusion can be made from VAR(4) (Model 4.1) with seasonal dummies. 'Easter' is significant to 'LnNTn' and 'LnTn' only.

	Model	AIC	BIC
<b>LnBn</b>	ARIMA(1, 0, 0) (1, 1, 0) <sub>4</sub>	-237.37	-232.606
	ARIMAX(1,1,0)**	-267.814	-255.719
<b>LnFn</b>	ARIMA(0, 1, 1) (0, 1, 1) <sub>4</sub> **	-194.711	-187.603
	ARIMAX(2,1,0)	-191.712	-184.456
<b>LnNTn</b>	ARIMA(0,(1,4),(4))	-276.754	-272.015
	ARIMAX(0,1,(4))	-309.821	-302.564
	ARIMA((3,4),(1,4),0)	-273.715	-266.607
	ARIMAX((4),1,0)**	-313.146	-305.889
<b>LnTn</b>	ARIMA(0,(1,4),(4))	-271.679	-266.94
	ARIMAX((2),1,0)**	-295.185	-280.672

**Table 4.21<sup>32</sup>: Summary table for ARIMA modelling**

As reported by Table 4.21, most of the first differenced ARIMAX models with seasonal dummies perform better in-sample than models using seasonal differencing. Fraud model is the only exception, seasonal differencing model is better than first differencing model. AIC and BIC can only compare between two or more models when the same modelling methodology is used on a dataset. It means that dataset needs to have same sample size and same variable format. For example, AIC and BIC cannot compare models using log-transformed variables and untransformed variables. ARIMA and VAR model cannot be compared using AIC and BIC either. MSE, MAE or MPE can compare the accuracy of the model when different modelling methods or a different form of the variable are used. For example, MSE, MAE or MPE can compare the performance of VAR, VECM and ARIMA model. Models using logarithm variable and original variable formats are comparable using MSE, MAE or MPE. Perfect goodness of fit model may not produce good forecast. In-sample and out-of-sample comparison of MSE, MAE or MPE is in the next section and can be used to pick the best forecast.

---

<sup>32</sup> Models with \*\* are the best model for each crime variable based on AIC and BIC.

# Chapter 5 Model Comparison and Forecasting

In-sample forecasts usually have a high accuracy rate. The reason is that in-sample forecasts are generated by the same set of data which was used to estimate the model parameters. One way to validate the accuracy of forecasts is to compare out-of-sample forecasts produced in a holdout sample, which is a sample created by holding some observations from the original dataset (Brooks, 2002, ch.5). In this study the holdout sample uses data from 2013 to 2014.

Mean Square Error (MSE), Mean Absolut Error (MAE) or Mean Percentage Error (MPE) are used to compare the accuracy of models. In some cases the best model picked by information criteria may not be the one with the lowest MSE, MAE and MPE. A model that captures the correct trend, seasonal pattern or turning point will be more valuable to make a decision in business or government strategy (Brooks, 2002, ch.5). MSE, MAE and MPE can be calculated as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_t - Y_t)^2$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{Y}_t - Y_t|,$$

$$MPE = \frac{100\%}{n} \sum_{i=1}^n \frac{\hat{Y}_t - Y_t}{Y_t}$$

where n is the number of observations. It is too difficult to pick the best forecast sometimes, each forecast has its own strength. The combination of forecasts is a method to combine multiple forecasts from different models, and it performs better than a single forecast in many cases. The easiest way to combine forecasts is an equal-weighted average.

It is simply just the average of multiple forecasts,

$$f_{ew} = \frac{1}{n} \sum_{i=1}^n f_i$$

where  $f_i$  is the forecast produced by different models. ' $f_{ew}$ ' is the new forecast produced by using the equal-weighted average. ' $n$ ' is number of forecasts (Elliott & Timmermann, 2016).

## 5.1 In-Sample and Out-of-sample comparison

Data between 1992 and 2012 are used to build models. This is the in-sample period for this study. Multi-step-ahead-forecasts are used to compare out-of-sample forecasts in this study. There are a total of three out-of-sample forecast periods to be compared. Models with the best out-of-sample forecast will be used to produce the final forecasts.

	Recursive window period	Out-of-sample forecasting period
<b>Forecast 1</b>	1992q1-2010q4	2011q1-2012q4
<b>Forecast 2</b>	1992q1-2011q4	2012q1-2013q4
<b>Forecast 3</b>	1992q1-2012q4	2013q1-2014q4

**Table 5.1: Recursive window and out-of-sample forecasting period for multi-step-ahead-forecast**

There are total of three steps for producing multi-step-ahead-forecasts,

### 1. Select models for multi-step-ahead-forecasts

This is based on in-sample model diagnostics, all models with good model diagnostics (listed in Table 5.2) are chosen to produce multi-step-ahead-forecasts.

### 2. Set recursive window period and out-of-sample forecasting period

In this case, the number of observations for recursive window increases by four quarters each time. Out-of-sample forecasting period is always equal to eight quarters. Table 5.1 listed the recursive window period and Out-of-sample forecasting period for each forecast.

### 3. Produce out-of-sample forecasts

All models listed in Table 5.2 will have parameters re-estimated using data from the recursive window periods. Those newly estimated parameters will be used to produce out-of-sample forecasts.



		MSE	MAE	MPE
Non traffic	ARIMA(0, (1,4), (4))	896146.4	767.83	0.23%
	ARIMAX(0, 1, (4))	699365.8	680.54	0.09%
	ARIMA((3,4), (1,4), 0)	903703.1	754.13	0.11%
	ARIMAX((4), 1, 0)	675126.8	664.13	0.13%
	VAR(4) (Model 4.2)	787441.5	695.96	0.07%
	VECM(4) <sup>33</sup>	912645.1	799.25	0.09%
	VECM(6)	731825.9	699.10	0.07%
Burglary	ARIMA(1, 0, 0)(1, 1, 0) <sub>4</sub>	7373.78	69.84	0.33%
	ARIMAX(1, 1, 0)	5233.41	55.67	0.12%
	VAR(4) (Model 4.3)	3601.57	43.75	0.07%
	VECM(5) (Model 4.6)	3260.88	43.85	0.07%
Fraud	VAR(4) (Model 4.3)	4429.38	54.14	0.13%
	ARIMA(0, 1, 1)(0, 1, 1) <sub>4</sub>	7178.62	67.94	0.21%
	ARIMAX(2, 1, 0)	8809.43	75.52	0.17%
	VECM(5) (Model 4.6)	4864.83	55.50	0.16%
Theft	ARIMA(0, (1,4), (4))	34341.49	148.51	0.05%
	ARIMAX((2), 1, 0)	27909.74	130.33	0.08%
	VAR(4) (Model 4.5)	28591.81	134.18	0.07%
	VECM(4) <sup>34</sup>	22547.77	114.36	0.06%
	VECM(4) <sup>35</sup>	41195.49	152.15	0.10%
	VECM(5) (Model 4.6)	18945.94	107.49	0.05%

Table 5.2: In-sample forecasts comparison, 1992-2012

Table 5.2 points out that VECM and VAR models perform better in-sample than ARIMA and ARIMAX models in general. Model 4.6<sup>36</sup> is the best in-sample model for Theft and Burglary, and the second best in-sample model for Fraud. Crime and the state of the economy model has the best in-sample accuracy for Non-traffic and Fraud.

The irregular component can be a big issue in time series forecasting, it is the random fluctuations in time series data with various causes (Newbold & Bos, 1994, ch.5). In this case, the irregular component is irregular seasonal patterns in crime variables especially between 2011 and 2014. No models can capture correct seasonal patterns in out-of-sample forecasts. Multivariate models have better in-sample accuracy than ARIMA models. Nonetheless, VECM and VAR are not stable when the models are fitted to recursive window periods for

<sup>33</sup> VECM(4) & VECM(6) in Table 5.2 are non-traffic and unemployment model with seasonal dummy variables

<sup>34</sup> VECM(4) is theft and unemployment model with seasonal dummy variables

<sup>35</sup> VECM(4) is theft and unemployment model with seasonal differencing

<sup>36</sup> Model 4.6 contains 'LnTn', 'LnBn' and 'LnFn' as endogenous variable. Seasonal dummies are used to model seasonality.

Forecast 1 and Forecast 2. There are only three VAR or VECM models that can fit the recursive window period well for both Forecast 1 and Forecast 2. They are VECM(4) using seasonal dummies for theft and unemployment model, VECM(4) and VECM(6) models for non-traffic and unemployment model using seasonal differencing. Although all models produced good univariate models in VECM or VAR models, cross-correlation between univariate error terms is an issue in this study. For ARIMA, all models can fit recursive window period well. A consistent model can be reused when new data becomes available. Otherwise, a new model is required every time when forecasts need to be updated.

### **Non-Traffic**

According to Table 5.3, ARIMA(0, (1,4), (4)), ARIMA((3,4), (1,4), 0) and VAR(4) have much lower MSE, MAE and MPE than other models. Forecast 1 from ARIMA(0, (1,4), (4)) and ARIMA((3,4), (1,4), 0) are not as good as VAR(4). All three models produced good Forecast 2 and Forecast 3. VAR(4) consistently produced good forecasts, all univariate models in VAR(4) are good. It can be used to forecast number of people charged with non-traffic offences (non-traffic) despite significant cross-correlations between univariate error terms. ARIMA(0, (1,4), (4)) and ARIMA((3,4), (1,4), 0) fits out-of-sample data well. This makes the ARIMA models more attractive to be used as long-term forecasting models.

		ARIMA (0,(1,4),(4))	ARIMA (0,1,(4))	ARIMA ((3,4),(1,4),0)	ARIMA ((4),1,0)	VAR(4) (Model 4.2)	VECM(4) <sup>37</sup>	VECM(6)
For 1	MSE	10055976	17937520	3151795	14199116	1465572	48161155	33046675
	MAE	2869.07	3678.41	1643.07	3302.48	1062.79	6264.52	5157
	MPE	13.13%	16.99%	7.39%	15.23%	4.62%	28.72%	23.68%
For 2	MSE	1922502	8217776	571411	4759768	626616	29107681	9619896
	MAE	1119.46	2498.97	653.48	1853.62	613.23	4852.73	2708.76
	MPE	6.12%	13.58%	-0.81%	10.10%	1.48%	26.20%	14.71%
For 3	MSE	1308032	5696302	850478	3540897	2021485	14419080	4511142
	MAE	878.89	2025.18	740.84	1612.01	1107.26	3394.21	1874.36
	MPE	4.90%	11.65%	0.61%	8.73%	6.66%	20.62%	11.35%

**Table 5.3<sup>38</sup>: Out-of-sample comparison for number of people charged with non-traffic offence**

<sup>37</sup> VECM (4), VECM (6) in Table 5.3 are non-traffic and unemployment model with seasonal dummy variables

<sup>38</sup> For 1, For 2 and For 3 are Forecast 1, Forecast 2 and Forecast 3

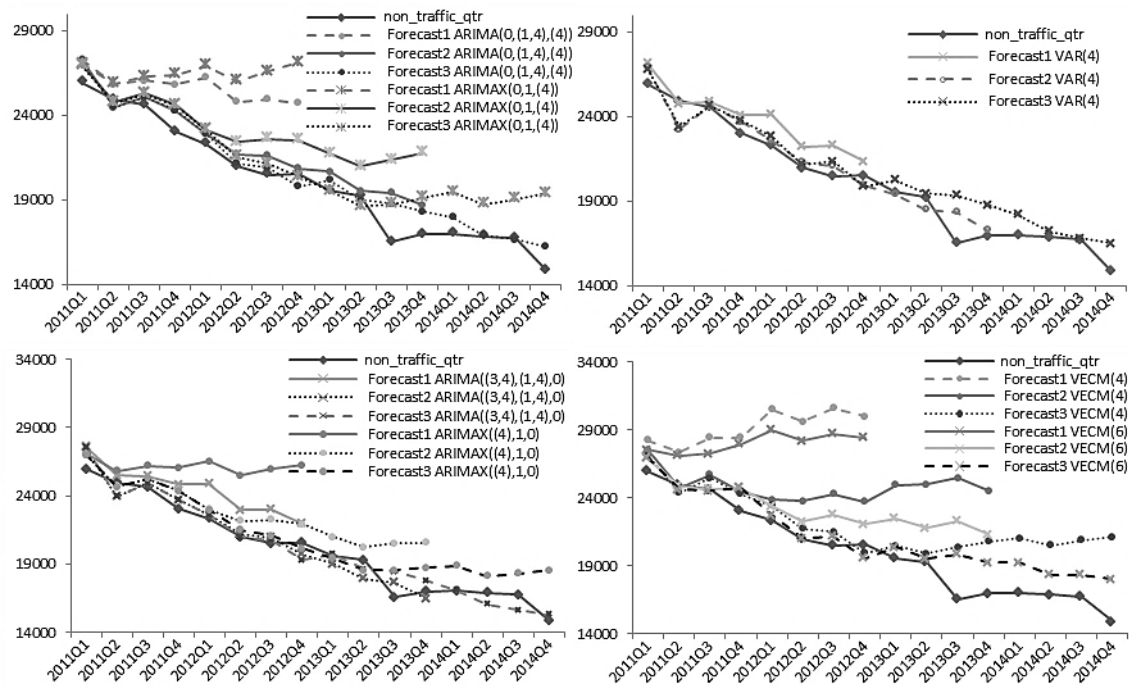


Figure 5.1: Out-of-sample forecasts for number of people charged with non-traffic offences

Figure 5.1 suggests that ARIMA(0, (1,4), (4)), ARIMA((3,4), (1,4), 0) and VAR(4) are clearly better than other models in terms of the accuracy of the forecast. Forecasts from other models have incorrect trend compared with actual data. VAR(4) and ARIMAX(0, 1, (4)) are two of the best in-sample models. All other well performed in-sample models failed to produce better forecasts than ARIMA(0, (1,4), (4)), ARIMA((3,4), (1,4), 0) and VAR(4).

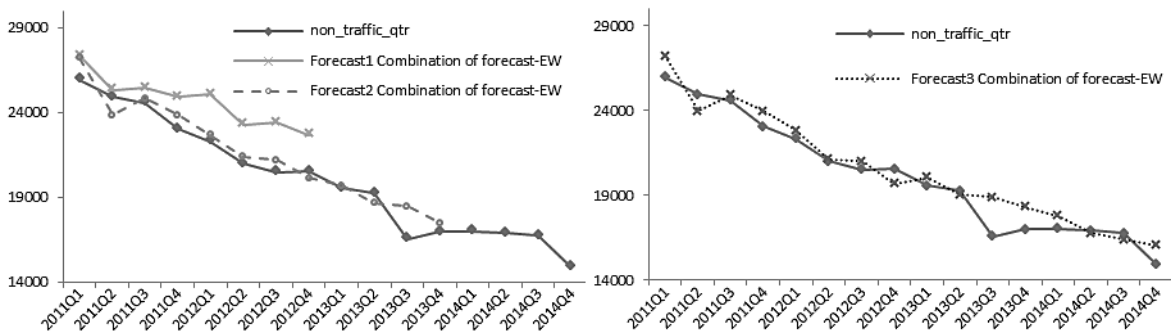


Figure 5.2: Out-of-sample forecasts using Equal-Weighted-Average for number of people charged with non-traffic offences

	Forecast1	Forecast2	Forecast3
<b>MSE</b>	4098015	681072.5	1193381
<b>MAE</b>	1843.68	639.16	850.09
<b>MPE</b>	8.38%	2.26%	4.06%

**Table 5.4: Out-of-sample forecasts for number of people charged with non-traffic offences: MAE, MSE and MPE for Equal-Weighted-Average**

A sharp decline in the third quarter of 2013 changed the trend of number of people charged with non-traffic, and none of the models were able to forecast this big drop. Out-of-sample forecasts from ARIMA(0, (1,4), (4)), ARIMA((3,4), (1,4), 0) and VAR(4) have correct trend before and after the third quarter of 2013. As a result, the final forecast will be the Equal-Weighted-Average for ARIMA(0, (1,4), (4)), ARIMA((3,4), (1,4), 0) and VAR(4) because it is too difficult to identify the best model based on the out-of-sample forecasts. Equal-Weighted-Average forecast in Figure 5.2 shows the correct trend. The performance of forecasting using Equal-Weighted-Average in out-of-sample period is not as good as ARIMA((3,4), (1,4), 0). However, Equal-Weighted-Average is still a good method to use, the final forecast will have impact from macroeconomic variables, AR and MA terms included.

### **Burglary**

Number of people charged with burglary has a relatively flat trend in general. The decreasing trend since 2010 is not as notable as the number of people charged with theft, non-traffic and fraud. Macroeconomic variables in VAR(4) have a very small impact on the forecast value of 'LnBn' because the impacts of macroeconomic variables to 'LnBn' are too small to make any notable changes. Thus, the main components that contribute to the forecast of 'LnBn' from VAR(4) are the AR terms of 'LnBn'. VECM(5) contains 'LnBn', 'LnTn' and 'LnFn' as endogenous variables and the conclusion is that 'LnBn' has bi-directional causality with both 'LnTn' and 'LnFn'. It is fair to conclude that the sharp decrease in trend since 2010 for 'LnTn' and 'LnFn' contributed to the decreasing trend of the forecast to 'LnBn' from VECM(5).

		ARIMA (1, 0, 0)(1, 1, 0) <sub>4</sub>	ARIMAX (1,1,0)	VAR(4) (Model 4.3)	VECM(5) (Model 4.6)
Forecast 1	MSE	24737.3	21373.39	40128.08	19262.71
	MAE	136.8581	117.43	179.68	111.11
	MPE	8.63%	7.48%	11.77%	7.36%
Forecast 2	MSE	80255.03	85647.25	66492.63	39557.52
	MAE	238.8054	247.42	220.56	162.55
	MPE	18.55%	19.21%	17.10%	12.68%
Forecast 3	MSE	110434.5	94247.88	105301.7	30461.92
	MAE	305.5641	272.87	296.81	149.83
	MPE	27.66%	24.87%	26.84%	12.44%

Table 5.5: Out-of-sample forecasts comparison for number of people charged with burglary

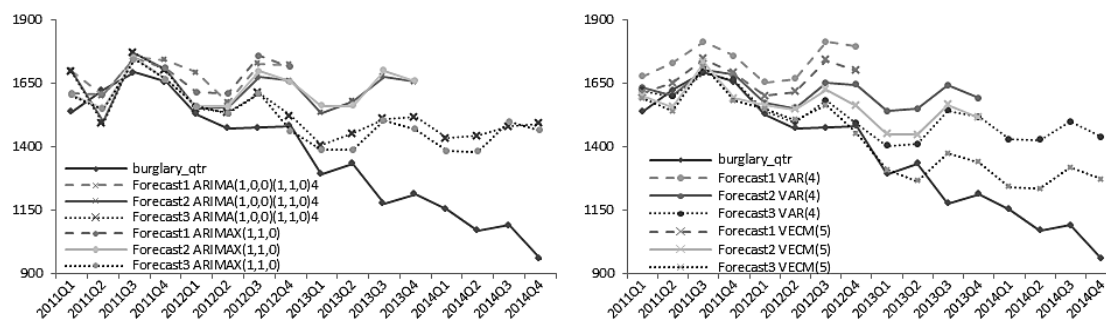


Figure 5.3: Out-of-sample forecasts for number of people charged with burglary

VECM(5) is the best model to produce out-of-sample forecasts for the number of people charged with burglary (Burglary). VECM(5) has lower MSE, MAE and MPE for Forecast 1 to Forecast 3. Actually, VECM(5) performed better in-sample as well. Figure 5.3 shows that burglary forecasts from VECM(5) have a minor decrease in trend from 2013 to 2014. VECM(5) capture the correct trend for 'Burglary' even though the actual 'Burglary' has steeper decreasing trend than the predicted value. The performance of VECM(5) in Forecast 1 to 3 is consistent while all other models' MAE, MSE and MPE increase sharply for Forecast 3. ARIMA(1, 0, 0)(1, 1, 0)<sub>4</sub> and ARIMAX(1, 1, 0) have flat trend in out-of-sample forecasts. This is different compared with actual 'Burglary'. Forecast 1 and Forecast 2 from two ARIMA models are better than Forecast 3. It is reasonable to believe that the accuracy of the forecasting from two ARIMA models will improve when new data becomes available. As a result, VECM(5), ARIMA(1, 0, 0)(1, 1, 0)<sub>4</sub> and ARIMAX(1, 1, 0) will be used to produce the final forecast.

## Fraud

Number of people charged with ‘Fraud’ is unstable from 2011 to 2014. There are some irregular seasonal movements and sharp increases or decreases. Producing accurate forecasts with the matching trend and seasonal pattern could be a challenge. Table 5.6 shows that ARIMA(0, 1, 1)(0, 1, 1)<sub>4</sub> and ARIMAX(2, 1, 0) are good and reliable models because out-of-sample forecasts are constantly good in Forecast 1 to Forecast 3. However, both ARIMA models are not as good as VECM(5) and VAR(4) for in-sample performance.

		<b>VAR(4) (Model 4.3)</b>	<b>ARIMA (0, 1, 1)(0, 1, 1)<sub>4</sub></b>	<b>ARIMAX (2,1,0)</b>	<b>VECM(5) (Model 4.6)</b>
<b>Forecast 1</b>	<b>MSE</b>	14389.81	5673.31	11766.94	43707.57
	<b>MAE</b>	104.03	64.6	81.65	188.6
	<b>MPE</b>	11.55%	2.50%	9.81%	22.01%
<b>Forecast 2</b>	<b>MSE</b>	22847.16	8683.79	17817.64	22702.8
	<b>MAE</b>	128.87	78.73	115.17	132.61
	<b>MPE</b>	16.84%	3.66%	11.94%	13.32%
<b>Forecast 3</b>	<b>MSE</b>	20573.01	18835.53	13778.51	16552.51
	<b>MAE</b>	119.58	110.97	95.55	99.44
	<b>MPE</b>	-6.18%	-6.43%	9.02%	-1.92%

Table 5.6: Out-of-sample forecasts comparison for number of people charged with fraud

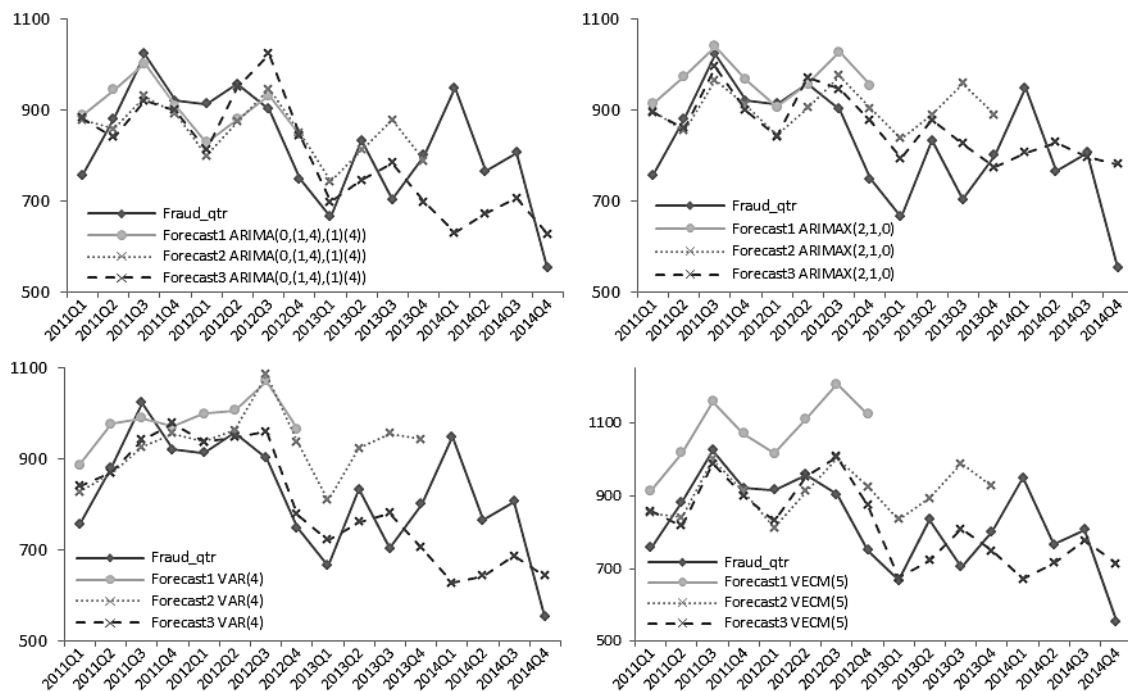
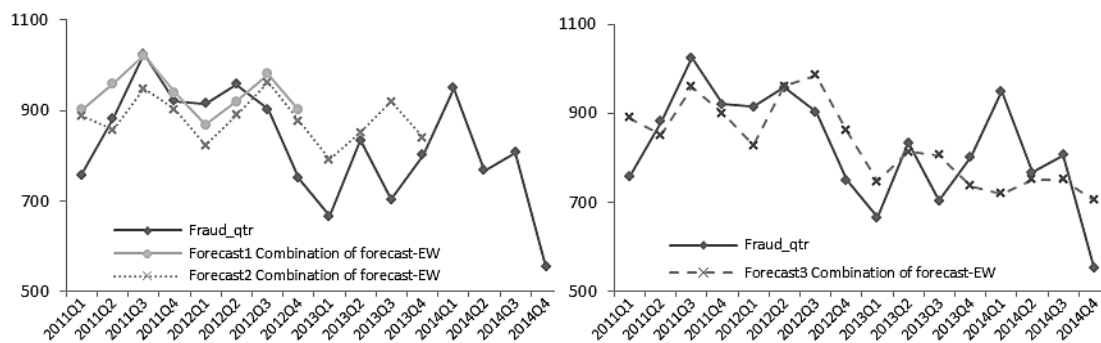


Figure 5.4: Out-of-sample forecasts for number of people charged with fraud

None of the models predicted seasonal patterns well because seasonal patterns in ‘LnFn’ are not consistent between 2011 and 2014. All models have captured the decreasing trend between 2011 and 2014. In fact, forecasts from VAR(4), VECM(5) and ARIMA(0, 1, 1)(0, 1, 1)<sub>4</sub> are identical. The seasonal patterns are the same in all forecasts. Forecasts from VAR(4), VECM(5) and ARIMA(0, 1, 1)(0, 1, 1)<sub>4</sub> peak at quarter three and the first and fourth quarters are lower. ARIMAX(2, 1, 0) has forecast peaks at quarter two. VAR(4), VECM(5) and ARIMA(0, 1, 1)(0, 1, 1)<sub>4</sub> have seasonal components modelled either through AR or MA terms. Those are the further adjustments to seasonality on top of the seasonal dummy or seasonal differencing.

Both ARIMA models can fit data well in the recursive window period. VECM(5) and VAR(4) cannot fit the data well in recursive window period due to cross-correlation between univariate error terms. For this reason, ARIMA(0, 1, 1)(0, 1, 1)<sub>4</sub> and ARIMAX(2, 1, 0) are picked to produce final forecasts. It is too difficult to choose the best model out of ARIMA(0, 1, 1)(0, 1, 1)<sub>4</sub> and ARIMAX(2, 1, 0), a combination of forecasts using Equal-weighted average is the final forecast for the number of people charged with Fraud.



**Figure 5.5: Out-of-sample forecasts for number of people charged with fraud: MAE, MSE and MPE for Equal-Weighted-Average**

	Forecast1	Forecast2	Forecast3
<b>MSE</b>	7490.836	12044.92	12511.7
<b>MAE</b>	69.79559	92.64823	89.72257
<b>MPE</b>	6.16%	7.80%	1.29%

**Table 5.7: Out-of-sample forecasts using Equal-Weighted-Average for number of people charged with fraud**

ARIMAX(2, 1, 0)’s forecast is higher than the actual value in general. ARIMA(0, 1, 1)(0, 1, 1)<sub>4</sub> has forecast slightly lower than the actual value especially at 2014. Table 5.7 recommends

that combination of forecasts is better than single model forecasts because it is consistent through Forecast 1 to Forecast 3. It performs worse than  $ARIMA(0, 1, 1)(0, 1, 1)_4$  in Forecast 1 and Forecast 2. In general, Figure 5.5 indicates that Equal-Weighted-Average forecast fits out-of-sample data better than single model forecast and Forecast 3 is the best example.

### **Theft**

‘LnTn’ has irregular seasonal pattern from 2013 to 2014. This makes it difficult for any models to pick the correct seasonal patterns. According to Figure 5.6,  $ARIMA(0, (1,4), (4))$ ,  $VAR(4)$  and  $VECM(5)$  predicted the correct trend and regular seasonal patterns for Forecast 1 to Forecast 3. Table 5.8 indicates that  $ARIMA(0, (1,4), (4))$ ,  $VAR(4)$  and  $VECM(5)$  have lower MSE, MAE and MPE than  $ARIMAX((2), 1, 0)$  and two  $VECM(4)$  models.

		<b>ARIMA (0,(1,4),(4))</b>	<b>ARIMAX ((2),1,0)</b>	<b>VAR(4) (Model 4.5)</b>	<b>VECM(4)<sup>39</sup></b>	<b>VECM(4)<sup>40</sup></b>	<b>VECM(5) (Model 4.6)</b>
<b>Forecast 1</b>	<b>MSE</b>	250711.3	422865.5	23956.93	291977.1	654234	498494.3
	<b>MAE</b>	449.92	582.26	133.15	475.47	736.47	651.71
	<b>MPE</b>	11.51%	14.90%	3.31%	12.21%	18.67%	16.57%
<b>Forecast 2</b>	<b>MSE</b>	267613.7	411791.7	35168.62	428670.7	523962.9	235992.7
	<b>MAE</b>	468.5	595.01	156	599.39	652.91	435.54
	<b>MPE</b>	13.48%	17.07%	-2.17%	17.24%	18.77%	12.53%
<b>Forecast 3</b>	<b>MSE</b>	43504.41	107535.9	50984.88	215527.1	83868.84	63562.49
	<b>MAE</b>	173.6	263.98	167.64	386.32	236.34	215.94
	<b>MPE</b>	1.68%	6.86%	1.31%	11.42%	4.57%	-3.40%

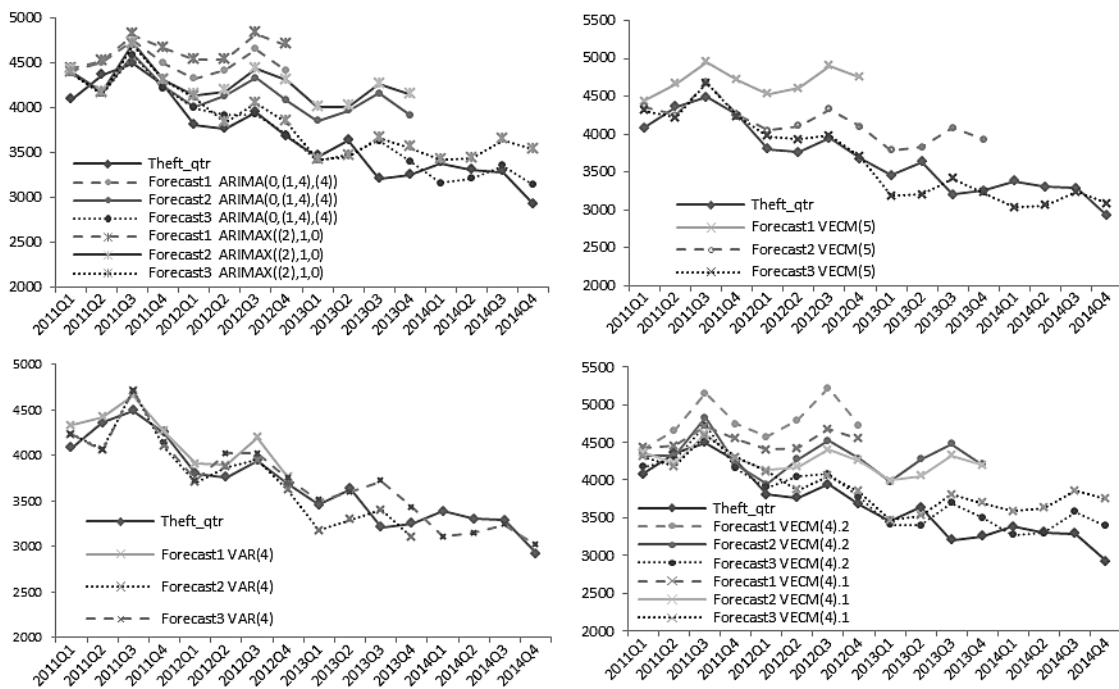
**Table 5.8: Out-of-sample comparison for number of people charged with theft**

Based on Figure 5.6,  $VECM(4)$  using seasonal dummies and  $ARIMAX((2), 1, 0)$  have flat trend for Forecast 1 to Forecast 3, this is different from the descending trend in actual data. For this reason,  $ARIMAX((2), 1, 0)$  and the two  $VECM(4)$  models are not good.

<sup>39</sup>  $VECM(4)$  is theft and unemployment model with seasonal dummy variables

<sup>40</sup>  $VECM(4)$  is theft and unemployment model with seasonal differencing



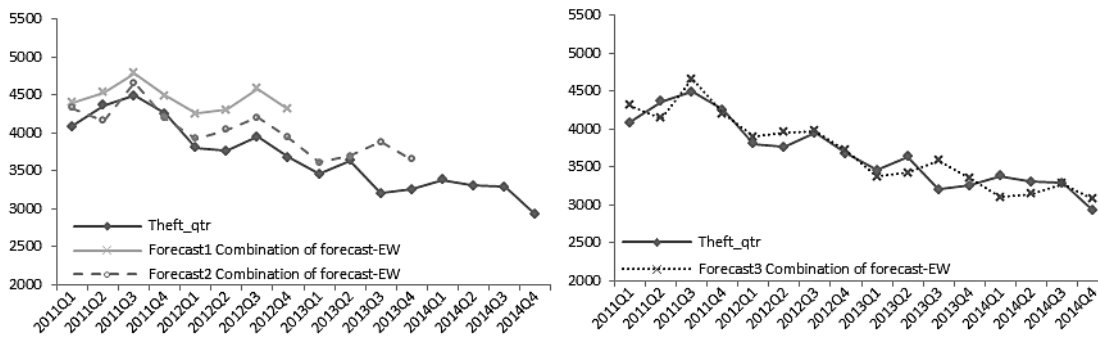


**Figure 5.641: Out-of-sample forecasts for number of people charged with theft**

VAR(4) is the best model because all three forecast fits the actual data well. In fact, VAR(4) has the lowest MSE, MAE and MPE for Forecast 1 to Forecast 3. It means that VAR(4) is a stable model and consistently produces good forecasts. VECM(5) performs the best in-sample. As with other VAR or VECM models in this section, VAR(4) and VECM(5) have issue with cross-correlation between univariate error terms when models fit to out-of-sample data. Nonetheless, it does not stop the forecasting of ‘LnTn’ from VAR(4) or VECM(5) to be used, because the univariate model of ‘LnTn’ is a good model.

Forecast 1 and Forecast 2 of ARIMA(0, (1,4), (4)) and VECM(5) are not good. However, Forecast 3 has improved when new data are included. ‘LnTn’ and ‘LnU’ are cointegrated in VECM(4) using seasonal differencing. This cointegration relationship becomes insignificant in recursive window period for Forecast 1 and Forecast 2. Therefore, VECM is not necessary in this case.

<sup>41</sup> In Figure 5.6, VECM(4).1 is theft and unemployment model with seasonal dummy variables and VECM(4).2 is theft and unemployment model with seasonal differencing



**Figure 5.7: Out-of-sample forecasts for number of people charged with theft: MAE, MSE and MPE for Equal-Weighted-Average**

	Forecast1	Forecast2	Forecast3
<b>MSE</b>	197932.1	108335.8	42390.92
<b>MAE</b>	411.5958	275.3556	173.9105
<b>MPE</b>	10.46%	7.95%	-0.14%

**Table 5.9: Out-of-sample forecasts using Equal-Weighted-Average for number of people charged with theft**

Figure 5.7 displays combination of forecasts using Equal-Weighted-Average. Forecast 1 to Forecast 3 having accurate trend compared with actual data. Seasonal patterns are forecast better than single model forecasts. The same conclusion can be made using data from Table 5.9. Forecast 3 using Equal-Weighted-Average has smaller MAE, MSE and MPE than all models listed in Table 5.8. The forecast will not be used for final forecast if the model cannot fit the data well and a new model is not required if at least one of VAR(4), ARIMA(0, (1,4), (4)) and VECM(5) fits the new dataset well.

## 5.2 Forecasting

Data between 1992 and 2014 are used to produce the final forecast of 2015 to 2019 for the quarterly number of people charged with burglary, fraud, theft and non-traffic offences. The extra two years of data from 2013 to 2014 are added to the dataset. It is common that a model becomes poorly fitted after adding new data, hence model diagnostics are monitored in order to ensure the chosen model fits the data well. In this case, models that have bad fits to the data will not be selected to produce final forecasts. Equal-Weighted-Average as a method of combination of the forecast will be used to produce final forecast.

<b>Non traffic</b>	ARIMA(0, (1,4), (4))
	ARIMA((3,4), (1,4), 0)
	VAR(4) (Model 4.2)
<b>Burglary</b>	ARIMA(1, 0, 0)(1, 1, 0) <sub>4</sub>
	ARIMAX(1, 1, 0)
	VECM(5) (Model 4.6)
<b>Fraud</b>	VAR(4) (Model 4.3)
	ARIMA(0,1,1)(0,1,1) <sub>4</sub>
	ARIMAX(2, 1, 0)
<b>Theft</b>	ARIMA(0, (1,4), (4))
	VAR(4) (Model 4.5)
	VECM(5) (Model 4.6)

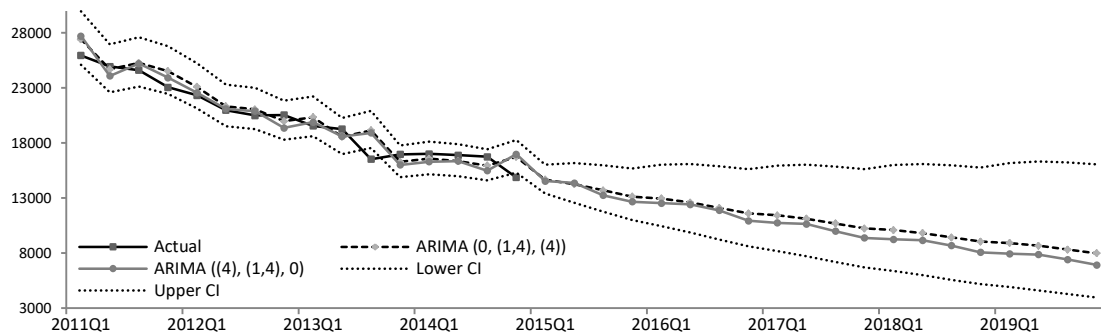
Table 5.10: Models used for final forecasts

Table 5.10 listed all models used to produce final forecasts. Most of the ARIMA models do not need modification or only minor modification because the model fits the data well. This is also the case in out-of-sample forecasts in Section 5.1. VAR and VECM failed to fit the data well when 2013 and 2014 data are included. Cross-correlations between univariate error terms is the issue for all VAR(4) models (Crime and the state of the economy model). ‘LnD’ created the problem of non-white-noise error term for VAR(4) models for theft, fraud and non-traffic offences. Most of the parameters are not significant for non-traffic and the state of the economy model (VAR(4)). Therefore, VAR(4) will not be used to produce final forecasts. Heteroscedasticity in univariate error terms of ‘LnTn’ and ‘LnFn’ is the issue for VECM(5) (Model 4.6). For this reason, forecasts of ‘LnTn’ and ‘LnFn’ from VECM(5) are not used. Univariate model for ‘LnBn’ in VECM(5) is still a good model, hence the forecast of ‘LnBn’ can be used.

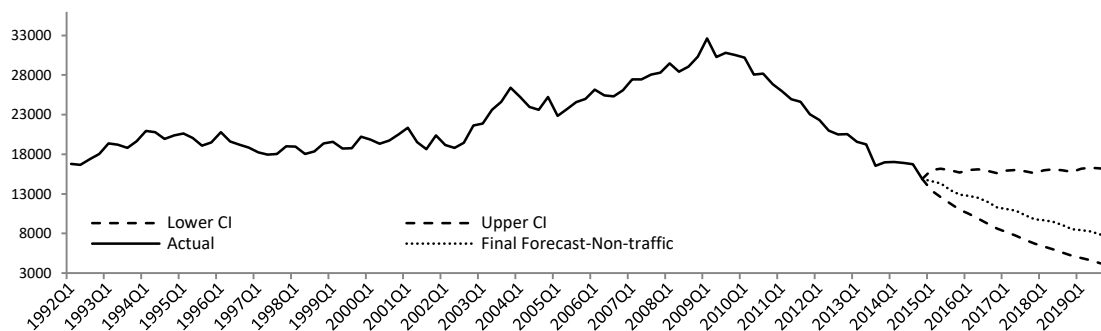
### **Non-Traffic**

Based on data from 1992 to 2014, ARIMA((3,4), (1,4), 0) reduces to ARIMA((4), (1,4), 0) because lag three of the AR term is not significant. Constant term is also removed from ARIMA((4), (1,4), 0) and ARIMA(0, (1,4), (4)) because it is insignificant. Both models can fit the data well, and ARIMA(0, (1,4), (4)) is better than ARIMA((4), (1,4), 0) because of lower AIC and BIC. From Figure 5.8, a number of people charged with non-traffic offences in the fourth quarter of 2013 and 2014 were slightly outside the confidence interval. Forecasts for the rest of the period are within the confidence interval. As a result, the confidence

interval from ARIMA(0, (1,4), (4)) performed well based on the performance between 2011 and 2014, and it is selected to be the confidence interval for the final forecasts.



**Figure 5.8: Number of people charged with non-traffic offences: 2015-2019 forecasts by ARIMA((4), (1,4), 0) and ARIMA(0, (1,4), (4))**

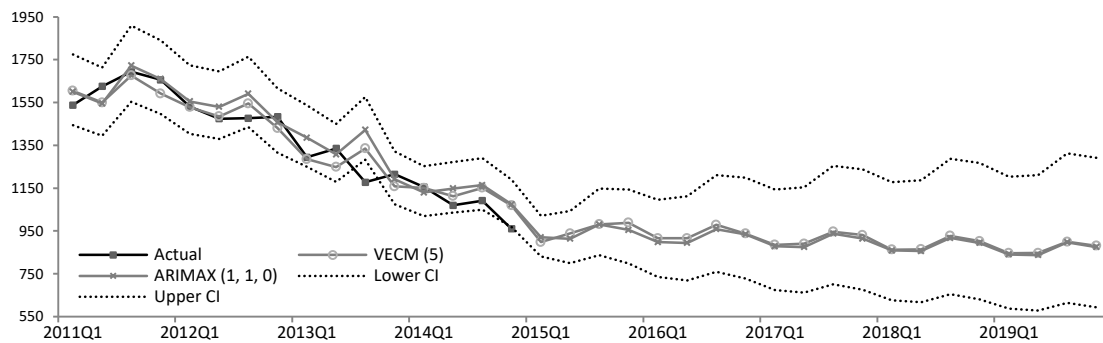


**Figure 5.9: Final forecast for number of people charged with non-traffic offences, 2015-2019**

Forecast produced by ARIMA((4), (1,4), 0) has better accuracy in seasonal patterns, but the level of forecast is lower than ARIMA(0, (1,4), (4)). Both models are forecasting a decreasing trend for number of people charged with Non-traffic offences between 2015 and 2019. Based on trend from 1992 to 2014, predicted number of people charged with the Non-traffic offences can drop to below 4000 in 2019. The highest level can be just above 16000, which is only marginally lower than the first quarter of 2014.

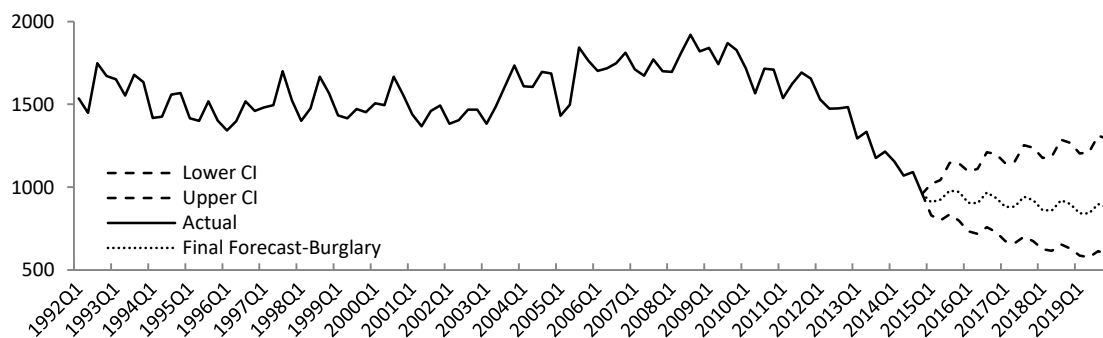
### **Burglary**

ARIMA(1, 0, 0)(1, 1, 0)<sub>4</sub> cannot be used to make final forecasts due to non-stationarity of 'Δ<sup>4</sup>LnBn'. 'LnBn' needs both seasonal and first differencing applied in order to become stationary when 2013 and 2014 data are included. This suggests that a new model is required. 'LnBn' forecasts from VECM(5) and ARIMAX(1, 1, 0) can still be used to produce a combination of forecasts using Equal-Weighted-Average.



**Figure 5.10: Number of people charged with burglary: 2015-2019 forecasts by VECM(5) and ARIMAX(1,1,0)**

ARIMAX(1, 1, 0) is purely based on the historical movement of 'LnBn', hence it is a good choice to be the confidence interval of the final forecast. Figure 5.10 shows that the actual number of people charged with burglary is out of the confidence interval at quarter three in 2013. The actual number is only eight people less than the lower range of the confidence interval at quarter four in 2014. Overall, the chosen confidence interval from ARIMAX(1, 1, 0) covers the data well and thus it is chosen to be the confidence interval for the final forecast.

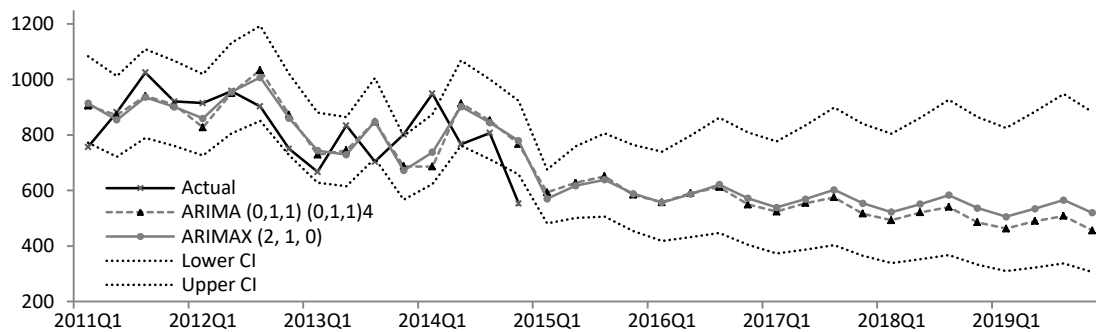


**Figure 5.11: Final forecast for number of people charged with burglary, 2015-2019**

Equal-Weighted-Average based on forecasting from VECM(5) and ARIMAX(1, 1, 0) is the final forecast for 'LnBn'. According to Figure 5.11, the final forecast suggests that the quarterly number of people charged with burglary will retain its decreasing trend until 2019, but at a much lower rate of decrease compared with 2010-2014. Number of people charged with burglary per quarter in 2019 could be as low as 592 people, and the upper limit is 1,291.

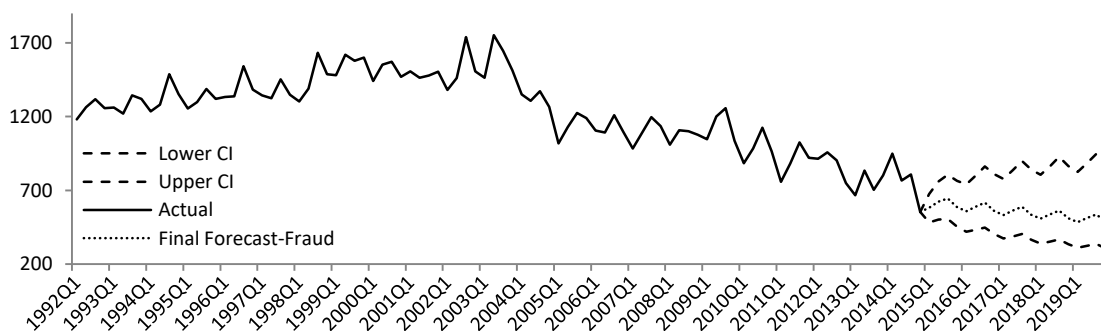
## Fraud

The constant term from  $ARIMA(0,1,1)(0,1,1)_4$  is removed because it is insignificant. Both  $ARIMA(0,1,1)(0,1,1)_4$  and  $ARIMAX(2, 1, 0)$  fit the data well with an extra two years of data.



**Figure 5.12: Number of people charged with fraud: 2015-2019 forecasts by  $ARIMA(0,1,1)(0,1,1)_4$  and  $ARIMAX(2, 1, 0)$**

Based on Figure 5.12, forecasts from  $ARIMA(0,1,1)(0,1,1)_4$  and  $ARIMAX(2, 1, 0)$  have the same trend and seasonal patterns.  $ARIMAX(2, 1, 0)$  has slightly higher forecast values than  $ARIMA(0,1,1)(0,1,1)_4$  from the third quarter of 2016.  $ARIMAX(2, 1, 0)$  has lower AIC and BIC, this suggests that  $ARIMAX(2, 1, 0)$  fits the data better than  $ARIMA(0,1,1)(0,1,1)_4$ . As a result, the confidence interval from  $ARIMAX(2, 1, 0)$  will be used as the confidence interval for the final forecast. As with the number of people charged with burglary, the actual number of people charged with fraud is out of the confidence interval range in some quarters between 2013 and 2014. The confidence interval from  $ARIMAX(2, 1, 0)$  managed to cover most of the unusual movements, hence it is a good confidence interval to be used in the final forecast.



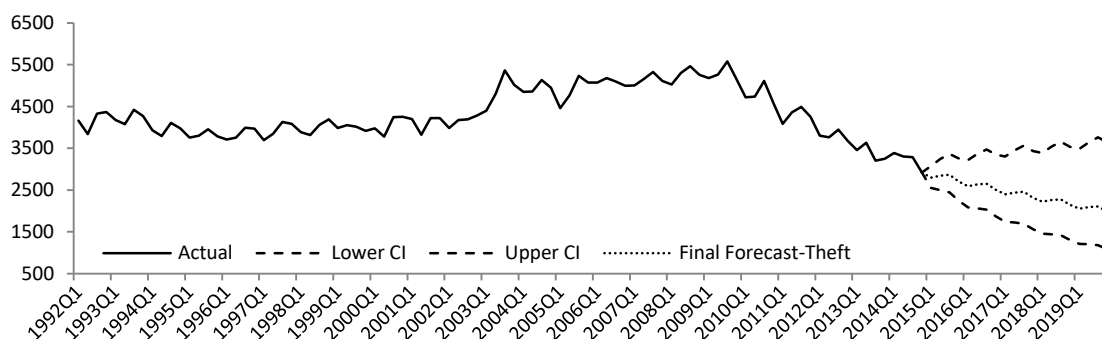
**Figure 5.13: Final forecast for number of people charged with fraud, 2015-2019**

Figure 5.13 is the final forecast produced by Equal-Weighted-Average from  $ARIMA(0,1,1)(0,1,1)_4$  and  $ARIMAX(2, 1, 0)$ . The graph shows that the quarterly number of people charged

with fraud will decrease until 2019, the rate of decrease will be lower than in 2009-2014. The final forecast suggests that the number of people charged with fraud is unlikely to reach beyond 1,000 before 2019 and the lowest level could be just above 300.

### **Theft**

ARIMA(0, (1,4), (4)) is the only model that fits the data well for number of people charged with theft. Therefore, combination of forecasts and choosing the confidence interval are not necessary in this case. Forecasts from ARIMA(0, (1,4), (4)) is the final forecast for the number of people charged with theft.



**Figure 5.14: Final forecast for number of people charged with theft, 2015-2019**

Final forecast from 2015 to 2019 in Figure 5.14 suggests that the quarterly number of people charged with theft will keep falling at a similar rate compared with 2012 – 2014. The highest level is likely to reach 3,641 in 2019, this is the same level as 2012. The lowest level in 2019 could be 1,084 and this is more than 50% lower than the forecast from the first quarter of 2015.

# Chapter 6 Discussion and Conclusions

Redoing Small and Lewis's (1996) research is the first part of this study. Unemployment and crime is modelled at both national and regional levels. At a national level, the unemployment and crime model suggested that the growth rate of the unemployment rate ( $\text{LnU}$ ) has a long-term relationship to the growth rate of the number of people charged with theft ( $\text{LnTn}$ ). The growth rate of the number of people charged with burglary ( $\text{LnBn}$ ) is not cointegrated with ' $\text{LnU}$ '. At a regional level, cointegration relationships are found between ' $\text{LnBn}$ ' and ' $\text{LnU}$ ' in all four regions. ' $\text{LnU}$ ' cointegrated with ' $\text{LnTn}$ ' and ' $\text{LnFn}$ ' in Auckland and Northland regions. The national level analysis suggested that there is no cointegration between ' $\text{LnU}$ ' and ' $\text{LnFn}$ '. In this case, associations between fraud and unemployment, or burglary and unemployment have different results in the national and regional level analysis.

Causality between crime and unemployment is much weaker at a regional level. The Granger-Causality test found no causality between ' $\text{LnFn}$ ' and ' $\text{LnU}$ ' at a national level. However, movements of ' $\text{LnU}$ ' cause the movements of ' $\text{LnFn}$ ' in Northland. Figure 4.8 suggests that the unemployment rate in Northland is much higher than the unemployment rate at a national level, and it is also the highest in the selected regions for this study. Therefore, it is reasonable to assume that the higher unemployment rate in Northland could be the reason for ' $\text{LnU}$ ' causing ' $\text{LnFn}$ '. This is confirmed by parameter estimates, the VECM(4) model suggested that the long-term relationship between ' $\text{LnU}$ ' and ' $\text{LnFn}$ ' is positive in Northland. The unemployment rate in other regions or at a national level may not be high enough to move ' $\text{LnFn}$ '. At a national level, movement of ' $\text{LnTn}$ ' causes the movement of ' $\text{LnU}$ ', and vice versa. Bi-directional causality is only found between ' $\text{LnBn}$ ' and ' $\text{LnU}$ ' in Nelson/Marlborough/West Coast. ' $\text{LnU}$ ' causes ' $\text{LnTn}$ ' in Northland and Auckland.

The unemployment rate is a very important macroeconomic indicator. It may cause the movement of other macroeconomic variables or it may be caused by the movement of other macroeconomic indicators. The Crime and the state of the economy model is in a better position to measure the impact of the macroeconomics condition on crime because the model



covers every field in macroeconomics. In this case, the causality relationship between the unemployment rate and crime disappeared when more macroeconomic variables were included in the model. The average weekly income (LnW), and household debt and disposable income ratio (LnD) become the variables which cause movement in crime, especially for property related crime such as burglary and theft. Movement of 'LnW' causes the movement of 'LnBn', 'LnBr'. 'LnW' and 'LnD' both have bi-directional causality with 'LnTr' and 'LnTn'. The impact of macroeconomic variables on the number of people charged with fraud behaves differently compared with the number of people charged with theft or burglary. Movement of inflation causes the movement of 'LnFr' and 'LnFn'. The number of people charged with non-traffic offences has the same pattern as theft and burglary: the movement of 'LnD' causes the movement of 'LnNTn' and 'LnNTr'. The impulse response explains the size of the impact, and also how long the impact will stay in the system. In some cases, the statistical significance is not as important as the size of the impact between two variables.

Variable	Average Quarterly change		Average Annual change		Average	
	Mean	Maximum	Mean	Maximum	Mean	Maximum
avg_incomelog	0.011	0.031	0.03	0.064	6.448	6.794
Hld_debt_disp_income_ptlog	0.012	0.031	0.048	0.105	4.701	5.032
inflation	0.42%	2.80%	1.05%	4.30%	2.34%	5.3%
real_gdp_growth	0.55%	1.80%	1.79%	5.43%	2.84%	6.6%
unemployment_rate_allog	0.044	0.145	0.112	0.437	1.782	2.389
burglary_qtrlog	0.055	0.208	0.061	0.173	7.364	7.561
burglary_ratelog	0.056	0.205	0.063	0.166	-0.926	-0.705
fraud_qtrlog	0.078	0.241	0.077	0.294	7.134	7.469
fraud_ratelog	0.079	0.243	0.079	0.309	-1.156	-0.825
Theft_qtrlog	0.046	0.119	0.052	0.245	8.385	8.627
Theft_ratelog	0.047	0.117	0.054	0.226	0.095	0.282
non_traffic_qtrlog	0.037	0.106	0.071	0.236	10.007	10.393
non_traffic_ratelog	0.036	0.102	0.069	0.216	1.717	2.026

Table 6.1: Average quarterly change and average annual change for each variable

$$\text{Average Quarterly change} = \overline{|Y_t - Y_{t-1}|}$$

$$\text{Average Annual change} = \overline{|Y_t - Y_{t-4}|}$$

The impulse response produced in this study is based on a one unit shock to each variable. In fact, one unit shock is not common for most of the variables used in this study. The quarterly movement of crime or macroeconomic variables is quite small most of the time. Table 6.1 shows the average quarterly change and average annual change for each variable. The impulse response of crime variables are much bigger to the shocks to itself comparing with shocks to macroeconomic variables. This has been shown with ARIMA models, where all crime variables have significant AR lags. The maximum impulse response of crime variables to one unit shocks to macroeconomic variables is from 0.001% to 0.33% of the average value of crime variables. This is much smaller than the impulse response of crime variables to the shock itself. Impulse response of growth rate for number of people charged with non-traffic offence and theft are 1.5% and 5% of the average value when size of the shock is equal to average quarterly change. For fraud and burglary offence, impulse response of growth rate for number of people charged with fraud and burglary offence are 28.4% and 21.6% of the average value when size of the shock equals to average quarterly change. The impulse response moves between positive and negative, and it started to drop to a very small number after two years in most cases. It means that the accumulated impulse response in the long-run tends to be small, because adding positive and negative impulse responses in different periods will have a very small accumulated impulse response in the long-run. The impacts of macroeconomic variables to crime variables are very small. This means that macroeconomic variables do not have big predictive power for crime variables.

VAR or VECM only have AR components included and the impacts of macroeconomic variables are too small to change the current value of crime variables. From the forecasting point of view, the accuracy of forecasts from the VAR or VECM models are not improved when macroeconomic variables are included. ARIMA have different combinations of AR and MA terms; this makes ARIMA model a better forecasting model for many cases in this study. In general, VAR or VECM perform better in-sample and ARIMA models perform better in out-of-sample forecasts in this study. VAR or VECM cannot fit the data well when extra data are added to produce the final forecast. However, most of the ARIMA models fit the data well. As a result, ARIMA model is more stable than VECM or VAR model in this study. Multivariate models such as VAR or VECM are hard to build because they involve many parameters and model diagnostics. This means that VAR or VECM have a bigger possibility of not fitting the data well when new data are added to produce updated forecast. Tjostheim (1984) suggested that the accuracy of VAR model forecasts should outperform ARIMA

forecasts in theory. However, forecasts produced by VAR models may be more sensitive than the univariate model (Riise & Tjostheim, 1984 as cited in Fildes & Makridakis, 1995). VAR is a theory-free model, which is good when used to test relationships between macroeconomic variables and to analyse the response to random shocks (Sims, 1980). VECM or VAR model has been used to test the association between crime and the state of the economy in New Zealand. Impulse response is also included in this study. Most of the final forecasts are produced by ARIMA model.

In conclusion, significant causality relationships are found between the state of the economy and crime. Different offence categories relate to different macroeconomic variables. In general, variables measuring personal and household financial conditions such as personal weekly average income, household debt and disposable income ratios have a closer association with property related crime (theft and burglary). The Granger-Causality tests found that the growth of the unemployment rate causes the growth of the number of people charged with burglary, theft and non-traffic offences at a national level. This differs from a recent study by Janko and Popli (2015) in Canada. They found no long-term relationship between the crime rate and the unemployment rate, and there was only a short-term relationship found in fraud and robbery. Results from this study reach the same conclusion as did Cantor and Land in 1985. Cantor and Land (1985) suggested that the impact of unemployment on burglary and theft are statistically significant. However, the size of the impact is small.

At a regional level, all results from this study are the same as Small and Lewis' (1996) conclusion. For example, both studies found that there are causality relationships between the unemployment rate and the number of people charged with theft in Auckland and Northland regions. There are the same results for causality relations between the unemployment rate and burglary offences in Nelson/Marlborough/West Coast and unemployment and fraud offences in the Northland region. The Napier/Gisborne region has a different Granger-Causality test result, and hence it is more difficult to aggregate results from Small and Lewis' study. This study combines Gisborne and Hawke's Bay in order to match the court region from the Ministry of Justice, and the geographic region from Statistics New Zealand. For this reason, the result from the Gisborne/Hawke's Bay region cannot be compared to Small and Lewis' study.

Cohen and Felson (1979) conducted research to investigate the reason for the large increase of residential burglary during the 1960s and 1970s in the US. The conclusion was that increased motivation and opportunity contributed to the big increase in residential burglary. A large increase in ownership of lightweight electronic devices (for example, TVs and VCRs) was one of the causes for the increased motivation. The increase of opportunity is partly because of a higher rate of woman in the workforce (Cohen & Felson, 1979). In Cohen and Felson's research, a higher rate of women in the workforce means there were more unoccupied houses during the day, which created easier opportunities for burglaries. TVs and VCRs can be taken out of the house and sold easily. In this case, a big increase in the ownership of lightweight electronic devices and an increase of women in the workforce have created more crime targets for offenders. The same arguments are used to explain results from this study. To compare against other macroeconomic variables, this study recommends that the average weekly income and the household debt and the disposable income ratio have much closer relationships with burglary. People might choose to purchase more goods when their financial condition improves. This means there might be more crime targets available at some households. At the same time, retailers need to increase their stock in order to cope with consumer demand, furthermore some retailers might choose to open more shops. This creates more crime targets especially for shop lifting offenders. The household debt and disposable income ratio changed from an increasing trend to a stable trend in 2008, and then it started to drop from 2010. If the lower household debt and disposable income ratio means more crime targets are available, then the number of people charged with theft and burglary should increase. However, the number of people charged with theft and burglary started to drop from 2010. The household debt and disposable income ratio did start to drop from 2010, but it is still at a high level. At the same time, unemployment is still at a high level. Therefore, financial condition for many households is still too tough to increase their household expenditure. As a result, there are not big increases in crime targets. In fact, the recent crime trend cannot be explained by only using criminal opportunity theory. Both opportunity and motivation are needed to explain criminal activities (Cohen & Felson, 1979).

Criminal motivation theory explains and identifies the factors which drive potential offenders to commit a crime. In other words, criminal motivation theory suggests a change of state of the economy may cause the movement of crime. It will increase or decrease the likelihood to commit a crime for each potential offender (Cantor & Land, 1985). Monetary benefit is one of the motivations when offenders decide to commit a crime (Burt & Simons, 2003). The

level of monetary benefit gained from crime, especially property related crime, highly depends on consumer spending. In general, consumers spend less during an economic downturn and may increase their consumption when the economy is good. The only explanation of the recent drop in burglary and theft charges is that the high level of household debt and disposable income ratio may reduce consumer spending. The recent drop in the household debt and disposable income ratio may not be enough to improve a household's financial condition. For this reason, people may still choose to spend less. Therefore, it is harder for potential offenders to sell their stolen goods, and the potential benefits gained from committing crime become uncertain. This may reduce the motivation to commit crime, and hence the overall crime level may drop. From an income point of view, increased income means more people may choose to buy new items instead of second-hand items. This increases the total consumption, hence more crime targets are available. However, the demand for second-hand goods is going down; this reduces the monetary benefit of selling stolen items. Increased international online shopping may also play an important role in this case. Consumers can buy the same product from overseas retailers at a much lower price; this makes second-hand goods less attractive. The movement of inflation causes the growth rate of the number people charged with fraud to change. Criminal motivation theory might be able to explain the causality between inflation and fraud. The price for goods and services are lower when inflation is low. Lower inflation usually means low interest rates, and this will decrease the household debt and disposable income ratio. Therefore, household and individual financial condition will improve; this means lower motivation for fraud. The Granger-Causality test confirms that there are associations between economically motivated crimes and macroeconomic variables measuring personal and household financial conditions. Those relationships are statistically significant. However, impulse response and parameter estimates confirm that the size of the impact is too small to give each variable a noticeable change, and the sign of the relationship is hard to conclude because they move between positive and negative. In fact, the sign of the relationship becomes less important when movements of macroeconomic variables are not big enough to change the crime variable and vice versa.

By using macroeconomic variables, both criminal opportunity and motivation theory cannot fully explain the recent crime downturn. Impulse responses confirm that the biggest impact of crime variables came from shocks to itself. Other factors such as government policy can have a huge impact on the movement of crime. Those factors are not included in this study. In

2014, the Ministry of Justice (2014) had no clear explanation for the fall in crime since 2010. Policing excellence was one of the policies which partly contributed to the reduction of prosecution volume. Forecasts produced by this study suggest that the number of people charged with burglary, theft, fraud and non-traffic offences will continue to decrease up until 2019, but at a lower rate of reduction. The Justice Sector Forecast from 2014 suggests that the number of court disposals changed from a decreasing trend to a flat trend. The forecast is dependent on how the criminal justice system dealt with crime. It is based on assumptions from other justice sector agencies, and those assumptions are not included in the forecasting models in this thesis (Ministry of Justice, 2014). This might be the reason for a different forecast trend between this study and the Ministry of Justice's forecasts.

# Bibliography

- Bober, S. (1968). *The economics of cycles and growth*. New York: Wiley.
- Breusch, T. (1978). Testing for Autocorrelation in Dynamic Linear Models. *Australian Economic Papers*, 17(31), 334-355.
- Brooks, C. (2002). *Introductory econometrics for finance*. Cambridge; New York: Cambridge University Press.
- Burnham, K., & Anderson, D. (2002). *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*.
- Burnham, K., & Anderson, D. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33(2), 261-304.
- Burt, C., & Simons, R. (2013). Self-Control, Thrill Seeking, and Crime. *Criminal Justice and Behavior*, 40(11), 1326-1348.
- Bushway, S, Cook, P. J. & Phillips, M. (2012). The Overall Effect of the Business Cycle on Crime. *German Economic Review*, 13(4): 436–446.
- Cantor, D., & Land, K. (1985). Unemployment and Crime Rates in the Post-World War II United States: A Theoretical and Empirical Analysis. *American Sociological Review*, 50(4), 317.
- Cohen, L., & Felson, M. (1979). Social Change and Crime Rate Trends: A Routine Activities Approach. *American Sociological Review*, 44(4), 588.
- Davidson, R., & MacKinnon, J. (2004). *Econometric theory and methods*. New York: Oxford University Press.
- Elliott, G., & Timmermann, A. (2016). Forecasting in Economics and Finance. *Annual Review of Economics*, 8, 81-110.

- Engle, R. (1982). Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4), 987.
- Fildes, R., & Makridakis, S. (1995). The impact of empirical accuracy studies on time series analysis and forecasting. *International Statistical Review = Revue Internationale De Statistique*, 63(3), 289.
- Godfrey, L. (1978). Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables. *Econometrica*, 46(6), 1293.
- Granger, C. (1980). *Forecasting in business and economics / C. W. J. Granger*. New York: Academic Press.
- Hannan, E., & Quinn, B. (1979). The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(2), 190-195.
- Hosking, J. (1980). The Multivariate Portmanteau Statistic. *Journal of the American Statistical Association*, 75(371), 602-608.
- Janko, Z. & Popli, G. (2015). Examining the link between crime and unemployment: a time-series analysis for Canada. *Applied Economics*, 47 (37). pp. 4007-4019.  
ISSN: 0003-6846
- Ljung, G., & Box, G. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 65(2), 297-303.
- Mankiw, N. (2000). *Principles of economics*. (2nd ed.). Fort Worth, Tex.; London: Harcourt Brace College.
- Ministry of Business, Innovation and Employment. (2014). *Regional Economic Activity Report*. ISSN: 2324-2698
- Ministry of Justice. (2012). *Justice Sector Forecast 2012-2022*. ISSN: 1179-1446.  
Retrieved 20 February 2017 from the World Wide Web:  
<https://www.justice.govt.nz/assets/Documents/Publications/2012-forecast-report.pdf>



- Ministry of Justice. (2014). *Justice Sector Forecast: Prison Population Forecast 2014-2024*. ISSN 1179-1446. Retrieved 20 February 2017 from the World Wide Web:  
<https://www.justice.govt.nz/assets/Documents/Publications/Prison-forecast-2014-2024.pdf>
- Newbold, P., & Bos, T. (1994). *Introductory business & economic forecasting* (2nd ed.). Cincinnati: South-Western Pub.
- Parry, G. & Kemp, S. (2000). *Exploring macroeconomics*. Perth: Tactic Publications.
- Riise, T., & Tjostheim, D. (1984). Theory and Practice of Multivariate ARMA Forecasting. *Journal of Forecasting*, 3(3), 309.
- SAS Institute Inc. (2014). *SAS/ETS 13.2 User's Guide*. Cary, NC: SAS Institute Inc.  
Retrieved 18 March 2017 from the World Wide Web:  
<http://support.sas.com/documentation/cdl/en/etsug/67525/PDF/default/etsug.pdf>
- Shumway, R., & Stoffer, D. (2011). *Time Series Analysis and Its Applications: With R Examples*, Springer eBooks.
- Sims, C. (1980). Macroeconomics and Reality. *Econometrica*, 48(1), 1.
- Small, J. P. & Lewis, C. (1996). *Economic crime and unemployment in New Zealand: causation or coincidence?* Auckland: Department of Economics, University of Auckland
- Verbeek, M. (2000). *A guide to modern econometrics*. New York: Wiley.
- Wooldridge, J. M. (2009). *Introductory econometrics: A modern approach*. Mason, OH: South Western, Cengage Learning.

# Appendix A Unemployment and crime model

## Non-traffic and unemployment model

Appendix Table 1: Parameter estimate for unemployment and non-traffic model, VECM (6)

Note: This table presents parameter estimate for VECM (6) using seasonal differencing

Equation	Estimate	Pr >  t	Variable
D_non_traffic_qtrlog	0.07593	0.0238	1
	-0.01419	0.1732	easter(t)
	-0.1494		non_traffic_qtrlog(t-1)
	-0.04301		unemployment_rate_all1log(t-1)
	0.20109	0.0874	D_non_traffic_qtrlog(t-1)
	0.05792	0.5352	D_unemployment_rate_all1log(t-1)
	0.13744	0.161	D_non_traffic_qtrlog(t-2)
	-0.03694	0.6877	D_unemployment_rate_all1log(t-2)
	0.24058	0.017	D_non_traffic_qtrlog(t-3)
	-0.02729	0.7884	D_unemployment_rate_all1log(t-3)
	-0.42476	0.0001	D_non_traffic_qtrlog(t-4)
	0.15825	0.088	D_unemployment_rate_all1log(t-4)
	0.03584	0.752	D_non_traffic_qtrlog(t-5)
	-0.16271	0.0845	D_unemployment_rate_all1log(t-5)
D_unemployment_rate_all1log	0.0942	0.0171	1
	0.00728	0.5486	easter(t)
	-0.19327		non_traffic_qtrlog(t-1)
	-0.05564		unemployment_rate_all1log(t-1)
	-0.12697	0.3532	D_non_traffic_qtrlog(t-1)
	0.04068	0.7101	D_unemployment_rate_all1log(t-1)
	0.19407	0.0926	D_non_traffic_qtrlog(t-2)
	0.47504	0.0001	D_unemployment_rate_all1log(t-2)
	0.24517	0.0369	D_non_traffic_qtrlog(t-3)
	-0.02061	0.8627	D_unemployment_rate_all1log(t-3)
	-0.07627	0.5249	D_non_traffic_qtrlog(t-4)
	-0.10397	0.3351	D_unemployment_rate_all1log(t-4)
	-0.18783	0.1611	D_non_traffic_qtrlog(t-5)
	0.22408	0.0437	D_unemployment_rate_all1log(t-5)

## Burglary and unemployment model

Appendix Table 2: Parameter estimate for unemployment and burglary model, VAR (4)

Note: This table presents parameter estimate for VAR (4) using seasonal dummies

Equation	Estimate	Pr >  t	Variable
burglary_qtrlog	0.00221	0.925	1
	0.06835	0.0179	S_1t
	-0.01443	0.6624	S_2t
	-0.05818	0.029	S_3t
	-0.01539	0.4384	easter(t)
	-0.16538	0.1788	burglary_qtrlog(t-1)
	0.14219	0.2045	unemployment_rate_all1log(t-1)
	-0.13437	0.27	burglary_qtrlog(t-2)
	0.06446	0.5628	unemployment_rate_all1log(t-2)
	-0.05816	0.6285	burglary_qtrlog(t-3)
	-0.24624	0.0275	unemployment_rate_all1log(t-3)
	-0.04972	0.6713	burglary_qtrlog(t-4)
	-0.02023	0.8593	unemployment_rate_all1log(t-4)
unemployment_rate_all1log	-0.00827	0.7546	1
	-0.00792	0.8035	S_1t
	-0.00108	0.9769	S_2t
	0.02344	0.4277	S_3t
	0.01037	0.6427	easter(t)
	0.06984	0.6123	burglary_qtrlog(t-1)
	0.06366	0.6124	unemployment_rate_all1log(t-1)
	-0.05487	0.6883	burglary_qtrlog(t-2)
	0.50313	0.0001	unemployment_rate_all1log(t-2)
	-0.02354	0.8619	burglary_qtrlog(t-3)
	0.08984	0.4683	unemployment_rate_all1log(t-3)
	-0.00444	0.9731	burglary_qtrlog(t-4)
	-0.11625	0.3677	unemployment_rate_all1log(t-4)

## Fraud and unemployment rate

Appendix Table 3: Parameter estimate for unemployment and fraud model, VAR (2)

Note: This table presents parameter estimate for VAR (2) using seasonal dummies

Equation	Estimate	Pr >  t	Variable
Fraud_qtrlog	0.04516	0.1801	1
	0.01935	0.6118	S_1t
	-0.09887	0.0264	S_2t
	-0.11025	0.0018	S_3t
	-0.02197	0.4497	easter(t)
	-0.19059	0.1147	Fraud_qtrlog(t-1)
	-0.02556	0.8577	unemployment_rate_all1log(t-1)
	-0.23073	0.0606	Fraud_qtrlog(t-2)
	-0.02609	0.8526	unemployment_rate_all1log(t-2)
unemployment_rate_all1log	-0.00789	0.7576	1
	-0.00211	0.9421	S_1t
	0.00042	0.99	S_2t
	0.01192	0.648	S_3t
	0.01228	0.5796	easter(t)
	0.04645	0.612	Fraud_qtrlog(t-1)
	0.08768	0.4214	unemployment_rate_all1log(t-1)
	0.05285	0.5692	Fraud_qtrlog(t-2)
	0.44202	0.0001	unemployment_rate_all1log(t-2)

## Theft and Unemployment rate and model

Appendix Table 4: Parameter estimate for unemployment and theft model, VECM (4)

Note: This table presents parameter estimate for VECM (4) using seasonal dummies

Equation	Estimate	Pr >  t	Variable
D_Theft_qtrlog	1.00428	0.0041	1
	0.06395	0.0007	S_1t
	0.07738	0.0001	S_2t
	0.00203	0.8962	S_3t
	-0.03499	0.0207	easter(t)
	-0.11018		Theft_qtrlog(t-1)
	-0.06172		unemployment_rate_all1log(t-1)
	-0.00457	0.9647	D_Theft_qtrlog(t-1)
	0.07514	0.3734	D_unemployment_rate_all1log(t-1)
	-0.11703	0.242	D_Theft_qtrlog(t-2)
	0.06064	0.4278	D_unemployment_rate_all1log(t-2)
	-0.01522	0.8796	D_Theft_qtrlog(t-3)
	-0.08174	0.3343	D_unemployment_rate_all1log(t-3)
D_unemployment_rate_all1log	0.97064	0.0338	1
	-0.04581	0.0588	S_1t
	-0.03387	0.1208	S_2t
	-0.01376	0.5066	S_3t
	0.01351	0.4931	easter(t)
	-0.10157		Theft_qtrlog(t-1)
	-0.05689		unemployment_rate_all1log(t-1)
	0.11909	0.3868	D_Theft_qtrlog(t-1)
	0.0879	0.4327	D_unemployment_rate_all1log(t-1)
	-0.15379	0.2468	D_Theft_qtrlog(t-2)
	0.45244	0.0001	D_unemployment_rate_all1log(t-2)
	0.19388	0.1493	D_Theft_qtrlog(t-3)
	0.17407	0.1235	D_unemployment_rate_all1log(t-3)

# Appendix B Crime association model

## Number of people charged

Appendix Table 5: Parameter estimate for crime association model, VECM (5) (Model 4.6)

Notes: This table presents s parameter estimate for VECM (5) using seasonal dummies. Model is built based on number of people charged

Equation	Estimate	Pr >  t	Variable
D_Theft_qtrlog	0.62534	0.0039	1
	0.07051	0.001	S_1t
	0.08571	0.0005	S_2t
	0.01857	0.3362	S_3t
	-0.03911	0.0064	easter(t)
	0.25377		Theft_qtrlog(t-1)
	0.00725		Fraud_qtrlog(t-1)
	-0.38574		burglary_qtrlog(t-1)
	-0.36394	0.0266	D_Theft_qtrlog(t-1)
	0.17179	0.0137	D_Fraud_qtrlog(t-1)
	0.24584	0.0616	D_burglary_qtrlog(t-1)
	-0.49949	0.0029	D_Theft_qtrlog(t-2)
	0.01776	0.8092	D_Fraud_qtrlog(t-2)
	0.51534	0.0001	D_burglary_qtrlog(t-2)
	-0.21901	0.1375	D_Theft_qtrlog(t-3)
	0.09548	0.1857	D_Fraud_qtrlog(t-3)
	0.21607	0.0948	D_burglary_qtrlog(t-3)
	0.15651	0.2442	D_Theft_qtrlog(t-4)
	0.02778	0.6739	D_Fraud_qtrlog(t-4)
	-0.04367	0.7017	D_burglary_qtrlog(t-4)
D_Fraud_qtrlog	-0.62302	0.109	1
	0.15591	0.0001	S_1t
	0.11504	0.0096	S_2t
	-0.03001	0.3979	S_3t
	-0.03135	0.2226	easter(t)
	-0.21631		Theft_qtrlog(t-1)
	-0.00618		Fraud_qtrlog(t-1)
	0.32879		burglary_qtrlog(t-1)
	0.83946	0.006	D_Theft_qtrlog(t-1)
	-0.17521	0.1644	D_Fraud_qtrlog(t-1)
	-0.52932	0.0296	D_burglary_qtrlog(t-1)
	0.17041	0.5668	D_Theft_qtrlog(t-2)
	-0.23004	0.0929	D_Fraud_qtrlog(t-2)
	-0.11687	0.609	D_burglary_qtrlog(t-2)

	0.03519	0.8958	D_Theft_qtrlog(t-3)
	-0.01403	0.9152	D_Fraud_qtrlog(t-3)
	-0.30797	0.1935	D_burglary_qtrlog(t-3)
	0.56273	0.0251	D_Theft_qtrlog(t-4)
	-0.15442	0.2063	D_Fraud_qtrlog(t-4)
	-0.26681	0.2062	D_burglary_qtrlog(t-4)
D_burglary_qtrlog	0.98029	0.0002	1
	0.04931	0.042	S_1t
	0.11459	0.0001	S_2t
	0.04024	0.0776	S_3t
	-0.01635	0.3158	easter(t)
	0.3957		Theft_qtrlog(t-1)
	0.01131		Fraud_qtrlog(t-1)
	-0.60147		burglary_qtrlog(t-1)
	-0.17257	0.3601	D_Theft_qtrlog(t-1)
	0.07907	0.3215	D_Fraud_qtrlog(t-1)
	0.15207	0.3175	D_burglary_qtrlog(t-1)
	-0.47459	0.0143	D_Theft_qtrlog(t-2)
	0.04131	0.6311	D_Fraud_qtrlog(t-2)
	0.20564	0.1598	D_burglary_qtrlog(t-2)
	-0.10527	0.5383	D_Theft_qtrlog(t-3)
	0.16595	0.0512	D_Fraud_qtrlog(t-3)
	0.06166	0.6801	D_burglary_qtrlog(t-3)
	-0.05018	0.7482	D_Theft_qtrlog(t-4)
	-0.04532	0.5574	D_Fraud_qtrlog(t-4)
	0.07401	0.579	D_burglary_qtrlog(t-4)

## Crime rate

Appendix Table 6: Parameter estimate for crime association model, VAR (1, 2, 4) (Model 4.7)

Notes: This table presents s parameter estimate for VAR (1, 2, 4) using seasonal dummies. Model is built based on crime rate

Equation	Estimate	Pr >  t	Variable
Theft_ratelog	0.04532	0.0244	1
	0.00496	0.8356	S_1t
	-0.07518	0.01	S_2t
	-0.08332	0.0004	S_3t
	-0.04102	0.012	easter(t)
	-0.03203	0.7937	Theft_ratelog(t-1)
	0.11672	0.1301	fraud_ratelog(t-1)
	-0.03449	0.7465	burglary_ratelog(t-1)
	-0.18099	0.1901	Theft_ratelog(t-2)
	-0.06873	0.3425	fraud_ratelog(t-2)

	0.27945	0.0114	burglary_ratelog(t-2)
	0.33449	0.0146	Theft_ratelog(t-4)
	-0.01413	0.8469	fraud_ratelog(t-4)
	-0.20927	0.0547	burglary_ratelog(t-4)
fraud_ratelog	0.07091	0.0476	1
	-0.0205	0.631	S_1t
	-0.14479	0.0057	S_2t
	-0.13059	0.0016	S_3t
	-0.03025	0.29	easter(t)
	0.5106	0.0221	Theft_ratelog(t-1)
	-0.12126	0.3757	fraud_ratelog(t-1)
	-0.25992	0.1755	burglary_ratelog(t-1)
	-0.12417	0.6126	Theft_ratelog(t-2)
	-0.17278	0.1829	fraud_ratelog(t-2)
	0.11363	0.555	burglary_ratelog(t-2)
	0.43595	0.0717	Theft_ratelog(t-4)
	-0.14172	0.2802	fraud_ratelog(t-4)
	-0.09872	0.607	burglary_ratelog(t-4)
burglary_ratelog	0.00876	0.7304	1
	0.05603	0.0718	S_1t
	-0.03149	0.3911	S_2t
	-0.06336	0.0297	S_3t
	-0.02007	0.3296	easter(t)
	0.21897	0.1678	Theft_ratelog(t-1)
	-0.00121	0.9902	fraud_ratelog(t-1)
	-0.22908	0.0987	burglary_ratelog(t-1)
	-0.06488	0.7134	Theft_ratelog(t-2)
	-0.04205	0.6509	fraud_ratelog(t-2)
	-0.07968	0.5657	burglary_ratelog(t-2)
	0.20094	0.2458	Theft_ratelog(t-4)
	-0.10468	0.2685	fraud_ratelog(t-4)
	-0.11442	0.4088	burglary_ratelog(t-4)



# Appendix C Crime and the state of the economy model

## Non-traffic and the state of the economy model

### Full model

Appendix Table 7: Parameter estimate for economic and non-traffic full model, VAR (4)

Notes: This table presents parameter estimate for VAR (4) using seasonal differencing.

Endogenous variable: non\_traffic\_qtrlog(1,4), non\_traffic\_ratelog(1,4), avg\_incomelog(1,4), Hld\_debt\_disp\_income\_pt log(1,4), real\_gdp\_growth (1), inflation (1), unemployment\_rate\_all (1)

Exogenous variable: Easter

Equation	Estimate	Pr >  t	Variable
non_traffic_qtrlog	-0.00252	0.649	1
	-0.36903	0.9753	non_traffic_qtrlog(t-1)
	0.29563	0.9802	non_traffic_ratelog(t-1)
	0.55516	0.4909	avg_incomelog(t-1)
	0.01048	0.6594	unemployment_rate_all1(t-1)
	-0.01383	0.4989	real_gdp_growth1(t-1)
	-0.00939	0.3979	inflation1(t-1)
	2.15128	0.0509	Hld_debt_disp_income_ptlog(t-1)
	7.63969	0.6548	non_traffic_qtrlog(t-2)
	-7.55944	0.6584	non_traffic_ratelog(t-2)
	1.21195	0.141	avg_incomelog(t-2)
	-0.01115	0.6238	unemployment_rate_all1(t-2)
	0.03105	0.2493	real_gdp_growth1(t-2)
	0.00762	0.4821	inflation1(t-2)
	-0.25589	0.8515	Hld_debt_disp_income_ptlog(t-2)
	-2.64821	0.8766	non_traffic_qtrlog(t-3)
	2.9206	0.864	non_traffic_ratelog(t-3)
	1.24935	0.0947	avg_incomelog(t-3)
	-0.01338	0.5286	unemployment_rate_all1(t-3)
	-0.00721	0.7818	real_gdp_growth1(t-3)
	0.00048	0.9619	inflation1(t-3)
	0.52013	0.7101	Hld_debt_disp_income_ptlog(t-3)
	-1.71011	0.885	non_traffic_qtrlog(t-4)
	1.37112	0.9078	non_traffic_ratelog(t-4)
	0.02136	0.9784	avg_incomelog(t-4)
	0.0313	0.1973	unemployment_rate_all1(t-4)
	-0.01324	0.5138	real_gdp_growth1(t-4)

	-0.00227	0.8214	inflation1(t-4)
	-1.34564	0.225	Hld_debt_disp_income_ptlog(t-4)
non_traffic_ratelog	-0.00248	0.6541	1
	-1.31707	0.9122	non_traffic_qtrlog(t-1)
	1.24478	0.917	non_traffic_ratelog(t-1)
	0.54919	0.4966	avg_incomelog(t-1)
	0.01013	0.6707	unemployment_rate_all1(t-1)
	-0.01399	0.495	real_gdp_growth1(t-1)
	-0.00932	0.4023	inflation1(t-1)
	2.15488	0.051	Hld_debt_disp_income_ptlog(t-1)
	7.73521	0.6515	non_traffic_qtrlog(t-2)
	-7.65514	0.6551	non_traffic_ratelog(t-2)
	1.19712	0.1467	avg_incomelog(t-2)
	-0.01092	0.6316	unemployment_rate_all1(t-2)
	0.03125	0.2474	real_gdp_growth1(t-2)
	0.00764	0.4815	inflation1(t-2)
	-0.27528	0.8407	Hld_debt_disp_income_ptlog(t-2)
	-2.44664	0.8862	non_traffic_qtrlog(t-3)
	2.71852	0.8736	non_traffic_ratelog(t-3)
	1.24286	0.0972	avg_incomelog(t-3)
	-0.01296	0.5427	unemployment_rate_all1(t-3)
	-0.00736	0.7781	real_gdp_growth1(t-3)
	0.00033	0.9738	inflation1(t-3)
	0.52174	0.7099	Hld_debt_disp_income_ptlog(t-3)
	-1.79703	0.8795	non_traffic_qtrlog(t-4)
	1.46209	0.9019	non_traffic_ratelog(t-4)
	0.01506	0.9848	avg_incomelog(t-4)
	0.03138	0.1972	unemployment_rate_all1(t-4)
	-0.01297	0.5234	real_gdp_growth1(t-4)
	-0.0021	0.8352	inflation1(t-4)
	-1.35144	0.2241	Hld_debt_disp_income_ptlog(t-4)
avg_incomelog	0.00038	0.6751	1
	-0.57568	0.7683	non_traffic_qtrlog(t-1)
	0.5871	0.764	non_traffic_ratelog(t-1)
	-0.20921	0.1172	avg_incomelog(t-1)
	-0.00253	0.5159	unemployment_rate_all1(t-1)
	-0.00749	0.0291	real_gdp_growth1(t-1)
	-0.00153	0.3998	inflation1(t-1)
	-0.12277	0.4887	Hld_debt_disp_income_ptlog(t-1)
	-0.49896	0.8585	non_traffic_qtrlog(t-2)
	0.51183	0.8549	non_traffic_ratelog(t-2)
	0.06608	0.6207	avg_incomelog(t-2)
	-0.00074	0.8424	unemployment_rate_all1(t-2)
	0.00878	0.0499	real_gdp_growth1(t-2)
	0.00217	0.2237	inflation1(t-2)
	-0.10901	0.6269	Hld_debt_disp_income_ptlog(t-2)

	3.06377	0.276	non_traffic_qtrlog(t-3)
	-3.0721	0.2748	non_traffic_ratelog(t-3)
	-0.03916	0.7457	avg_incomelog(t-3)
	-0.00661	0.0618	unemployment_rate_all1(t-3)
	0.00015	0.9724	real_gdp_growth1(t-3)
	0.00238	0.1564	inflation1(t-3)
	0.19958	0.3857	Hld_debt_disp_income_ptlog(t-3)
	-2.20255	0.259	non_traffic_qtrlog(t-4)
	2.18949	0.2623	non_traffic_ratelog(t-4)
	-0.55584	0.0001	avg_incomelog(t-4)
	-0.00013	0.9746	unemployment_rate_all1(t-4)
	-0.00286	0.3908	real_gdp_growth1(t-4)
	0.00014	0.9321	inflation1(t-4)
	0.0983	0.5862	Hld_debt_disp_income_ptlog(t-4)
unemployment_rate_all1	-0.03521	0.3359	1
	49.32584	0.5308	non_traffic_qtrlog(t-1)
	-49.8536	0.5266	non_traffic_ratelog(t-1)
	-8.85999	0.0994	avg_incomelog(t-1)
	-0.02465	0.8749	unemployment_rate_all1(t-1)
	-0.22451	0.1	real_gdp_growth1(t-1)
	-0.10337	0.1607	inflation1(t-1)
	-9.06009	0.2065	Hld_debt_disp_income_ptlog(t-1)
	-10.9557	0.9224	non_traffic_qtrlog(t-2)
	12.14653	0.9141	non_traffic_ratelog(t-2)
	1.34611	0.8018	avg_incomelog(t-2)
	0.39993	0.01	unemployment_rate_all1(t-2)
	0.00311	0.9859	real_gdp_growth1(t-2)
	-0.01899	0.7898	inflation1(t-2)
	5.17026	0.5665	Hld_debt_disp_income_ptlog(t-2)
	-24.3393	0.8285	non_traffic_qtrlog(t-3)
	24.99711	0.824	non_traffic_ratelog(t-3)
	-8.24458	0.0943	avg_incomelog(t-3)
	0.09931	0.4782	unemployment_rate_all1(t-3)
	0.20159	0.2434	real_gdp_growth1(t-3)
	0.12327	0.07	inflation1(t-3)
	-6.75487	0.4648	Hld_debt_disp_income_ptlog(t-3)
	-10.0097	0.8978	non_traffic_qtrlog(t-4)
	9.01011	0.908	non_traffic_ratelog(t-4)
	-5.62237	0.2826	avg_incomelog(t-4)
	-0.12445	0.4339	unemployment_rate_all1(t-4)
	-0.23267	0.086	real_gdp_growth1(t-4)
	0.0342	0.6067	inflation1(t-4)
	-1.58901	0.8265	Hld_debt_disp_income_ptlog(t-4)
real_gdp_growth1	-0.02319	0.5709	1
	-166.086	0.0643	non_traffic_qtrlog(t-1)
	165.8798	0.0647	non_traffic_ratelog(t-1)

	8.04701	0.18	avg_incomelog(t-1)
	-0.20238	0.2524	unemployment_rate_all1(t-1)
	0.90953	0.0001	real_gdp_growth1(t-1)
	-0.01117	0.8913	inflation1(t-1)
	-10.2588	0.2024	Hld_debt_disp_income_ptlog(t-1)
	186.3547	0.1444	non_traffic_qtrlog(t-2)
	-186.885	0.1434	non_traffic_ratelog(t-2)
	7.60451	0.21	avg_incomelog(t-2)
	-0.18906	0.2631	unemployment_rate_all1(t-2)
	-0.052	0.7927	real_gdp_growth1(t-2)
	-0.0805	0.3161	inflation1(t-2)
	14.58099	0.1534	Hld_debt_disp_income_ptlog(t-2)
	16.48704	0.8959	non_traffic_qtrlog(t-3)
	-15.507	0.9021	non_traffic_ratelog(t-3)
	12.00487	0.0316	avg_incomelog(t-3)
	0.31531	0.0488	unemployment_rate_all1(t-3)
	-0.24011	0.2159	real_gdp_growth1(t-3)
	-0.0158	0.833	inflation1(t-3)
	-2.47832	0.8105	Hld_debt_disp_income_ptlog(t-3)
	-50.9073	0.5609	non_traffic_qtrlog(t-4)
	50.30181	0.566	non_traffic_ratelog(t-4)
	5.66413	0.334	avg_incomelog(t-4)
	0.15652	0.3807	unemployment_rate_all1(t-4)
	-0.14183	0.3454	real_gdp_growth1(t-4)
	0.00747	0.9201	inflation1(t-4)
	0.0583	0.9943	Hld_debt_disp_income_ptlog(t-4)
inflation1	-0.03998	0.5186	1
	228.5249	0.0915	non_traffic_qtrlog(t-1)
	-225.33	0.0961	non_traffic_ratelog(t-1)
	1.32004	0.8833	avg_incomelog(t-1)
	0.22859	0.3915	unemployment_rate_all1(t-1)
	-0.06121	0.7886	real_gdp_growth1(t-1)
	0.15438	0.216	inflation1(t-1)
	-25.8125	0.0368	Hld_debt_disp_income_ptlog(t-1)
	-313.17	0.1059	non_traffic_qtrlog(t-2)
	313.7602	0.1055	non_traffic_ratelog(t-2)
	-3.88997	0.6694	avg_incomelog(t-2)
	0.54622	0.0358	unemployment_rate_all1(t-2)
	0.44061	0.1457	real_gdp_growth1(t-2)
	-0.00162	0.9893	inflation1(t-2)
	5.27782	0.73	Hld_debt_disp_income_ptlog(t-2)
	115.0742	0.547	non_traffic_qtrlog(t-3)
	-116.674	0.5416	non_traffic_ratelog(t-3)
	-23.2305	0.0068	avg_incomelog(t-3)
	-0.81934	0.0011	unemployment_rate_all1(t-3)
	-0.54384	0.0668	real_gdp_growth1(t-3)

	-0.07372	0.5166	inflation1(t-3)
	-6.72124	0.6677	Hld_debt_disp_income_ptlog(t-3)
	-22.3535	0.8658	non_traffic_qtrlog(t-4)
	21.00239	0.874	non_traffic_ratelog(t-4)
	-13.9061	0.12	avg_incomelog(t-4)
	-0.21691	0.4218	unemployment_rate_all1(t-4)
	0.4676	0.0434	real_gdp_growth1(t-4)
	-0.43698	0.0003	inflation1(t-4)
	0.22281	0.9856	Hld_debt_disp_income_ptlog(t-4)
Hld_debt_disp_income_ptlog	-0.00081	0.2528	1
	0.61171	0.6869	non_traffic_qtrlog(t-1)
	-0.62647	0.6799	non_traffic_ratelog(t-1)
	-0.11874	0.2491	avg_incomelog(t-1)
	0.00664	0.032	unemployment_rate_all1(t-1)
	0.00121	0.6402	real_gdp_growth1(t-1)
	-0.00074	0.6008	inflation1(t-1)
	0.60784	0.0001	Hld_debt_disp_income_ptlog(t-1)
	0.95182	0.6615	non_traffic_qtrlog(t-2)
	-0.97951	0.6525	non_traffic_ratelog(t-2)
	-0.05036	0.627	avg_incomelog(t-2)
	0.00412	0.1585	unemployment_rate_all1(t-2)
	-0.00038	0.9114	real_gdp_growth1(t-2)
	-0.00071	0.6035	inflation1(t-2)
	0.08709	0.6168	Hld_debt_disp_income_ptlog(t-2)
	0.42856	0.8434	non_traffic_qtrlog(t-3)
	-0.43733	0.8403	non_traffic_ratelog(t-3)
	-0.0382	0.6837	avg_incomelog(t-3)
	-0.00369	0.1751	unemployment_rate_all1(t-3)
	0.00294	0.3763	real_gdp_growth1(t-3)
	0.00126	0.3306	inflation1(t-3)
	0.06774	0.7035	Hld_debt_disp_income_ptlog(t-3)
	-1.72839	0.2539	non_traffic_qtrlog(t-4)
	1.75123	0.2483	non_traffic_ratelog(t-4)
	0.1	0.3218	avg_incomelog(t-4)
	-0.00653	0.0372	unemployment_rate_all1(t-4)
	0.00003	0.9897	real_gdp_growth1(t-4)
	0.00007	0.956	inflation1(t-4)
	-0.4016	0.0059	Hld_debt_disp_income_ptlog(t-4)

## Reduced Model

**Appendix Table 8: Parameter estimate for reduced non-traffic and the state of the economy model, VAR (4) (Model 4.2)**

Endogenous variable: non\_traffic\_qtrlog(1,4), non\_traffic\_ratelog (1,4),

Hld\_debt\_disp\_income\_ptlog(1,4)

Exogenous variable: Prin1 Prin2 Easter

**Note: This table presents parameter estimate for VAR (4) using seasonal differencing**

Equation	Estimate	Pr >  t	Variable
non_traffic_qtrlog	0.00094	0.8764	1
	-0.01248	0.2791	easter(t)
	-0.00016	0.9726	Prin1(t)
	0.00302	0.5263	Prin2(t)
	1.23111	0.9022	non_traffic_qtrlog(t-1)
	-1.26939	0.8992	non_traffic_ratelog(t-1)
	1.58077	0.083	Hld_debt_disp_income_ptlog(t-1)
	10.44018	0.4782	non_traffic_qtrlog(t-2)
	-10.4445	0.4782	non_traffic_ratelog(t-2)
	-0.06662	0.9534	Hld_debt_disp_income_ptlog(t-2)
	-3.42863	0.8153	non_traffic_qtrlog(t-3)
	3.66317	0.803	non_traffic_ratelog(t-3)
	0.26612	0.8124	Hld_debt_disp_income_ptlog(t-3)
	-4.17156	0.6826	non_traffic_qtrlog(t-4)
	3.8251	0.7077	non_traffic_ratelog(t-4)
	-0.99452	0.2677	Hld_debt_disp_income_ptlog(t-4)
non_traffic_ratelog	0.00094	0.8769	1
	-0.01247	0.2797	easter(t)
	-0.00015	0.9736	Prin1(t)
	0.00303	0.5255	Prin2(t)
	0.1796	0.9857	non_traffic_qtrlog(t-1)
	-0.21785	0.9827	non_traffic_ratelog(t-1)
	1.59158	0.0812	Hld_debt_disp_income_ptlog(t-1)
	10.58663	0.4725	non_traffic_qtrlog(t-2)
	-10.5898	0.4725	non_traffic_ratelog(t-2)
	-0.07943	0.9445	Hld_debt_disp_income_ptlog(t-2)
	-3.27625	0.8235	non_traffic_qtrlog(t-3)
	3.51093	0.8112	non_traffic_ratelog(t-3)
	0.25678	0.819	Hld_debt_disp_income_ptlog(t-3)
	-4.18742	0.6818	non_traffic_qtrlog(t-4)
	3.8446	0.7065	non_traffic_ratelog(t-4)
	-0.9988	0.2661	Hld_debt_disp_income_ptlog(t-4)
Hld_debt_disp_income_ptlog	-0.00119	0.1414	1
	0.00144	0.3446	easter(t)
	-0.00013	0.8325	Prin1(t)

	0.00097	0.1282	Prin2(t)
	1.27029	0.3383	non_traffic_qtrlog(t-1)
	-1.26716	0.3397	non_traffic_ratelog(t-1)
	0.75309	0.0001	Hld_debt_disp_income_ptlog(t-1)
	0.27724	0.8862	non_traffic_qtrlog(t-2)
	-0.30118	0.8765	non_traffic_ratelog(t-2)
	-0.06903	0.6467	Hld_debt_disp_income_ptlog(t-2)
	0.74072	0.7021	non_traffic_qtrlog(t-3)
	-0.74695	0.6998	non_traffic_ratelog(t-3)
	0.04321	0.7701	Hld_debt_disp_income_ptlog(t-3)
	-1.90588	0.1599	non_traffic_qtrlog(t-4)
	1.93199	0.1543	non_traffic_ratelog(t-4)
	-0.4558	0.0003	Hld_debt_disp_income_ptlog(t-4)

## Theft and the state of the economy model

### Full Model

**Appendix Table 9: Parameter estimate for full theft and the state of the economy model, VAR (4)**

Endogenous variable: Theft\_qtrlog(1,4), Theft\_ratelog(1,4), avg\_incomelog(1,4),

Hld\_debt\_disp\_income\_ptlog(1,4), real\_gdp\_growth (1), inflation (1), unemployment\_rate\_all (1)

Exogenous variable: Easter

Notes: This table presents s parameter estimate for VAR (4) using seasonal differencing.

Equation	Estimate	Pr >  t	Variable
<b>Theft_qtrlog</b>	0.00565	0.4193	1
	-0.01583	0.2529	easter(t)
	3.35665	0.7958	Theft_qtrlog(t-1)
	-3.41238	0.7927	Theft_ratelog(t-1)
	1.50052	0.1098	avg_incomelog(t-1)
	0.01483	0.5618	unemployment_rate_all1(t-1)
	0.02648	0.2343	real_gdp_growth1(t-1)
	0.00361	0.7682	inflation1(t-1)
	1.15942	0.3173	Hld_debt_disp_income_ptlog(t-1)
	14.47234	0.4369	Theft_qtrlog(t-2)
	-14.76399	0.4285	Theft_ratelog(t-2)
	-1.04051	0.2895	avg_incomelog(t-2)
	0.01327	0.5812	unemployment_rate_all1(t-2)
	-0.01159	0.6836	real_gdp_growth1(t-2)
	0.01153	0.3201	inflation1(t-2)
	1.6308	0.2749	Hld_debt_disp_income_ptlog(t-2)
	-21.38851	0.2547	Theft_qtrlog(t-3)
	21.30476	0.2575	Theft_ratelog(t-3)
	0.79622	0.3636	avg_incomelog(t-3)
	-0.03244	0.1727	unemployment_rate_all1(t-3)

	-0.02558	0.349	real_gdp_growth1(t-3)
	-0.00651	0.5529	inflation1(t-3)
	-1.15312	0.4435	Hld_debt_disp_income_ptlog(t-3)
	10.94143	0.398	Theft_qtrlog(t-4)
	-11.29887	0.384	Theft_ratelog(t-4)
	0.74187	0.4193	avg_incomelog(t-4)
	0.01326	0.6168	unemployment_rate_all1(t-4)
	0.02653	0.2147	real_gdp_growth1(t-4)
	0.01005	0.3464	inflation1(t-4)
	0.76065	0.5102	Hld_debt_disp_income_ptlog(t-4)
<b>Theft_ratelog</b>	0.00569	0.4167	1
	-0.01588	0.2519	easter(t)
	2.39318	0.8537	Theft_qtrlog(t-1)
	-2.44454	0.8508	Theft_ratelog(t-1)
	1.50294	0.1096	avg_incomelog(t-1)
	0.01442	0.573	unemployment_rate_all1(t-1)
	0.02639	0.2363	real_gdp_growth1(t-1)
	0.00361	0.7687	inflation1(t-1)
	1.1596	0.3177	Hld_debt_disp_income_ptlog(t-1)
	14.6088	0.4331	Theft_qtrlog(t-2)
	-14.89983	0.4248	Theft_ratelog(t-2)
	-1.06558	0.2787	avg_incomelog(t-2)
	0.01372	0.5689	unemployment_rate_all1(t-2)
	-0.01148	0.6867	real_gdp_growth1(t-2)
	0.01161	0.3173	inflation1(t-2)
	1.60711	0.2823	Hld_debt_disp_income_ptlog(t-2)
	-21.25793	0.2581	Theft_qtrlog(t-3)
	21.17457	0.2609	Theft_ratelog(t-3)
	0.79308	0.3659	avg_incomelog(t-3)
	-0.03208	0.1778	unemployment_rate_all1(t-3)
	-0.02571	0.347	real_gdp_growth1(t-3)
	-0.00671	0.5414	inflation1(t-3)
	-1.15597	0.4428	Hld_debt_disp_income_ptlog(t-3)
	10.8898	0.4007	Theft_qtrlog(t-4)
	-11.2445	0.3867	Theft_ratelog(t-4)
	0.73492	0.4241	avg_incomelog(t-4)
	0.01341	0.6131	unemployment_rate_all1(t-4)
	0.02688	0.2092	real_gdp_growth1(t-4)
	0.01017	0.3412	inflation1(t-4)
	0.76047	0.5107	Hld_debt_disp_income_ptlog(t-4)
<b>avg_incomelog</b>	-0.00044	0.6676	1
	0.00299	0.1412	easter(t)
	0.22519	0.9055	Theft_qtrlog(t-1)
	-0.22763	0.9046	Theft_ratelog(t-1)
	-0.20811	0.1289	avg_incomelog(t-1)
	-0.00121	0.7467	unemployment_rate_all1(t-1)



	-0.0061	0.0639	real_gdp_growth1(t-1)
	-0.00196	0.2761	inflation1(t-1)
	-0.04869	0.7729	Hld_debt_disp_income_ptlog(t-1)
	-0.86172	0.751	Theft_qtrlog(t-2)
	0.89645	0.7417	Theft_ratelog(t-2)
	0.20728	0.1513	avg_incomelog(t-2)
	-0.00183	0.6034	unemployment_rate_all1(t-2)
	0.00757	0.0735	real_gdp_growth1(t-2)
	0.00127	0.4528	inflation1(t-2)
	-0.17965	0.4096	Hld_debt_disp_income_ptlog(t-2)
	1.8453	0.4998	Theft_qtrlog(t-3)
	-1.80751	0.5094	Theft_ratelog(t-3)
	-0.03647	0.7751	avg_incomelog(t-3)
	-0.00515	0.1393	unemployment_rate_all1(t-3)
	0.00134	0.7356	real_gdp_growth1(t-3)
	0.00328	0.0454	inflation1(t-3)
	0.20484	0.3528	Hld_debt_disp_income_ptlog(t-3)
	-1.30102	0.4915	Theft_qtrlog(t-4)
	1.2738	0.5015	Theft_ratelog(t-4)
	-0.66003	0.0001	avg_incomelog(t-4)
	0.00035	0.9271	unemployment_rate_all1(t-4)
	-0.00306	0.327	real_gdp_growth1(t-4)
	0.00023	0.8844	inflation1(t-4)
	-0.00669	0.9683	Hld_debt_disp_income_ptlog(t-4)
<b>unemployment_rate_all1</b>	-0.0607	0.1666	1
	0.10182	0.2382	easter(t)
	57.79618	0.4756	Theft_qtrlog(t-1)
	-58.11286	0.4737	Theft_ratelog(t-1)
	-9.01038	0.123	avg_incomelog(t-1)
	-0.04809	0.7625	unemployment_rate_all1(t-1)
	-0.2249	0.1074	real_gdp_growth1(t-1)
	-0.09848	0.201	inflation1(t-1)
	-6.89141	0.3399	Hld_debt_disp_income_ptlog(t-1)
	11.44383	0.9212	Theft_qtrlog(t-2)
	-10.57317	0.9273	Theft_ratelog(t-2)
	4.10155	0.5013	avg_incomelog(t-2)
	0.35709	0.0206	unemployment_rate_all1(t-2)
	0.03696	0.8347	real_gdp_growth1(t-2)
	-0.04601	0.5229	inflation1(t-2)
	2.86769	0.7566	Hld_debt_disp_income_ptlog(t-2)
	-68.43649	0.5566	Theft_qtrlog(t-3)
	70.27526	0.5469	Theft_ratelog(t-3)
	-7.44565	0.1752	avg_incomelog(t-3)
	0.18523	0.2108	unemployment_rate_all1(t-3)
	0.16022	0.3466	real_gdp_growth1(t-3)
	0.09568	0.1655	inflation1(t-3)

	-5.06219	0.5887	Hld_debt_disp_income_ptlog(t-3)
	4.05377	0.9598	Theft_qtrlog(t-4)
	-4.37901	0.9566	Theft_ratelog(t-4)
	-8.58799	0.137	avg_incomelog(t-4)
	-0.07205	0.6625	unemployment_rate_all1(t-4)
	-0.18428	0.1675	real_gdp_growth1(t-4)
	0.08525	0.202	inflation1(t-4)
	-5.66941	0.4314	Hld_debt_disp_income_ptlog(t-4)
<b>real_gdp_growth1</b>	-0.01117	0.8159	1
	-0.03111	0.7426	easter(t)
	-163.19665	0.0726	Theft_qtrlog(t-1)
	162.40869	0.0743	Theft_ratelog(t-1)
	4.96225	0.4375	avg_incomelog(t-1)
	-0.13242	0.4525	unemployment_rate_all1(t-1)
	0.93598	0.0001	real_gdp_growth1(t-1)
	0.02695	0.7494	inflation1(t-1)
	-7.84333	0.3258	Hld_debt_disp_income_ptlog(t-1)
	194.72605	0.1324	Theft_qtrlog(t-2)
	-195.62814	0.1312	Theft_ratelog(t-2)
	9.24103	0.1737	avg_incomelog(t-2)
	-0.27208	0.105	unemployment_rate_all1(t-2)
	-0.09986	0.6103	real_gdp_growth1(t-2)
	-0.09618	0.2296	inflation1(t-2)
	11.69744	0.2557	Hld_debt_disp_income_ptlog(t-2)
	-3.12592	0.9806	Theft_qtrlog(t-3)
	4.14856	0.9743	Theft_ratelog(t-3)
	14.26421	0.0212	avg_incomelog(t-3)
	0.27406	0.0961	unemployment_rate_all1(t-3)
	-0.20402	0.2789	real_gdp_growth1(t-3)
	-0.01303	0.8628	inflation1(t-3)
	1.15464	0.911	Hld_debt_disp_income_ptlog(t-3)
	-31.1618	0.7258	Theft_qtrlog(t-4)
	29.77925	0.7381	Theft_ratelog(t-4)
	2.02268	0.7484	avg_incomelog(t-4)
	0.18727	0.3069	unemployment_rate_all1(t-4)
	-0.17203	0.2422	real_gdp_growth1(t-4)
	0.02691	0.7132	inflation1(t-4)
	-0.17961	0.9819	Hld_debt_disp_income_ptlog(t-4)
<b>inflation1</b>	-0.0167	0.8045	1
	-0.07366	0.5806	easter(t)
	239.51312	0.0614	Theft_qtrlog(t-1)
	-235.78632	0.0656	Theft_ratelog(t-1)
	9.12395	0.3111	avg_incomelog(t-1)
	0.40794	0.1038	unemployment_rate_all1(t-1)
	0.13051	0.5426	real_gdp_growth1(t-1)
	0.14724	0.2182	inflation1(t-1)

	-16.87122	0.1355	Hld_debt_disp_income_ptlog(t-1)
	-358.18445	0.051	Theft_qtrlog(t-2)
	359.04501	0.0508	Theft_ratelog(t-2)
	-1.50707	0.8733	avg_incomelog(t-2)
	0.56681	0.0182	unemployment_rate_all1(t-2)
	0.2427	0.3797	real_gdp_growth1(t-2)
	-0.0248	0.8242	inflation1(t-2)
	0.72191	0.9599	Hld_debt_disp_income_ptlog(t-2)
	67.43435	0.709	Theft_qtrlog(t-3)
	-68.15584	0.7066	Theft_ratelog(t-3)
	-22.82364	0.0093	avg_incomelog(t-3)
	-0.9696	0.0001	unemployment_rate_all1(t-3)
	-0.41849	0.1168	real_gdp_growth1(t-3)
	0.00119	0.9911	inflation1(t-3)
	-8.86857	0.5422	Hld_debt_disp_income_ptlog(t-3)
	73.14454	0.5587	Theft_qtrlog(t-4)
	-77.19059	0.5382	Theft_ratelog(t-4)
	-20.79191	0.0227	avg_incomelog(t-4)
	-0.21544	0.4022	unemployment_rate_all1(t-4)
	0.35189	0.0916	real_gdp_growth1(t-4)
	-0.47215	0.0001	inflation1(t-4)
	1.902	0.8646	Hld_debt_disp_income_ptlog(t-4)
<b>Hld_debt_disp_income_ptlog</b>	-0.0012	0.164	1
	0.00114	0.4958	easter(t)
	0.25644	0.8711	Theft_qtrlog(t-1)
	-0.23821	0.8803	Theft_ratelog(t-1)
	-0.05733	0.6117	avg_incomelog(t-1)
	0.00617	0.0523	unemployment_rate_all1(t-1)
	0.0019	0.482	real_gdp_growth1(t-1)
	-0.00099	0.51	inflation1(t-1)
	0.6519	0.0001	Hld_debt_disp_income_ptlog(t-1)
	1.92974	0.3955	Theft_qtrlog(t-2)
	-1.94915	0.3915	Theft_ratelog(t-2)
	-0.09871	0.4087	avg_incomelog(t-2)
	0.0038	0.1982	unemployment_rate_all1(t-2)
	-0.00129	0.7088	real_gdp_growth1(t-2)
	-0.00024	0.865	inflation1(t-2)
	-0.07706	0.6704	Hld_debt_disp_income_ptlog(t-2)
	-0.78611	0.7296	Theft_qtrlog(t-3)
	0.80606	0.7236	Theft_ratelog(t-3)
	0.02349	0.8252	avg_incomelog(t-3)
	-0.00455	0.1178	unemployment_rate_all1(t-3)
	0.00253	0.4473	real_gdp_growth1(t-3)
	0.00062	0.6404	inflation1(t-3)
	0.06425	0.7255	Hld_debt_disp_income_ptlog(t-3)
	-1.22405	0.4377	Theft_qtrlog(t-4)

	1.22939	0.4368	Theft_ratelog(t-4)
	0.00184	0.9868	avg_incomelog(t-4)
	-0.00481	0.1403	unemployment_rate_all1(t-4)
	0.00037	0.8866	real_gdp_growth1(t-4)
	0.00005	0.9714	inflation1(t-4)
	-0.4081	0.0054	Hld_debt_disp_income_ptlog(t-4)

## Reduced Model

**Appendix Table 10: Parameter estimate for reduced theft and the state of the economy model, VAR (4) (Model 4.5)**

Notes: This table presents s parameter estimate for VAR (4) using seasonal differencing.

Equation	Estimate	Pr >  t	Variable
Theft_qtrlog	0.00315	0.6306	1
	-0.01276	0.3132	easter(t)
	0.00195	0.7135	Prin1(t)
	0.00142	0.7855	Prin2(t)
	6.29615	0.5771	Theft_qtrlog(t-1)
	-6.49302	0.5655	Theft_ratelog(t-1)
	1.21788	0.0871	avg_incomelog(t-1)
	1.21705	0.2146	Hld_debt_disp_income_ptlog(t-1)
	1.64487	0.9199	Theft_qtrlog(t-2)
	-1.84217	0.9104	Theft_ratelog(t-2)
	-0.11199	0.8777	avg_incomelog(t-2)
	1.37788	0.29	Hld_debt_disp_income_ptlog(t-2)
	-9.57865	0.5587	Theft_qtrlog(t-3)
	9.43066	0.5656	Theft_ratelog(t-3)
	0.59262	0.4042	avg_incomelog(t-3)
	-0.79299	0.5299	Hld_debt_disp_income_ptlog(t-3)
	7.17825	0.5279	Theft_qtrlog(t-4)
	-7.54115	0.5078	Theft_ratelog(t-4)
	1.47391	0.0509	avg_incomelog(t-4)
	0.81355	0.4151	Hld_debt_disp_income_ptlog(t-4)
Theft_ratelog	0.00316	0.6302	1
	-0.01277	0.3134	easter(t)
	0.00196	0.7131	Prin1(t)
	0.00143	0.7839	Prin2(t)
	5.28904	0.6398	Theft_qtrlog(t-1)
	-5.48255	0.6277	Theft_ratelog(t-1)
	1.21974	0.087	avg_incomelog(t-1)
	1.22151	0.2135	Hld_debt_disp_income_ptlog(t-1)
	1.74249	0.9152	Theft_qtrlog(t-2)
	-1.93853	0.9059	Theft_ratelog(t-2)
	-0.12211	0.8669	avg_incomelog(t-2)
	1.36241	0.296	Hld_debt_disp_income_ptlog(t-2)

	-9.42086	0.5657	Theft_qtrlog(t-3)
	9.27237	0.5726	Theft_ratelog(t-3)
	0.58106	0.4139	avg_incomelog(t-3)
	-0.80232	0.5256	Hld_debt_disp_income_ptlog(t-3)
	7.17682	0.5285	Theft_qtrlog(t-4)
	-7.53937	0.5085	Theft_ratelog(t-4)
	1.46284	0.0529	avg_incomelog(t-4)
	0.81668	0.4139	Hld_debt_disp_income_ptlog(t-4)
avg_incomelog	-0.00075	0.4817	1
	0.00339	0.1024	easter(t)
	0.00127	0.1472	Prin1(t)
	0.00099	0.2454	Prin2(t)
	-0.37073	0.84	Theft_qtrlog(t-1)
	0.37646	0.8376	Theft_ratelog(t-1)
	-0.12268	0.2858	avg_incomelog(t-1)
	-0.05369	0.735	Hld_debt_disp_income_ptlog(t-1)
	-0.70738	0.7905	Theft_qtrlog(t-2)
	0.74014	0.7814	Theft_ratelog(t-2)
	0.23184	0.0543	avg_incomelog(t-2)
	-0.13726	0.516	Hld_debt_disp_income_ptlog(t-2)
	1.58294	0.5528	Theft_qtrlog(t-3)
	-1.54013	0.5644	Theft_ratelog(t-3)
	-0.01136	0.9215	avg_incomelog(t-3)
	0.08721	0.671	Hld_debt_disp_income_ptlog(t-3)
	-0.83843	0.6504	Theft_qtrlog(t-4)
	0.80208	0.6649	Theft_ratelog(t-4)
	-0.68529	0.0001	avg_incomelog(t-4)
	-0.01434	0.9295	Hld_debt_disp_income_ptlog(t-4)
Hld_debt_disp_income_ptlog	-0.00126	0.129	1
	0.00143	0.3651	easter(t)
	-0.0001	0.8777	Prin1(t)
	0.00096	0.145	Prin2(t)
	0.94782	0.5021	Theft_qtrlog(t-1)
	-0.92607	0.5121	Theft_ratelog(t-1)
	0.03408	0.698	avg_incomelog(t-1)
	0.78301	0.0001	Hld_debt_disp_income_ptlog(t-1)
	0.94314	0.6447	Theft_qtrlog(t-2)
	-0.96557	0.6374	Theft_ratelog(t-2)
	-0.08752	0.3377	avg_incomelog(t-2)
	-0.20789	0.2025	Hld_debt_disp_income_ptlog(t-2)
	0.18599	0.9275	Theft_qtrlog(t-3)
	-0.16279	0.9367	Theft_ratelog(t-3)
	0.00315	0.9716	avg_incomelog(t-3)
	0.07299	0.6433	Hld_debt_disp_income_ptlog(t-3)
	-1.82403	0.2019	Theft_qtrlog(t-4)
	1.83321	0.2002	Theft_ratelog(t-4)

	-0.01734	0.8517	avg_incomelog(t-4)
	-0.44662	0.0007	Hld_debt_disp_income_ptlog(t-4)

## Burglary and the state of the economy model

### Full Model1

Endogenous variable: burglary\_qtrlog (1), burglary\_qtrlog (1), avg\_incomelog(1),  
Hld\_debt\_disp\_income\_pt log(1,4), real\_gdp\_growth (1), inflation (1), unemployment\_rate\_all (1)

Exogenous variable: Easter, qtr1, qtr2, qtr3

**Appendix Table 11: Parameter estimate for full burglary and the state of the economy model, VAR (4)**

**Note:** This table presents parameter estimate for VAR (4) using seasonal dummies

Equation	Estimate	Pr >  t	Variable
burglary_qtrlog	-0.085	0.086	1
	0.08997	0.098	S_1t
	0.07145	0.3316	S_2t
	0.06391	0.2858	S_3t
	-0.00512	0.8438	easter(t)
	-6.75165	0.633	burglary_qtrlog(t-1)
	6.56423	0.6422	burglary_ratelog(t-1)
	1.37662	0.2759	avg_incomelog(t-1)
	0.04333	0.1061	unemployment_rate_all1(t-1)
	0.02683	0.267	real_gdp_growth1(t-1)
	0.00099	0.936	inflation1(t-1)
	-0.77447	0.4928	Hld_debt_disp_income_ptlog(t-1)
	-2.15961	0.9055	burglary_qtrlog(t-2)
	1.99036	0.9129	burglary_ratelog(t-2)
	1.01447	0.3937	avg_incomelog(t-2)
	0.02981	0.2531	unemployment_rate_all1(t-2)
	-0.01044	0.7225	real_gdp_growth1(t-2)
	0.01755	0.1534	inflation1(t-2)
	0.82244	0.5748	Hld_debt_disp_income_ptlog(t-2)
	37.43683	0.0496	burglary_qtrlog(t-3)
	-37.5354	0.0491	burglary_ratelog(t-3)
	-0.25204	0.8358	avg_incomelog(t-3)
	-0.02737	0.261	unemployment_rate_all1(t-3)
	-0.00323	0.9149	real_gdp_growth1(t-3)
	0.00236	0.8277	inflation1(t-3)
	0.40056	0.7852	Hld_debt_disp_income_ptlog(t-3)
	-22.63036	0.0912	burglary_qtrlog(t-4)
	22.61053	0.0915	burglary_ratelog(t-4)
	-0.34845	0.7342	avg_incomelog(t-4)

	-0.03292	0.25	unemployment_rate_all1(t-4)
	0.01134	0.6262	real_gdp_growth1(t-4)
	0.00296	0.7751	inflation1(t-4)
	0.80695	0.4767	Hld_debt_disp_income_ptlog(t-4)
burglary_ratelog	-0.08434	0.0889	1
	0.08808	0.1057	S_1t
	0.07046	0.339	S_2t
	0.0639	0.2866	S_3t
	-0.00499	0.8481	easter(t)
	-7.62552	0.5904	burglary_qtrlog(t-1)
	7.43988	0.5992	burglary_ratelog(t-1)
	1.38625	0.2734	avg_incomelog(t-1)
	0.04331	0.1068	unemployment_rate_all1(t-1)
	0.02697	0.2654	real_gdp_growth1(t-1)
	0.00094	0.9391	inflation1(t-1)
	-0.77786	0.4917	Hld_debt_disp_income_ptlog(t-1)
	-2.18771	0.9044	burglary_qtrlog(t-2)
	2.02097	0.9117	burglary_ratelog(t-2)
	1.00258	0.4	avg_incomelog(t-2)
	0.03027	0.2467	unemployment_rate_all1(t-2)
	-0.01051	0.7212	real_gdp_growth1(t-2)
	0.01761	0.1527	inflation1(t-2)
	0.81348	0.5796	Hld_debt_disp_income_ptlog(t-2)
	37.67937	0.0485	burglary_qtrlog(t-3)
	-37.77646	0.0481	burglary_ratelog(t-3)
	-0.26548	0.8274	avg_incomelog(t-3)
	-0.02699	0.2683	unemployment_rate_all1(t-3)
	-0.0033	0.9132	real_gdp_growth1(t-3)
	0.00219	0.8402	inflation1(t-3)
	0.40059	0.7855	Hld_debt_disp_income_ptlog(t-3)
	-22.83487	0.0889	burglary_qtrlog(t-4)
	22.81765	0.0892	burglary_ratelog(t-4)
	-0.37026	0.7187	avg_incomelog(t-4)
	-0.03295	0.2503	unemployment_rate_all1(t-4)
	0.01164	0.6175	real_gdp_growth1(t-4)
	0.00304	0.7694	inflation1(t-4)
	0.8016	0.4803	Hld_debt_disp_income_ptlog(t-4)
avg_incomelog	-0.00546	0.3918	1
	0.00467	0.505	S_1t
	0.00695	0.468	S_2t
	0.02015	0.0124	S_3t
	0.00678	0.0507	easter(t)
	1.20652	0.5136	burglary_qtrlog(t-1)
	-1.22651	0.5064	burglary_ratelog(t-1)
	-0.12671	0.4405	avg_incomelog(t-1)
	0.00078	0.8202	unemployment_rate_all1(t-1)

	-0.00988	0.0028	real_gdp_growth1(t-1)
	-0.00084	0.6015	inflation1(t-1)
	0.04962	0.7357	Hld_debt_disp_income_ptlog(t-1)
	-5.35989	0.0283	burglary_qtrlog(t-2)
	5.376	0.0278	burglary_ratelog(t-2)
	0.38136	0.0171	avg_incomelog(t-2)
	-0.00425	0.2121	unemployment_rate_all1(t-2)
	0.01325	0.0012	real_gdp_growth1(t-2)
	0.00164	0.3048	inflation1(t-2)
	0.00316	0.9868	Hld_debt_disp_income_ptlog(t-2)
	2.4005	0.326	burglary_qtrlog(t-3)
	-2.39279	0.3278	burglary_ratelog(t-3)
	0.10593	0.5053	avg_incomelog(t-3)
	-0.00349	0.2716	unemployment_rate_all1(t-3)
	-0.0023	0.5595	real_gdp_growth1(t-3)
	0.0024	0.0946	inflation1(t-3)
	-0.00508	0.9788	Hld_debt_disp_income_ptlog(t-3)
	1.49711	0.3857	burglary_qtrlog(t-4)
	-1.52988	0.3755	burglary_ratelog(t-4)
	0.19968	0.1407	avg_incomelog(t-4)
	0.00344	0.3554	unemployment_rate_all1(t-4)
	-0.00378	0.2168	real_gdp_growth1(t-4)
	-0.00103	0.445	inflation1(t-4)
	0.18615	0.2112	Hld_debt_disp_income_ptlog(t-4)
unemployment_rate_all1	-0.26582	0.4093	1
	0.11515	0.7443	S_1t
	-0.03677	0.9393	S_2t
	0.30582	0.4373	S_3t
	0.2011	0.2446	easter(t)
	-14.01923	0.8804	burglary_qtrlog(t-1)
	14.84495	0.8733	burglary_ratelog(t-1)
	8.43921	0.3108	avg_incomelog(t-1)
	-0.201	0.2521	unemployment_rate_all1(t-1)
	-0.41419	0.0118	real_gdp_growth1(t-1)
	-0.12363	0.1342	inflation1(t-1)
	-2.83912	0.7026	Hld_debt_disp_income_ptlog(t-1)
	-155.08082	0.2007	burglary_qtrlog(t-2)
	155.32578	0.2	burglary_ratelog(t-2)
	11.3294	0.1521	avg_incomelog(t-2)
	0.3255	0.0619	unemployment_rate_all1(t-2)
	0.35351	0.0737	real_gdp_growth1(t-2)
	0.00344	0.9658	inflation1(t-2)
	3.28313	0.734	Hld_debt_disp_income_ptlog(t-2)
	145.26218	0.241	burglary_qtrlog(t-3)
	-145.57842	0.2402	burglary_ratelog(t-3)
	-0.4551	0.9547	avg_incomelog(t-3)



	0.19696	0.2208	unemployment_rate_all1(t-3)
	0.00894	0.9642	real_gdp_growth1(t-3)
	0.01997	0.7801	inflation1(t-3)
	-0.36473	0.97	Hld_debt_disp_income_ptlog(t-3)
	5.12926	0.9529	burglary_qtrlog(t-4)
	-6.85951	0.9371	burglary_ratelog(t-4)
	-3.40836	0.6149	avg_incomelog(t-4)
	-0.03085	0.8692	unemployment_rate_all1(t-4)
	-0.31347	0.0459	real_gdp_growth1(t-4)
	0.03836	0.5746	inflation1(t-4)
	-5.81681	0.437	Hld_debt_disp_income_ptlog(t-4)
real_gdp_growth1	0.48113	0.1678	1
	-0.55297	0.1499	S_1t
	-0.76024	0.1479	S_2t
	-0.58079	0.1731	S_3t
	-0.23219	0.2119	easter(t)
	-46.64467	0.6417	burglary_qtrlog(t-1)
	46.57435	0.6419	burglary_ratelog(t-1)
	3.34846	0.7068	avg_incomelog(t-1)
	-0.31219	0.1008	unemployment_rate_all1(t-1)
	0.89994	0.0001	real_gdp_growth1(t-1)
	-0.03881	0.6578	inflation1(t-1)
	-5.98223	0.4552	Hld_debt_disp_income_ptlog(t-1)
	69.43601	0.591	burglary_qtrlog(t-2)
	-70.16702	0.5871	burglary_ratelog(t-2)
	-3.02749	0.7186	avg_incomelog(t-2)
	-0.23666	0.2015	unemployment_rate_all1(t-2)
	-0.15668	0.4536	real_gdp_growth1(t-2)
	0.00356	0.967	inflation1(t-2)
	11.07277	0.2891	Hld_debt_disp_income_ptlog(t-2)
	-92.31467	0.4858	burglary_qtrlog(t-3)
	89.29413	0.5003	burglary_ratelog(t-3)
	5.60509	0.5165	avg_incomelog(t-3)
	0.37293	0.0342	unemployment_rate_all1(t-3)
	-0.15147	0.4807	real_gdp_growth1(t-3)
	0.02283	0.7665	inflation1(t-3)
	-9.35444	0.3714	Hld_debt_disp_income_ptlog(t-3)
	52.12704	0.5773	burglary_qtrlog(t-4)
	-52.86881	0.572	burglary_ratelog(t-4)
	4.19356	0.5648	avg_incomelog(t-4)
	0.10774	0.5931	unemployment_rate_all1(t-4)
	-0.17761	0.2844	real_gdp_growth1(t-4)
	-0.00768	0.9166	inflation1(t-4)
	7.17846	0.3727	Hld_debt_disp_income_ptlog(t-4)
inflation1	0.38982	0.4965	1
	-0.10266	0.8703	S_1t

	-0.15143	0.8603	S_2t
	0.06432	0.9267	S_3t
	-0.18523	0.5451	easter(t)
	58.6271	0.7241	burglary_qtrlog(t-1)
	-57.13809	0.7306	burglary_ratelog(t-1)
	-11.61687	0.4325	avg_incomelog(t-1)
	0.35976	0.25	unemployment_rate_all1(t-1)
	-0.02809	0.9207	real_gdp_growth1(t-1)
	0.16497	0.2591	inflation1(t-1)
	-29.51372	0.0302	Hld_debt_disp_income_ptlog(t-1)
	-92.42449	0.6659	burglary_qtrlog(t-2)
	92.80785	0.6646	burglary_ratelog(t-2)
	-7.3842	0.5964	avg_incomelog(t-2)
	0.8642	0.0066	unemployment_rate_all1(t-2)
	0.38578	0.2676	real_gdp_growth1(t-2)
	-0.07211	0.614	inflation1(t-2)
	12.01583	0.4861	Hld_debt_disp_income_ptlog(t-2)
	65.89172	0.7635	burglary_qtrlog(t-3)
	-64.49323	0.7685	burglary_ratelog(t-3)
	-14.43487	0.3152	avg_incomelog(t-3)
	-0.89895	0.0027	unemployment_rate_all1(t-3)
	-0.30235	0.3965	real_gdp_growth1(t-3)
	-0.0343	0.7878	inflation1(t-3)
	2.16448	0.9003	Hld_debt_disp_income_ptlog(t-3)
	-43.13647	0.7806	burglary_qtrlog(t-4)
	43.99625	0.7763	burglary_ratelog(t-4)
	-5.16775	0.6684	avg_incomelog(t-4)
	-0.389	0.2475	unemployment_rate_all1(t-4)
	0.33671	0.2219	real_gdp_growth1(t-4)
	-0.42112	0.0012	inflation1(t-4)
	-9.87742	0.4587	Hld_debt_disp_income_ptlog(t-4)
Hld_debt_disp_income_ptlog	-0.00485	0.4019	1
	0.00557	0.3813	S_1t
	0.0071	0.4142	S_2t
	0.00355	0.6148	S_3t
	0.00636	0.0437	easter(t)
	0.27028	0.8716	burglary_qtrlog(t-1)
	-0.248	0.882	burglary_ratelog(t-1)
	-0.06538	0.6603	avg_incomelog(t-1)
	0.00761	0.0186	unemployment_rate_all1(t-1)
	0.00111	0.6966	real_gdp_growth1(t-1)
	-0.00206	0.1629	inflation1(t-1)
	0.60469	0.0001	Hld_debt_disp_income_ptlog(t-1)
	-0.99984	0.643	burglary_qtrlog(t-2)
	1.0283	0.6336	burglary_ratelog(t-2)
	0.09845	0.484	avg_incomelog(t-2)

	0.0026	0.3991	unemployment_rate_all1(t-2)
	0.00027	0.9376	real_gdp_growth1(t-2)
	-0.00014	0.9208	inflation1(t-2)
	0.02193	0.8993	Hld_debt_disp_income_ptlog(t-2)
	1.45926	0.5094	burglary_qtrlog(t-3)
	-1.44017	0.5151	burglary_ratelog(t-3)
	-0.10822	0.4537	avg_incomelog(t-3)
	-0.00433	0.1355	unemployment_rate_all1(t-3)
	0.00223	0.5342	real_gdp_growth1(t-3)
	0.00075	0.5573	inflation1(t-3)
	0.03304	0.8494	Hld_debt_disp_income_ptlog(t-3)
	-1.22288	0.4347	burglary_qtrlog(t-4)
	1.22409	0.4342	burglary_ratelog(t-4)
	0.02017	0.8681	avg_incomelog(t-4)
	-0.00725	0.0359	unemployment_rate_all1(t-4)
	-0.00114	0.6799	real_gdp_growth1(t-4)
	-0.00123	0.3194	inflation1(t-4)
	-0.46319	0.0012	Hld_debt_disp_income_ptlog(t-4)

## Full Model2

Endogenous variable: burglary\_qtrlog (4), burglary\_qtrlog (4), avg\_incomelog (1, 4),  
Hld\_debt\_disp\_income\_pt log(1,4), real\_gdp\_growth (1), inflation (1), unemployment\_rate\_all (1)

Exogenous variable: Easter

### Appendix Table 12: Parameter estimate for full burglary and the state of the economy model, VAR (4)

Note: This table presents parameter estimate for VAR (4) using seasonal differencing

Equation	Estimate	Pr >  t	Variable
burglary_qtrlog	-0.04514	0.1131	1
	-0.00588	0.7221	easter(t)
	13.08398	0.4461	burglary_qtrlog(t-1)
	-12.4325	0.4691	burglary_ratelog(t-1)
	-0.09641	0.9297	avg_incomelog(t-1)
	0.06066	0.0552	unemployment_rate_all1(t-1)
	0.04985	0.08	real_gdp_growth1(t-1)
	-0.00183	0.9062	inflation1(t-1)
	-0.69437	0.6151	Hld_debt_disp_income_ptlog(t-1)
	-19.9607	0.5839	burglary_qtrlog(t-2)
	19.8625	0.5861	burglary_ratelog(t-2)
	0.03575	0.9732	avg_incomelog(t-2)
	0.05634	0.0635	unemployment_rate_all1(t-2)
	-0.05145	0.1443	real_gdp_growth1(t-2)
	0.02414	0.0953	inflation1(t-2)
	1.15799	0.506	Hld_debt_disp_income_ptlog(t-2)

	6.5085	0.862	burglary_qtrlog(t-3)
	-6.49417	0.8625	burglary_ratelog(t-3)
	0.16716	0.8698	avg_incomelog(t-3)
	-0.02377	0.4024	unemployment_rate_all1(t-3)
	0.02455	0.4898	real_gdp_growth1(t-3)
	0.00364	0.7801	inflation1(t-3)
	-0.98812	0.5717	Hld_debt_disp_income_ptlog(t-3)
	5.08696	0.7741	burglary_qtrlog(t-4)
	-5.33458	0.7638	burglary_ratelog(t-4)
	0.0163	0.9878	avg_incomelog(t-4)
	-0.04402	0.1711	unemployment_rate_all1(t-4)
	0.01517	0.5941	real_gdp_growth1(t-4)
	0.00071	0.9566	inflation1(t-4)
	2.21792	0.1174	Hld_debt_disp_income_ptlog(t-4)
burglary_ratelog	-0.04579	0.1082	1
	-0.0058	0.7256	easter(t)
	11.3526	0.5082	burglary_qtrlog(t-1)
	-10.6975	0.533	burglary_ratelog(t-1)
	-0.09512	0.9306	avg_incomelog(t-1)
	0.06038	0.0562	unemployment_rate_all1(t-1)
	0.05008	0.0787	real_gdp_growth1(t-1)
	-0.00189	0.903	inflation1(t-1)
	-0.69176	0.6164	Hld_debt_disp_income_ptlog(t-1)
	-19.0368	0.6013	burglary_qtrlog(t-2)
	18.93681	0.6036	burglary_ratelog(t-2)
	0.01913	0.9857	avg_incomelog(t-2)
	0.05657	0.0625	unemployment_rate_all1(t-2)
	-0.05174	0.1421	real_gdp_growth1(t-2)
	0.02418	0.0948	inflation1(t-2)
	1.14655	0.5102	Hld_debt_disp_income_ptlog(t-2)
	6.23767	0.8677	burglary_qtrlog(t-3)
	-6.22284	0.8682	burglary_ratelog(t-3)
	0.15809	0.8768	avg_incomelog(t-3)
	-0.02333	0.4111	unemployment_rate_all1(t-3)
	0.02465	0.488	real_gdp_growth1(t-3)
	0.00353	0.7869	inflation1(t-3)
	-0.99351	0.5696	Hld_debt_disp_income_ptlog(t-3)
	5.22601	0.7681	burglary_qtrlog(t-4)
	-5.47343	0.7578	burglary_ratelog(t-4)
	0.00108	0.9992	avg_incomelog(t-4)
	-0.04381	0.1731	unemployment_rate_all1(t-4)
	0.01554	0.5853	real_gdp_growth1(t-4)
	0.00082	0.9495	inflation1(t-4)
	2.21219	0.1183	Hld_debt_disp_income_ptlog(t-4)
avg_incomelog	0.00516	0.1016	1
	0.00165	0.3671	easter(t)

	-1.9655	0.3017	burglary_qtrlog(t-1)
	1.93442	0.3095	burglary_ratelog(t-1)
	-0.29557	0.0178	avg_incomelog(t-1)
	-0.00299	0.3854	unemployment_rate_all1(t-1)
	-0.01081	0.001	real_gdp_growth1(t-1)
	-0.00095	0.5803	inflation1(t-1)
	-0.08217	0.5904	Hld_debt_disp_income_ptlog(t-1)
	1.53436	0.703	burglary_qtrlog(t-2)
	-1.51708	0.7065	burglary_ratelog(t-2)
	0.09291	0.4311	avg_incomelog(t-2)
	0.00105	0.7501	unemployment_rate_all1(t-2)
	0.0103	0.01	real_gdp_growth1(t-2)
	0.00099	0.53	inflation1(t-2)
	-0.20054	0.2991	Hld_debt_disp_income_ptlog(t-2)
	1.94209	0.6393	burglary_qtrlog(t-3)
	-1.92643	0.6423	burglary_ratelog(t-3)
	0.01013	0.9284	avg_incomelog(t-3)
	-0.00709	0.0274	unemployment_rate_all1(t-3)
	0.00195	0.6197	real_gdp_growth1(t-3)
	0.00276	0.0608	inflation1(t-3)
	0.26253	0.1776	Hld_debt_disp_income_ptlog(t-3)
	-2.01789	0.3056	burglary_qtrlog(t-4)
	1.97437	0.3168	burglary_ratelog(t-4)
	-0.55998	0.0001	avg_incomelog(t-4)
	-0.00095	0.7876	unemployment_rate_all1(t-4)
	-0.00573	0.0731	real_gdp_growth1(t-4)
	0.00051	0.7248	inflation1(t-4)
	0.11226	0.4686	Hld_debt_disp_income_ptlog(t-4)
unemployment_rate_all1	-0.09812	0.4904	1
	0.06864	0.4125	easter(t)
	55.97134	0.5183	burglary_qtrlog(t-1)
	-56.8354	0.5121	burglary_ratelog(t-1)
	-7.35032	0.1872	avg_incomelog(t-1)
	-0.04694	0.7643	unemployment_rate_all1(t-1)
	-0.30139	0.0375	real_gdp_growth1(t-1)
	-0.10211	0.1961	inflation1(t-1)
	-9.17219	0.1921	Hld_debt_disp_income_ptlog(t-1)
	-44.3632	0.8093	burglary_qtrlog(t-2)
	45.71454	0.8038	burglary_ratelog(t-2)
	2.88772	0.5916	avg_incomelog(t-2)
	0.42352	0.0069	unemployment_rate_all1(t-2)
	0.13974	0.4284	real_gdp_growth1(t-2)
	-0.05056	0.4835	inflation1(t-2)
	6.0348	0.4926	Hld_debt_disp_income_ptlog(t-2)
	-76.0268	0.6879	burglary_qtrlog(t-3)
	75.82086	0.689	burglary_ratelog(t-3)

	-9.71103	0.0643	avg_incomelog(t-3)
	0.15592	0.2782	unemployment_rate_all1(t-3)
	0.07672	0.6687	real_gdp_growth1(t-3)
	0.09869	0.1388	inflation1(t-3)
	-4.8167	0.5851	Hld_debt_disp_income_ptlog(t-3)
	67.75121	0.4504	burglary_qtrlog(t-4)
	-68.7513	0.4444	burglary_ratelog(t-4)
	-6.96612	0.2015	avg_incomelog(t-4)
	-0.11816	0.4636	unemployment_rate_all1(t-4)
	-0.17718	0.221	real_gdp_growth1(t-4)
	0.06602	0.3177	inflation1(t-4)
	-3.04547	0.6663	Hld_debt_disp_income_ptlog(t-4)
real_gdp_growth1	0.17873	0.227	1
	-0.0533	0.5377	easter(t)
	-226.276	0.0144	burglary_qtrlog(t-1)
	225.8211	0.0146	burglary_ratelog(t-1)
	3.83996	0.5021	avg_incomelog(t-1)
	-0.16661	0.3061	unemployment_rate_all1(t-1)
	0.79985	0.0001	real_gdp_growth1(t-1)
	-0.00533	0.9475	inflation1(t-1)
	-9.04048	0.2134	Hld_debt_disp_income_ptlog(t-1)
	334.8269	0.0832	burglary_qtrlog(t-2)
	-335.641	0.0828	burglary_ratelog(t-2)
	4.93815	0.3763	avg_incomelog(t-2)
	-0.24979	0.1133	unemployment_rate_all1(t-2)
	0.00424	0.9814	real_gdp_growth1(t-2)
	-0.00214	0.9771	inflation1(t-2)
	12.80568	0.1626	Hld_debt_disp_income_ptlog(t-2)
	-32.4001	0.8684	burglary_qtrlog(t-3)
	31.30917	0.873	burglary_ratelog(t-3)
	11.31928	0.038	avg_incomelog(t-3)
	0.33165	0.0289	unemployment_rate_all1(t-3)
	-0.21628	0.2464	real_gdp_growth1(t-3)
	-0.0217	0.7502	inflation1(t-3)
	-4.25911	0.6405	Hld_debt_disp_income_ptlog(t-3)
	-93.8788	0.3131	burglary_qtrlog(t-4)
	95.05905	0.3078	burglary_ratelog(t-4)
	11.38671	0.0463	avg_incomelog(t-4)
	0.0799	0.6312	unemployment_rate_all1(t-4)
	-0.24396	0.1054	real_gdp_growth1(t-4)
	-0.01855	0.785	inflation1(t-4)
	5.67315	0.4384	Hld_debt_disp_income_ptlog(t-4)
inflation1	-0.00423	0.9862	1
	-0.04833	0.737	easter(t)
	284.0136	0.0616	burglary_qtrlog(t-1)
	-281.811	0.0637	burglary_ratelog(t-1)

	3.13593	0.7418	avg_incomelog(t-1)
	0.2369	0.3818	unemployment_rate_all1(t-1)
	-0.0073	0.9761	real_gdp_growth1(t-1)
	0.06835	0.6128	inflation1(t-1)
	-21.2144	0.0823	Hld_debt_disp_income_ptlog(t-1)
	-646.356	0.046	burglary_qtrlog(t-2)
	645.8821	0.0463	burglary_ratelog(t-2)
	-4.9329	0.5949	avg_incomelog(t-2)
	0.52061	0.0494	unemployment_rate_all1(t-2)
	0.34828	0.2539	real_gdp_growth1(t-2)
	-0.03116	0.8017	inflation1(t-2)
	10.80118	0.4762	Hld_debt_disp_income_ptlog(t-2)
	457.9943	0.1645	burglary_qtrlog(t-3)
	-459.416	0.1637	burglary_ratelog(t-3)
	-24.3655	0.0083	avg_incomelog(t-3)
	-0.87892	0.0008	unemployment_rate_all1(t-3)
	-0.43259	0.1656	real_gdp_growth1(t-3)
	-0.03848	0.7347	inflation1(t-3)
	-9.34647	0.5389	Hld_debt_disp_income_ptlog(t-3)
	-97.1514	0.5296	burglary_qtrlog(t-4)
	97.73627	0.5278	burglary_ratelog(t-4)
	-13.8275	0.1424	avg_incomelog(t-4)
	-0.20115	0.4691	unemployment_rate_all1(t-4)
	0.43094	0.0867	real_gdp_growth1(t-4)
	-0.46278	0.0002	inflation1(t-4)
	-7.45479	0.5407	Hld_debt_disp_income_ptlog(t-4)
Hld_debt_disp_income_ptlog	-0.00159	0.5614	1
	0.00197	0.2255	easter(t)
	1.47713	0.3775	burglary_qtrlog(t-1)
	-1.46637	0.3811	burglary_ratelog(t-1)
	-0.03542	0.7393	avg_incomelog(t-1)
	0.00696	0.0249	unemployment_rate_all1(t-1)
	0.00251	0.3586	real_gdp_growth1(t-1)
	-0.0021	0.169	inflation1(t-1)
	0.56976	0.0001	Hld_debt_disp_income_ptlog(t-1)
	0.00698	0.9984	burglary_qtrlog(t-2)
	0.00548	0.9988	burglary_ratelog(t-2)
	0.00938	0.9279	avg_incomelog(t-2)
	0.0035	0.2306	unemployment_rate_all1(t-2)
	-0.00225	0.5083	real_gdp_growth1(t-2)
	-0.00073	0.5994	inflation1(t-2)
	0.05248	0.7564	Hld_debt_disp_income_ptlog(t-2)
	-3.84169	0.2949	burglary_qtrlog(t-3)
	3.84387	0.2952	burglary_ratelog(t-3)
	-0.07813	0.4327	avg_incomelog(t-3)
	-0.00379	0.1735	unemployment_rate_all1(t-3)

	0.00359	0.3016	real_gdp_growth1(t-3)
	0.00123	0.3355	inflation1(t-3)
	-0.01653	0.9224	Hld_debt_disp_income_ptlog(t-3)
	2.37772	0.1725	burglary_qtrlog(t-4)
	-2.39416	0.1703	burglary_ratelog(t-4)
	-0.03684	0.7238	avg_incomelog(t-4)
	-0.00635	0.0452	unemployment_rate_all1(t-4)
	-0.00002	0.9938	real_gdp_growth1(t-4)
	-0.00102	0.4214	inflation1(t-4)
	-0.42347	0.0031	Hld_debt_disp_income_ptlog(t-4)

## Reduced Model

Dependent Variables: avg\_incomelog(1), real\_gdp\_growth1(1), Hld\_debt\_disp\_income\_ptlog(1,4), burglary\_qtrlog(1), burglary\_ratelog(1)

Independent Variables: Easter, qtr1, qtr2, qtr3 prin1

**Appendix Table 13: Parameter estimate for reduced burglary and the state of the economy model, VAR (4)**

**Note:** This table presents parameter estimate for VAR (4) using seasonal dummies

Equation	Estimate	Pr >  t	Variable
burglary_qtrlog	-0.09377	0.0156	1
	0.10692	0.0088	S_1t
	0.0794	0.1917	S_2t
	0.06103	0.2049	S_3t
	-0.00299	0.6548	Prin1(t)
	-4.18381	0.7401	burglary_qtrlog(t-1)
	4.01278	0.7502	burglary_ratelog(t-1)
	1.57328	0.1603	avg_incomelog(t-1)
	0.01232	0.5389	real_gdp_growth1(t-1)
	-0.03849	0.9679	Hld_debt_disp_income_ptlog(t-1)
	-6.0848	0.7155	burglary_qtrlog(t-2)
	5.94506	0.7217	burglary_ratelog(t-2)
	1.65892	0.1311	avg_incomelog(t-2)
	-0.01498	0.5739	real_gdp_growth1(t-2)
	0.0725	0.9522	Hld_debt_disp_income_ptlog(t-2)
	37.39843	0.0291	burglary_qtrlog(t-3)
	-37.48578	0.0288	burglary_ratelog(t-3)
	-0.56628	0.5998	avg_incomelog(t-3)
	0.00544	0.8296	real_gdp_growth1(t-3)
	0.2138	0.8662	Hld_debt_disp_income_ptlog(t-3)
	-21.77124	0.0811	burglary_qtrlog(t-4)
	21.63772	0.083	burglary_ratelog(t-4)
	-0.68323	0.4788	avg_incomelog(t-4)



	0.00431	0.8109	real_gdp_growth1(t-4)
	0.81326	0.4344	Hld_debt_disp_income_ptlog(t-4)
burglary_ratelog	-0.0931	0.0165	1
	0.10479	0.0102	S_1t
	0.07825	0.1987	S_2t
	0.06094	0.2062	S_3t
	-0.00293	0.6626	Prin1(t)
	-5.0832	0.6874	burglary_qtrlog(t-1)
	4.91394	0.6971	burglary_ratelog(t-1)
	1.58313	0.1584	avg_incomelog(t-1)
	0.01237	0.5377	real_gdp_growth1(t-1)
	-0.035	0.9708	Hld_debt_disp_income_ptlog(t-1)
	-6.15081	0.713	burglary_qtrlog(t-2)
	6.01404	0.719	burglary_ratelog(t-2)
	1.66109	0.1311	avg_incomelog(t-2)
	-0.01499	0.5743	real_gdp_growth1(t-2)
	0.06469	0.9574	Hld_debt_disp_income_ptlog(t-2)
	37.66059	0.0283	burglary_qtrlog(t-3)
	-37.74599	0.028	burglary_ratelog(t-3)
	-0.57289	0.596	avg_incomelog(t-3)
	0.00548	0.8285	real_gdp_growth1(t-3)
	0.21126	0.868	Hld_debt_disp_income_ptlog(t-3)
	-21.95766	0.079	burglary_qtrlog(t-4)
	21.82619	0.0808	burglary_ratelog(t-4)
	-0.70032	0.4685	avg_incomelog(t-4)
	0.00436	0.809	real_gdp_growth1(t-4)
	0.81089	0.4364	Hld_debt_disp_income_ptlog(t-4)
avg_incomelog	0.00766	0.1086	1
	-0.00545	0.2729	S_1t
	-0.00203	0.7885	S_2t
	0.01324	0.0308	S_3t
	-0.00339	0.0002	Prin1(t)
	1.18334	0.4543	burglary_qtrlog(t-1)
	-1.22618	0.4379	burglary_ratelog(t-1)
	-0.45575	0.0018	avg_incomelog(t-1)
	-0.00587	0.0222	real_gdp_growth1(t-1)
	0.006	0.9599	Hld_debt_disp_income_ptlog(t-1)
	-4.05856	0.0564	burglary_qtrlog(t-2)
	4.04669	0.057	burglary_ratelog(t-2)
	0.05472	0.687	avg_incomelog(t-2)
	0.00668	0.0487	real_gdp_growth1(t-2)
	-0.03981	0.7922	Hld_debt_disp_income_ptlog(t-2)
	0.55774	0.79	burglary_qtrlog(t-3)
	-0.55214	0.7922	burglary_ratelog(t-3)
	0.08653	0.5215	avg_incomelog(t-3)
	0.00102	0.7474	real_gdp_growth1(t-3)

	0.02216	0.8888	Hld_debt_disp_income_ptlog(t-3)
	2.13943	0.1682	burglary_qtrlog(t-4)
	-2.16297	0.1637	burglary_ratelog(t-4)
	0.12908	0.2857	avg_incomelog(t-4)
	-0.00305	0.1792	real_gdp_growth1(t-4)
	0.25326	0.055	Hld_debt_disp_income_ptlog(t-4)
real_gdp_growth1	-0.08379	0.6173	1
	0		S_1t
	0		S_2t
	0		S_3t
	0.0782	0.1072	Prin1(t)
	-99.65175	0.1405	burglary_qtrlog(t-1)
	100.00266	0.14	burglary_ratelog(t-1)
	12.89753	0.0605	avg_incomelog(t-1)
	0.94031	0.0001	real_gdp_growth1(t-1)
	-16.21173	0.0209	Hld_debt_disp_income_ptlog(t-1)
	116.82298	0.1471	burglary_qtrlog(t-2)
	-116.53471	0.1492	burglary_ratelog(t-2)
	-1.32059	0.8578	avg_incomelog(t-2)
	-0.03044	0.8729	real_gdp_growth1(t-2)
	23.67702	0.0087	Hld_debt_disp_income_ptlog(t-2)
	-5.1026	0.9508	burglary_qtrlog(t-3)
	2.69172	0.9741	burglary_ratelog(t-3)
	3.27878	0.63	avg_incomelog(t-3)
	-0.40801	0.0232	real_gdp_growth1(t-3)
	-12.06026	0.1895	Hld_debt_disp_income_ptlog(t-3)
	-33.41525	0.6131	burglary_qtrlog(t-4)
	33.11708	0.6167	burglary_ratelog(t-4)
	2.6557	0.6785	avg_incomelog(t-4)
	0.01411	0.9117	real_gdp_growth1(t-4)
	1.47259	0.839	Hld_debt_disp_income_ptlog(t-4)
Hld_debt_disp_income_ptlog	-0.00172	0.5851	1
	0		S_1t
	0		S_2t
	0		S_3t
	0.00108	0.2362	Prin1(t)
	1.11208	0.3789	burglary_qtrlog(t-1)
	-1.08694	0.3907	burglary_ratelog(t-1)
	-0.03324	0.7937	avg_incomelog(t-1)
	-0.00089	0.7408	real_gdp_growth1(t-1)
	0.70938	0.0001	Hld_debt_disp_income_ptlog(t-1)
	-0.89545	0.5514	burglary_qtrlog(t-2)
	0.90606	0.5478	burglary_ratelog(t-2)
	0.17789	0.2033	avg_incomelog(t-2)
	0.00039	0.9122	real_gdp_growth1(t-2)
	-0.17359	0.2922	Hld_debt_disp_income_ptlog(t-2)

	0.42869	0.7829	burglary_qtrlog(t-3)
	-0.39938	0.798	burglary_ratelog(t-3)
	0.07417	0.5628	avg_incomelog(t-3)
	0.0039	0.24	real_gdp_growth1(t-3)
	0.13632	0.4282	Hld_debt_disp_income_ptlog(t-3)
	-0.98457	0.4295	burglary_qtrlog(t-4)
	0.97913	0.4326	burglary_ratelog(t-4)
	0.0707	0.558	avg_incomelog(t-4)
	-0.00274	0.2563	real_gdp_growth1(t-4)
	-0.46267	0.0013	Hld_debt_disp_income_ptlog(t-4)

## Fraud and the state of the economy model

### Full Model

Endogenous variable: fraud\_ratelog (1,4), fraud\_qtrlog (1,4), avg\_incomelog(1,4),  
Hld\_debt\_disp\_income\_pt log(1,4), real\_gdp\_growth (1), inflation (1), unemployment\_rate\_all (1)

Exogenous variable: Easter

#### Appendix Table 14: Parameter estimate for full fraud and the state of the economy model, VAR (5)

Note: This table presents parameter estimate for VAR (5) using seasonal differencing

Equation	Estimate	Pr >  t	Variable
<b>Fraud_qtrlog</b>	-0.00739	0.5489	1
	0.01147	0.6261	easter(t)
	-12.5406	0.6014	Fraud_qtrlog(t-1)
	12.30754	0.608	fraud_ratelog(t-1)
	1.51384	0.4411	avg_incomelog(t-1)
	-0.0237	0.6004	unemployment_rate_all1(t-1)
	-0.01248	0.7439	real_gdp_growth1(t-1)
	-0.0256	0.3836	inflation1(t-1)
	0.84033	0.7112	Hld_debt_disp_income_ptlog(t-1)
	38.82446	0.2431	Fraud_qtrlog(t-2)
	-39.0358	0.2408	fraud_ratelog(t-2)
	1.09926	0.5065	avg_incomelog(t-2)
	0.00867	0.8555	unemployment_rate_all1(t-2)
	0.05418	0.3103	real_gdp_growth1(t-2)
	0.03041	0.1642	inflation1(t-2)
	-1.77329	0.4912	Hld_debt_disp_income_ptlog(t-2)
	-21.6309	0.5096	Fraud_qtrlog(t-3)
	21.64764	0.5096	fraud_ratelog(t-3)
	0.02507	0.9875	avg_incomelog(t-3)
	0.00288	0.953	unemployment_rate_all1(t-3)
	-0.03521	0.5205	real_gdp_growth1(t-3)

	0.01172	0.593	inflation1(t-3)
	4.6499	0.0835	Hld_debt_disp_income_ptlog(t-3)
	5.048	0.8766	Fraud_qtrlog(t-4)
	-5.60598	0.8633	fraud_ratelog(t-4)
	-1.26804	0.4034	avg_incomelog(t-4)
	0.04765	0.3136	unemployment_rate_all1(t-4)
	-0.0317	0.5472	real_gdp_growth1(t-4)
	0.04315	0.0292	inflation1(t-4)
	-5.23022	0.0494	Hld_debt_disp_income_ptlog(t-4)
	-5.10901	0.822	Fraud_qtrlog(t-5)
	4.91081	0.8289	fraud_ratelog(t-5)
	2.97863	0.1243	avg_incomelog(t-5)
	0.02289	0.6184	unemployment_rate_all1(t-5)
	0.02447	0.551	real_gdp_growth1(t-5)
	0.01145	0.6731	inflation1(t-5)
	0.83549	0.7177	Hld_debt_disp_income_ptlog(t-5)
<b>fraud_ratelog</b>	-0.00733	0.5519	1
	0.01144	0.6269	easter(t)
	-13.5356	0.573	Fraud_qtrlog(t-1)
	13.30375	0.5795	fraud_ratelog(t-1)
	1.50613	0.4434	avg_incomelog(t-1)
	-0.0239	0.5973	unemployment_rate_all1(t-1)
	-0.01249	0.7437	real_gdp_growth1(t-1)
	-0.02545	0.3865	inflation1(t-1)
	0.84619	0.7093	Hld_debt_disp_income_ptlog(t-1)
	38.98821	0.2412	Fraud_qtrlog(t-2)
	-39.1992	0.2389	fraud_ratelog(t-2)
	1.08195	0.5132	avg_incomelog(t-2)
	0.00904	0.8494	unemployment_rate_all1(t-2)
	0.05418	0.3104	real_gdp_growth1(t-2)
	0.03042	0.1641	inflation1(t-2)
	-1.78104	0.4893	Hld_debt_disp_income_ptlog(t-2)
	-21.5317	0.5116	Fraud_qtrlog(t-3)
	21.54807	0.5116	fraud_ratelog(t-3)
	0.02144	0.9893	avg_incomelog(t-3)
	0.00304	0.9505	unemployment_rate_all1(t-3)
	-0.03564	0.5155	real_gdp_growth1(t-3)
	0.01163	0.596	inflation1(t-3)
	4.64691	0.0838	Hld_debt_disp_income_ptlog(t-3)
	4.80028	0.8827	Fraud_qtrlog(t-4)
	-5.35642	0.8693	fraud_ratelog(t-4)
	-1.27482	0.401	avg_incomelog(t-4)
	0.04781	0.3121	unemployment_rate_all1(t-4)
	-0.03071	0.5598	real_gdp_growth1(t-4)

	0.04336	0.0284	inflation1(t-4)
	-5.24223	0.049	Hld_debt_disp_income_ptlog(t-4)
	-4.88932	0.8296	Fraud_qtrlog(t-5)
	4.69172	0.8364	fraud_ratelog(t-5)
	2.9958	0.1223	avg_incomelog(t-5)
	0.02275	0.6206	unemployment_rate_all1(t-5)
	0.024	0.5588	real_gdp_growth1(t-5)
	0.01126	0.6784	inflation1(t-5)
	0.85843	0.7103	Hld_debt_disp_income_ptlog(t-5)
<b>avg_incomelog</b>	-0.00001	0.9883	1
	0.00108	0.5743	easter(t)
	-1.06479	0.5868	Fraud_qtrlog(t-1)
	1.0764	0.5827	fraud_ratelog(t-1)
	-0.30055	0.066	avg_incomelog(t-1)
	-0.00551	0.1404	unemployment_rate_all1(t-1)
	-0.00882	0.0071	real_gdp_growth1(t-1)
	-0.00007	0.9758	inflation1(t-1)
	-0.03861	0.8348	Hld_debt_disp_income_ptlog(t-1)
	-1.73019	0.5211	Fraud_qtrlog(t-2)
	1.75601	0.5152	fraud_ratelog(t-2)
	0.05117	0.7042	avg_incomelog(t-2)
	-0.00536	0.1728	unemployment_rate_all1(t-2)
	0.00804	0.0693	real_gdp_growth1(t-2)
	0.00286	0.1109	inflation1(t-2)
	-0.25402	0.23	Hld_debt_disp_income_ptlog(t-2)
	3.09463	0.2506	Fraud_qtrlog(t-3)
	-3.08904	0.2517	fraud_ratelog(t-3)
	0.03372	0.7962	avg_incomelog(t-3)
	-0.00578	0.1529	unemployment_rate_all1(t-3)
	0.00146	0.7441	real_gdp_growth1(t-3)
	0.00291	0.1091	inflation1(t-3)
	0.17979	0.4048	Hld_debt_disp_income_ptlog(t-3)
	2.62051	0.3266	Fraud_qtrlog(t-4)
	-2.62964	0.3253	fraud_ratelog(t-4)
	-0.5027	0.0002	avg_incomelog(t-4)
	0.00298	0.4383	unemployment_rate_all1(t-4)
	-0.00106	0.8043	real_gdp_growth1(t-4)
	0.00039	0.8022	inflation1(t-4)
	0.28804	0.1787	Hld_debt_disp_income_ptlog(t-4)
	-3.68114	0.0529	Fraud_qtrlog(t-5)
	3.66354	0.0541	fraud_ratelog(t-5)
	0.00545	0.9721	avg_incomelog(t-5)
	0.00204	0.5863	unemployment_rate_all1(t-5)
	-0.00375	0.2657	real_gdp_growth1(t-5)

	0.00367	0.1032	inflation1(t-5)
	-0.03863	0.8376	Hld_debt_disp_income_ptlog(t-5)
<b>unemployment_rate_all1</b>	-0.04471	0.3147	1
	0.07134	0.4001	easter(t)
	28.45333	0.7411	Fraud_qtrlog(t-1)
	-28.7264	0.7386	fraud_ratelog(t-1)
	-4.91586	0.4857	avg_incomelog(t-1)
	0.06638	0.6826	unemployment_rate_all1(t-1)
	-0.28739	0.0417	real_gdp_growth1(t-1)
	0.10166	0.3359	inflation1(t-1)
	-10.7223	0.193	Hld_debt_disp_income_ptlog(t-1)
	-37.0826	0.7542	Fraud_qtrlog(t-2)
	36.79635	0.7561	fraud_ratelog(t-2)
	0.78837	0.8942	avg_incomelog(t-2)
	0.32034	0.0669	unemployment_rate_all1(t-2)
	0.11259	0.5553	real_gdp_growth1(t-2)
	-0.01769	0.8195	inflation1(t-2)
	10.65272	0.2522	Hld_debt_disp_income_ptlog(t-2)
	52.75565	0.6538	Fraud_qtrlog(t-3)
	-53.0067	0.6525	fraud_ratelog(t-3)
	-9.55434	0.1025	avg_incomelog(t-3)
	-0.00885	0.9598	unemployment_rate_all1(t-3)
	0.1264	0.5207	real_gdp_growth1(t-3)
	0.09189	0.2468	inflation1(t-3)
	-8.90479	0.3491	Hld_debt_disp_income_ptlog(t-3)
	-36.8743	0.7523	Fraud_qtrlog(t-4)
	36.68385	0.7538	fraud_ratelog(t-4)
	-1.21889	0.8223	avg_incomelog(t-4)
	0.0502	0.7661	unemployment_rate_all1(t-4)
	-0.15636	0.4094	real_gdp_growth1(t-4)
	0.04965	0.4717	inflation1(t-4)
	3.34679	0.7195	Hld_debt_disp_income_ptlog(t-4)
	13.08867	0.8725	Fraud_qtrlog(t-5)
	-12.3532	0.8797	fraud_ratelog(t-5)
	3.19762	0.641	avg_incomelog(t-5)
	0.13682	0.4085	unemployment_rate_all1(t-5)
	-0.07656	0.6033	real_gdp_growth1(t-5)
	0.27415	0.0074	inflation1(t-5)
	-4.74137	0.5683	Hld_debt_disp_income_ptlog(t-5)
<b>real_gdp_growth1</b>	-0.00181	0.9728	1
	-0.05176	0.6119	easter(t)
	-152.637	0.1475	Fraud_qtrlog(t-1)
	152.52	0.1477	fraud_ratelog(t-1)
	6.84359	0.4219	avg_incomelog(t-1)

	-0.18993	0.335	unemployment_rate_all1(t-1)
	0.95018	0.0001	real_gdp_growth1(t-1)
	-0.10412	0.4134	inflation1(t-1)
	-8.35532	0.3976	Hld_debt_disp_income_ptlog(t-1)
	196.525	0.1743	Fraud_qtrlog(t-2)
	-196.183	0.1753	fraud_ratelog(t-2)
	9.85366	0.1737	avg_incomelog(t-2)
	-0.2089	0.3144	unemployment_rate_all1(t-2)
	-0.0003	0.9989	real_gdp_growth1(t-2)
	-0.05096	0.5866	inflation1(t-2)
	9.44716	0.3983	Hld_debt_disp_income_ptlog(t-2)
	37.46416	0.7917	Fraud_qtrlog(t-3)
	-37.444	0.792	fraud_ratelog(t-3)
	8.47216	0.2264	avg_incomelog(t-3)
	0.36704	0.0896	unemployment_rate_all1(t-3)
	-0.20558	0.3879	real_gdp_growth1(t-3)
	-0.0474	0.6181	inflation1(t-3)
	-3.53069	0.7572	Hld_debt_disp_income_ptlog(t-3)
	-53.7348	0.7034	Fraud_qtrlog(t-4)
	53.8826	0.7028	fraud_ratelog(t-4)
	1.95356	0.7656	avg_incomelog(t-4)
	0.10442	0.6087	unemployment_rate_all1(t-4)
	-0.42006	0.0713	real_gdp_growth1(t-4)
	-0.01584	0.8486	inflation1(t-4)
	2.44876	0.8276	Hld_debt_disp_income_ptlog(t-4)
	-23.7112	0.8097	Fraud_qtrlog(t-5)
	22.82987	0.8168	fraud_ratelog(t-5)
	-3.6178	0.6621	avg_incomelog(t-5)
	0.07446	0.7084	unemployment_rate_all1(t-5)
	0.30579	0.0912	real_gdp_growth1(t-5)
	-0.0453	0.7003	inflation1(t-5)
	-2.61159	0.7942	Hld_debt_disp_income_ptlog(t-5)
<b>inflation1</b>	0.01876	0.7932	1
	-0.04317	0.7523	easter(t)
	298.0084	0.038	Fraud_qtrlog(t-1)
	-296.286	0.039	fraud_ratelog(t-1)
	-9.58533	0.4021	avg_incomelog(t-1)
	0.09817	0.7089	unemployment_rate_all1(t-1)
	-0.00253	0.9909	real_gdp_growth1(t-1)
	0.0091	0.9573	inflation1(t-1)
	-15.2435	0.2521	Hld_debt_disp_income_ptlog(t-1)
	-416.648	0.035	Fraud_qtrlog(t-2)
	417.8615	0.0346	fraud_ratelog(t-2)
	3.64208	0.7046	avg_incomelog(t-2)

	0.4796	0.0893	unemployment_rate_all1(t-2)
	0.37332	0.2308	real_gdp_growth1(t-2)
	0.04017	0.7491	inflation1(t-2)
	-1.40514	0.9251	Hld_debt_disp_income_ptlog(t-2)
	132.8212	0.4866	Fraud_qtrlog(t-3)
	-134.191	0.4825	fraud_ratelog(t-3)
	-11.4178	0.2245	avg_incomelog(t-3)
	-0.64033	0.0294	unemployment_rate_all1(t-3)
	-0.23706	0.4575	real_gdp_growth1(t-3)
	-0.00503	0.9685	inflation1(t-3)
	-4.06473	0.7908	Hld_debt_disp_income_ptlog(t-3)
	236.1794	0.2167	Fraud_qtrlog(t-4)
	-239.061	0.2116	fraud_ratelog(t-4)
	-14.7018	0.1005	avg_incomelog(t-4)
	-0.24732	0.368	unemployment_rate_all1(t-4)
	0.11484	0.7073	real_gdp_growth1(t-4)
	-0.56088	0.0001	inflation1(t-4)
	9.30822	0.5382	Hld_debt_disp_income_ptlog(t-4)
	-238.346	0.0774	Fraud_qtrlog(t-5)
	237.2747	0.0788	fraud_ratelog(t-5)
	-12.8225	0.252	avg_incomelog(t-5)
	-0.01036	0.969	unemployment_rate_all1(t-5)
	0.18843	0.4309	real_gdp_growth1(t-5)
	-0.10329	0.5138	inflation1(t-5)
	0.38743	0.977	Hld_debt_disp_income_ptlog(t-5)
<b>Hld_debt_disp_income_ptlog</b>	-0.00105	0.2599	1
	0.00217	0.2234	easter(t)
	1.30695	0.4694	Fraud_qtrlog(t-1)
	-1.31394	0.4669	fraud_ratelog(t-1)
	-0.17695	0.2333	avg_incomelog(t-1)
	0.0071	0.0418	unemployment_rate_all1(t-1)
	0.00167	0.5619	real_gdp_growth1(t-1)
	-0.00166	0.4526	inflation1(t-1)
	0.70547	0.0002	Hld_debt_disp_income_ptlog(t-1)
	0.53093	0.8302	Fraud_qtrlog(t-2)
	-0.53148	0.8301	fraud_ratelog(t-2)
	-0.0633	0.6101	avg_incomelog(t-2)
	0.00282	0.4312	unemployment_rate_all1(t-2)
	-0.00227	0.5701	real_gdp_growth1(t-2)
	-0.00093	0.5652	inflation1(t-2)
	0.03363	0.8616	Hld_debt_disp_income_ptlog(t-2)
	-0.89926	0.7146	Fraud_qtrlog(t-3)
	0.90617	0.7127	fraud_ratelog(t-3)
	-0.07146	0.553	avg_incomelog(t-3)



	-0.00503	0.1761	unemployment_rate_all1(t-3)
	0.00428	0.3011	real_gdp_growth1(t-3)
	0.00217	0.1914	inflation1(t-3)
	-0.06909	0.727	Hld_debt_disp_income_ptlog(t-3)
	-0.54235	0.8243	Fraud_qtrlog(t-4)
	0.55478	0.8205	fraud_ratelog(t-4)
	0.03003	0.7913	avg_incomelog(t-4)
	-0.00561	0.1179	unemployment_rate_all1(t-4)
	0.00122	0.7575	real_gdp_growth1(t-4)
	0.0007	0.6253	inflation1(t-4)
	-0.43782	0.0295	Hld_debt_disp_income_ptlog(t-4)
	-0.3445	0.84	Fraud_qtrlog(t-5)
	0.33932	0.8424	fraud_ratelog(t-5)
	-0.04063	0.7768	avg_incomelog(t-5)
	-0.00032	0.9249	unemployment_rate_all1(t-5)
	-0.00125	0.6846	real_gdp_growth1(t-5)
	-0.00054	0.7893	inflation1(t-5)
	0.21789	0.2138	Hld_debt_disp_income_ptlog(t-5)

## Reduced model

Endogenous variable: fraud\_qtrlog (1), fraud\_ratelog (1), avg\_incomelog(1), inflation(1),  
Hld\_debt\_disp\_income\_pt log(1,4)

Exogenous variable: unemployment\_rate\_all

**Appendix Table 15: Parameter estimate for reduced fraud and the state of the economy model, VAR (4) (Model 4.3)**

Note: This table presents parameter estimate for VAR (4) using seasonal dummies

Equation	Estimate	Pr >  t	Variable
Fraud_qtrlog	0.0255	0.6234	1
	0.09039	0.1022	S_1t
	-0.03821	0.6835	S_2t
	-0.11704	0.111	S_3t
	0		unemployment_rate_all(t)
	11.42058	0.4992	Fraud_qtrlog(t-1)
	-11.725	0.4878	fraud_ratelog(t-1)
	-1.47835	0.2555	avg_incomelog(t-1)
	-0.02467	0.0491	inflation1(t-1)
	0.53753	0.6747	Hld_debt_disp_income_ptlog(t-1)
	21.98245	0.335	Fraud_qtrlog(t-2)
	-22.2259	0.3302	fraud_ratelog(t-2)
	-2.11817	0.1106	avg_incomelog(t-2)
	0.0158	0.2159	inflation1(t-2)
	-1.3126	0.4301	Hld_debt_disp_income_ptlog(t-2)

	-27.6113	0.2072	Fraud_qtrlog(t-3)
	27.7326	0.2055	fraud_ratelog(t-3)
	0.74206	0.5554	avg_incomelog(t-3)
	0.0072	0.5836	inflation1(t-3)
	2.17782	0.188	Hld_debt_disp_income_ptlog(t-3)
	-0.91153	0.9565	Fraud_qtrlog(t-4)
	0.96515	0.9539	fraud_ratelog(t-4)
	-1.77395	0.1322	avg_incomelog(t-4)
	0.03972	0.0038	inflation1(t-4)
	-2.88819	0.034	Hld_debt_disp_income_ptlog(t-4)
fraud_ratelog	0.02636	0.6118	1
	0.08759	0.113	S_1t
	-0.03982	0.671	S_2t
	-0.11687	0.1116	S_3t
	0		unemployment_rate_all(t)
	10.44669	0.5364	Fraud_qtrlog(t-1)
	-10.7504	0.5245	fraud_ratelog(t-1)
	-1.47116	0.2578	avg_incomelog(t-1)
	-0.02459	0.0499	inflation1(t-1)
	0.54631	0.6698	Hld_debt_disp_income_ptlog(t-1)
	21.88001	0.3373	Fraud_qtrlog(t-2)
	-22.1237	0.3324	fraud_ratelog(t-2)
	-2.116	0.111	avg_incomelog(t-2)
	0.01581	0.2154	inflation1(t-2)
	-1.32459	0.426	Hld_debt_disp_income_ptlog(t-2)
	-27.2616	0.213	Fraud_qtrlog(t-3)
	27.38124	0.2112	fraud_ratelog(t-3)
	0.73706	0.5581	avg_incomelog(t-3)
	0.00701	0.5935	inflation1(t-3)
	2.17797	0.188	Hld_debt_disp_income_ptlog(t-3)
	-1.07048	0.9489	Fraud_qtrlog(t-4)
	1.12431	0.9463	fraud_ratelog(t-4)
	-1.7931	0.1282	avg_incomelog(t-4)
	0.03971	0.0038	inflation1(t-4)
	-2.887	0.0341	Hld_debt_disp_income_ptlog(t-4)
avg_incomelog	0.02472	0.0039	1
	-0.00559	0.3309	S_1t
	-0.00824	0.4017	S_2t
	0.00924	0.2259	S_3t
	-0.00213	0.0076	unemployment_rate_all(t)
	1.70104	0.3304	Fraud_qtrlog(t-1)
	-1.67326	0.3381	fraud_ratelog(t-1)
	-0.30147	0.0362	avg_incomelog(t-1)
	-0.00002	0.9847	inflation1(t-1)
	0.05313	0.684	Hld_debt_disp_income_ptlog(t-1)
	-3.92649	0.0988	Fraud_qtrlog(t-2)

	3.94441	0.0976	fraud_ratelog(t-2)
	0.01534	0.9149	avg_incomelog(t-2)
	0.00214	0.1027	inflation1(t-2)
	-0.03457	0.838	Hld_debt_disp_income_ptlog(t-2)
	0.00866	0.9969	Fraud_qtrlog(t-3)
	0.00436	0.9985	fraud_ratelog(t-3)
	-0.03253	0.8149	avg_incomelog(t-3)
	0.00154	0.2535	inflation1(t-3)
	-0.18579	0.2717	Hld_debt_disp_income_ptlog(t-3)
	1.94973	0.2604	Fraud_qtrlog(t-4)
	-1.95748	0.2587	fraud_ratelog(t-4)
	0.04673	0.7169	avg_incomelog(t-4)
	-0.00154	0.2557	inflation1(t-4)
	0.26236	0.0583	Hld_debt_disp_income_ptlog(t-4)
inflation1	1.87715	0.008	1
	0		S_1t
	0	1	S_2t
	0	1	S_3t
	-0.21136	0.0072	unemployment_rate_all(t)
	88.81705	0.4384	Fraud_qtrlog(t-1)
	-87.2902	0.4466	fraud_ratelog(t-1)
	-18.2096	0.1233	avg_incomelog(t-1)
	0.03913	0.7339	inflation1(t-1)
	-4.36922	0.721	Hld_debt_disp_income_ptlog(t-1)
	-184.058	0.1992	Fraud_qtrlog(t-2)
	186.3248	0.1937	fraud_ratelog(t-2)
	-2.30272	0.8448	avg_incomelog(t-2)
	-0.05837	0.6174	inflation1(t-2)
	-7.81697	0.6218	Hld_debt_disp_income_ptlog(t-2)
	87.41846	0.5328	Fraud_qtrlog(t-3)
	-86.3365	0.5374	fraud_ratelog(t-3)
	-36.7082	0.0019	avg_incomelog(t-3)
	0.06107	0.6168	inflation1(t-3)
	-14.79	0.3476	Hld_debt_disp_income_ptlog(t-3)
	2.20355	0.9847	Fraud_qtrlog(t-4)
	-1.94359	0.9865	fraud_ratelog(t-4)
	-29.2168	0.0142	avg_incomelog(t-4)
	-0.52281	0.0001	inflation1(t-4)
	0.0559	0.9965	Hld_debt_disp_income_ptlog(t-4)
Hld_debt_disp_income_ptlog	-0.00101	0.7236	1
	0		S_1t
	0		S_2t
	0		S_3t
	0		unemployment_rate_all(t)
	1.41409	0.2435	Fraud_qtrlog(t-1)
	-1.41697	0.2428	fraud_ratelog(t-1)

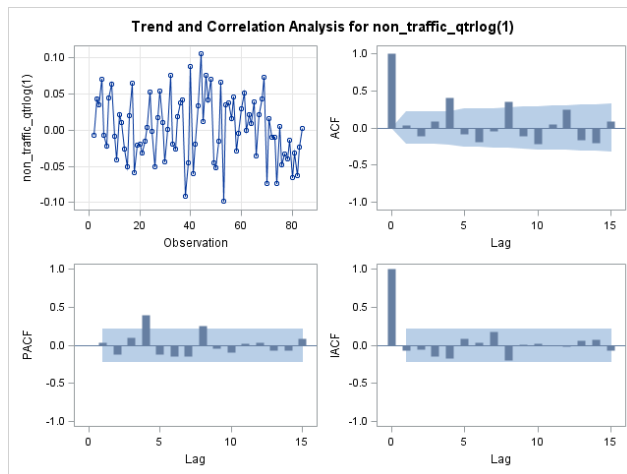
	-0.06967	0.5306	avg_incomelog(t-1)
	-0.00129	0.2828	inflation1(t-1)
	0.75213	0.0001	Hld_debt_disp_income_ptlog(t-1)
	-1.19684	0.425	Fraud_qtrlog(t-2)
	1.19506	0.4255	fraud_ratelog(t-2)
	0.06993	0.5471	avg_incomelog(t-2)
	-0.00144	0.2446	inflation1(t-2)
	-0.14949	0.3717	Hld_debt_disp_income_ptlog(t-2)
	0.74714	0.6127	Fraud_qtrlog(t-3)
	-0.75038	0.6107	fraud_ratelog(t-3)
	-0.04976	0.6472	avg_incomelog(t-3)
	0.00096	0.4566	inflation1(t-3)
	-0.022	0.8934	Hld_debt_disp_income_ptlog(t-3)
	-0.98389	0.4188	Fraud_qtrlog(t-4)
	0.98826	0.4158	fraud_ratelog(t-4)
	0.05677	0.6012	avg_incomelog(t-4)
	-0.00145	0.249	inflation1(t-4)
	-0.3936	0.0044	Hld_debt_disp_income_ptlog(t-4)

# Appendix D ARIMA model

## Seasonal Dummy Model

### Non-traffic

Appendix Figure 1: ACF and PACF for Non-traffic after first differencing



Appendix Table 16: Tentative Order Selection Test for Non-traffic after first differencing

ARMA(p+d,q) Tentative Order Selection Test					
SCAN			ESACF		
p+d	q	BIC	p+d	q	BIC
2	2	-6.22517	3	2	-6.18511
4	1	-6.2673	5	1	-6.2178
1	4	-6.15734	6	1	-6.20181
8	0	-6.28896	2	4	-6.22956
0	8	-6.15307	8	2	-6.19197
			0	8	-6.15307
			1	8	-6.11219

Appendix Table 17: Parameter estimate for ARIMA((4), 1,0) (Non-traffic)

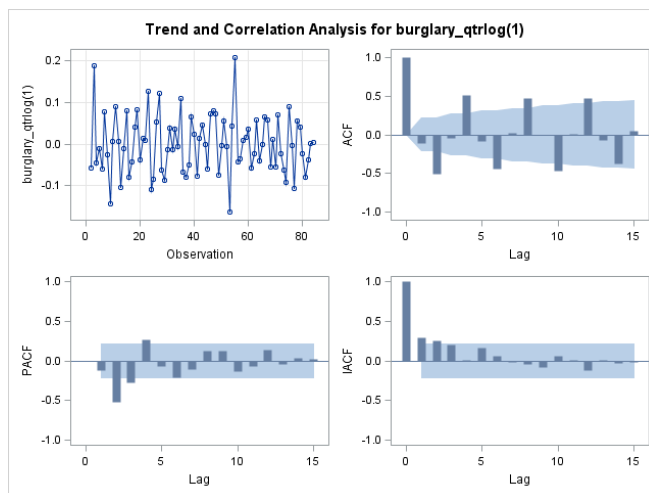
Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag
MU	0.01636	0.007073	2.31	0.0207	0
AR1,1	0.41682	0.10277	4.06	<.0001	4
NUM1	-0.05249	0.01044	-5.03	<.0001	0

Appendix Table 18: Parameter estimate for ARIMA(0,1,(4)) (Non-traffic)

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag
MU	0.01675	0.005953	2.81	0.0049	0
MA1,1	-0.33865	0.10811	-3.13	0.0017	4
NUM1	-0.05378	0.0102	-5.27	<.0001	0

## Burglary

Appendix Figure 2: ACF and PACF for burglary after first differencing



Appendix Table 19: Tentative Order Selection Test for burglary after first differencing

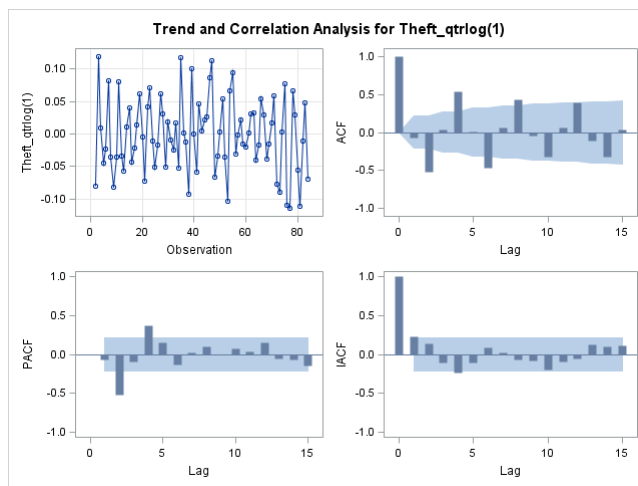
ARMA(p+d,q) Tentative Order Selection Test					
SCAN			ESACF		
p+d	q	BIC	p+d	q	BIC
3	2	-6.08818	2	3	-6.08812
2	3	-6.08812	3	3	-6.03748
8	1	-5.95826	7	2	-6.0161
12	0	-6.06671	8	2	-5.96975
			9	2	-6.01759
			11	3	-5.94272
			0	12	-5.54993
			1	12	-5.51384

Appendix Table 20: Parameter estimate for ARIMA(1,1,0) (Burglary)

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag	Variable	Shift
<b>MU</b>	-0.0602	0.01065	-5.65	<.0001	0	burglary_qtrlog	0
<b>AR1,1</b>	-0.18729	0.11143	-1.68	0.0928	1	burglary_qtrlog	0
<b>NUM1</b>	0.03667	0.01624	2.26	0.0239	0	qtr4	0
<b>NUM2</b>	0.14151	0.01437	9.85	<.0001	0	qtr3	0
<b>NUM3</b>	0.05832	0.01624	3.59	0.0003	0	qtr2	0

## Theft

Appendix Figure 3: ACF and PACF for theft after first differencing



Appendix Table 21: Tentative Order Selection Test for theft after first differencing

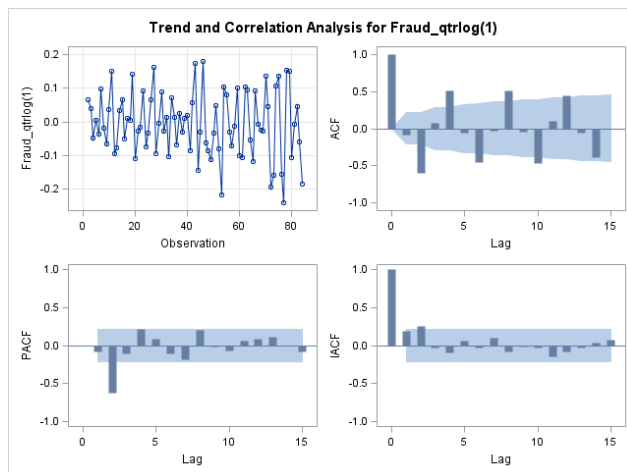
ARMA(p+d,q) Tentative Order Selection Test					
SCAN			ESACF		
p+d	q	BIC	p+d	q	BIC
2	2	-6.24199	2	2	-6.24199
4	0	-6.20361	3	2	-6.20102
			6	3	-6.05226
			7	2	-6.05407
			8	0	-6.11823
			0	8	-5.77255
			1	8	-5.71972

Appendix Table 22: Parameter estimate for ARIMA((2),1, 0) (Theft)

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag	Variable	Shift
MU	-0.03782	0.00984	-3.84	0.0001	0	Theft_qtrlog	0
AR1,1	-0.19299	0.11253	-1.71	0.0864	2	Theft_qtrlog	0
NUM1	0.008229	0.0134	0.61	0.539	0	qtr4	0
NUM2	0.0972	0.01552	6.26	<.0001	0	qtr3	0
NUM3	0.0778	0.01677	4.64	<.0001	0	qtr2	0
NUM4	-0.03728	0.01612	-2.31	0.0208	0	easter	0

## Fraud

Appendix Figure 4: ACF and PACF for fraud after first differencing



Appendix Table 23: Tentative Order Selection Test for fraud after first differencing

ARMA(p+d,q) Tentative Order Selection Test					
SCAN			ESACF		
p+d	q	BIC	p+d	q	BIC
2	2	-5.19713	2	2	-5.19713
3	1	-5.16111	3	2	-5.18576
8	0	-5.11397	4	2	-5.16951
			5	2	-5.14392
			8	4	-5.02725
			0	8	-4.63645
			1	8	-4.59352



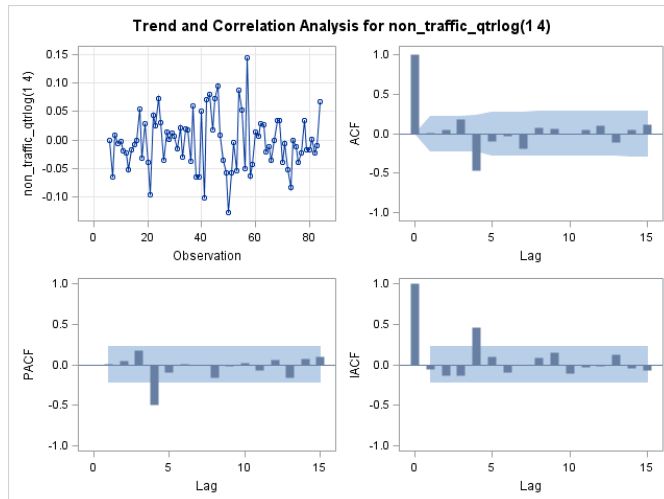
Appendix Table 24: Parameter estimate for ARIMA(2,1, 0) (Fraud)

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag	Variable	Shift
MU	-0.06928	0.01546	-4.48	<.0001	0	Fraud_qtrlog	0
AR1,1	-0.20537	0.11392	-1.8	0.0714	1	Fraud_qtrlog	0
AR1,2	-0.23257	0.11693	-1.99	0.0467	2	Fraud_qtrlog	0
NUM1	-0.01062	0.02264	-0.47	0.639	0	qtr4	0
NUM2	0.13749	0.0253	5.43	<.0001	0	qtr3	0
NUM3	0.12791	0.02265	5.65	<.0001	0	qtr2	0

## Seasonal Differencing

### Non-traffic

Appendix Figure 5: ACF and PACF for Non-traffic after seasonal and first differencing



Appendix Table 25: Tentative Order Selection Test for Non-traffic after seasonal and first differencing

ARMA(p+d,q) Tentative Order Selection Test					
SCAN			ESACF		
p+d	q	BIC	p+d	q	BIC
4	0	-6.43403	0	4	-6.29335
0	4	-6.29335	1	4	-6.23817
			2	4	-6.18323
			4	4	-6.24548
			5	4	-6.25605
			6	4	-6.22324
			7	3	-6.21173

Appendix Table 26: Parameter estimate for ARIMA(0,(1,4),(4)) (Non-traffic)

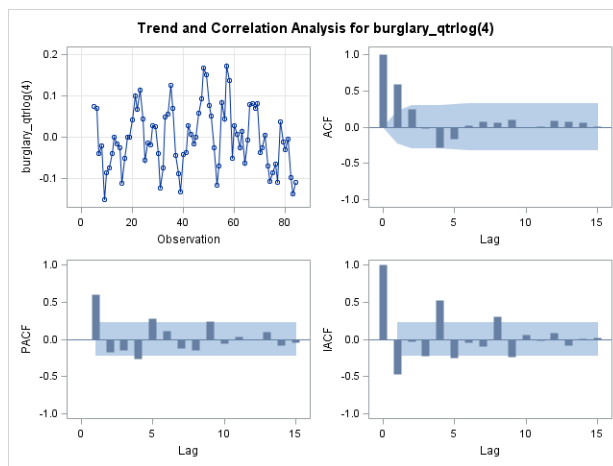
Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag
MU	-0.00232	0.001843	-1.26	0.2082	0
MA1,1	0.63496	0.09433	6.73	<.0001	4

Appendix Table 27: Parameter estimate for ARIMA((3,4)(1,4)(0)) (Non-traffic)

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag
MU	-0.00327	0.003649	-0.89	0.3708	0
AR1,1	0.18455	0.09807	1.88	0.0599	3
AR1,2	-0.47705	0.09809	-4.86	<.0001	4

## Burglary

Appendix Figure 6: ACF and PACF for burglary after seasonal differencing



Appendix Table 28: Tentative Order Selection Test for burglary after seasonal differencing

ARMA(p+d,q) Tentative Order Selection Test					
SCAN			ESACF		
p+d	q	BIC	p+d	q	BIC
0	3	-5.75216	4	2	-5.8181
4	2	-5.8181	0	3	-5.75216
5	1	-5.84242	2	3	-5.76705
			7	4	-5.67698
			8	4	-5.62938

Appendix Table 29: Parameter estimate for ARIMA(1,0,0)(1,1,0)<sub>4</sub> (Burglary)

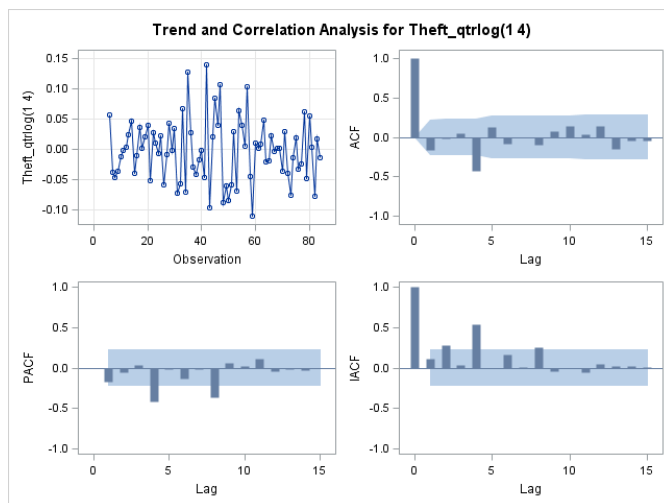
Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag
AR1,1	0.6867	0.08581	8	<.0001	1
AR2,1	-0.46285	0.10319	-4.49	<.0001	4

## Theft

Appendix Table 30: Tentative Order Selection Test for theft after seasonal and first differencing

ARMA(p+d,q) Tentative Order Selection Test					
SCAN			ESACF		
p+d	q	BIC	p+d	q	BIC
0	4	-6.15105	0	4	-6.15105
8	0	-6.11293	1	4	-6.13111
			2	4	-6.07655
			3	4	-6.02634
			4	4	-5.97192
			5	4	-5.91838
			6	4	-5.89357
			7	4	-5.86766
			8	3	-5.98183

Appendix Figure 7: ACF and PACF for theft after seasonal and first differencing

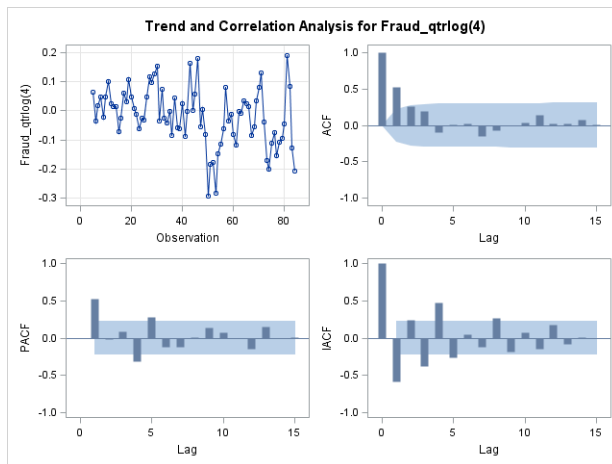


Appendix Table 31: Parameter estimate for ARIMA (0,(1,4), (4)) (Theft)

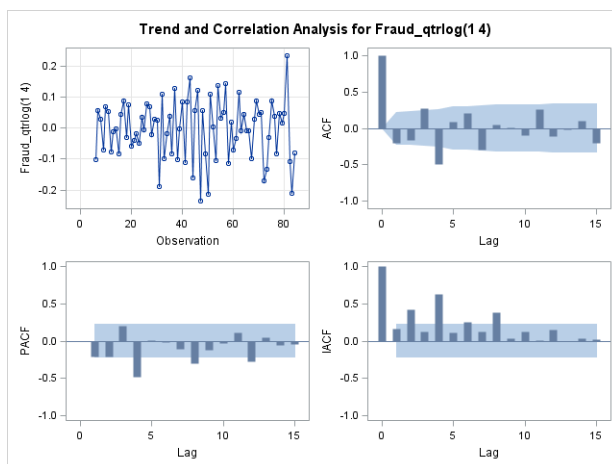
Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag
MU	-0.0009	0.001425	-0.63	0.5272	0
MA1,1	0.74821	0.09413	7.95	<.0001	4

## Fraud

Appendix Figure 8: ACF and PACF for fraud after seasonal differencing



Appendix Figure 9: ACF and PACF for fraud after seasonal and first differencing



Appendix Table 32: Tentative Order Selection Test for fraud after seasonal and first differencing

ARMA(p+d,q) Tentative Order Selection Test					
SCAN			ESACF		
p+d	q	BIC	p+d	q	BIC
2	4	-5.08937	0	6	-5.01337
5	1	-4.95536	2	6	-5.0037
0	7	-5.02413			
8	0	-5.0334			

Appendix Table 33: Parameter estimate for ARIMA(0,1,1)(0,1,1)<sub>4</sub> (Fraud)

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag
MU	-0.00171	0.001098	-1.56	0.119	0
MA1,1	0.28568	0.1105	2.59	0.0097	1
MA2,1	0.86492	0.0913	9.47	<.0001	4

# Appendix E SAS code

## Principal Component Analysis

```
/*This code performs Principal Component Analysis based on
unemployment_rate_all, real_gdp_growth and inflation. Output dataset
is 'pcl'*/
proc princomp data=ppl_chg_nz out=pcl;
var unemployment_rate_all real_gdp_growth inflation
;run;
```

## VAR and VECM

```
%let for=for;
%let no=3;
%let model=VAR_econ;
%let var=burglary;
%let normalize=burglary_qtrlog;
%let normal=burglary_ratelog;
%let p=4;
%let n=20;
%put &for&no&var&model&p=&&for&no&var&model&p;
%put test&var&model&p=&&test&var&model&p;
/*Above code is to setup macro variable, this also allows SAS to
automatically name the output dataset in a given format*/
ods graphics on;
proc varmax data=ppl_chg_nz plot=(all)
outstat=test&var&model&p; /*Output model diagnostics in to a
dataset*/
id qtrdate interval=qtr;
model
&normalize
&normal
avg_incomelog
real_gdp_growth1
Hld_debt_disp_income_ptlog
=
easter/*Specify endogenous and exogenous variable, easter is
exogenous variable in this code*/
/p=&p/*Specify p value*/
nseason=4/*Specify seasonal dummy, 4 means quarterly data*/
lagmax=20 dfest print=(roots)
dif=(&normalize(4) &normal(4) avg_incomelog(1)
Hld_debt_disp_income_ptlog(1,4) real_gdp_growth1(1))
method=ml/*Specify estimate to be calculated using Max Likelihood*/
/*ECM=(RANK=5 normalize=&normalize)*/
/*Only include this line when VECM is used*/
print=(diagnose estimates iarr impulse=(all));
restrict SD(4,1)=0, SD(4,2)=0, SD(4,3)=0,SD(5,1)=0, SD(5,2)=0,
SD(5,3)=0;
/*restrict is used when set some of the variable to 0. In this code,
some of seasonal dummies are set to 0*/
```

```
causal group1=(&normal) group2=(Hld_debt_disp_income_ptlog);
/*Causality test*/
output lead=&n out=for&no&var&model&p;
/*Specify number of quarters need to be forecasted, and also name of
the output dataset*/
run;
```

## **ARIMA**

```
%let for=for;
%let no=3;
%let model=ARIMA;
%let var=theft;
%let p=(0);
%let q=(4);
%let pq=0_14_4;
%let n=20;
%put &for&no&var&model&pq=&&for&no&var&model&pq;
/*Above code is to setup macro variable, this also allows SAS to
automatically name the output dataset in a given format*/
ods graphics on;
PROC ARIMA DATA=ppl_chg_nz;
IDENTIFY
VAR=theft_qtrlog(1,4)
/*Specify dependent variable and apply differencing.
In this code, both first and seasonal differencing applied*/
esacf scan p=(0:8) q=(0:8)
/*This is the automatic P and Q selection from SAS.
p=(0:8) q=(0:8) specify SAS to select lag length up to 8*/
crosscorr=(easter)/*Specify independent variable*/
nlag=15
STATIONARITY=(adf=(0,1,4));/*Stationary test using ADF test*/
ESTIMATE
input=(easter)
PLOT
P=&p Q=&q/*Specify P and Q value*/
outstat=test
(drop=_TYPE_ where=(_STAT_ in('AIC', 'SBC', 'SSE')))
/*Output model diagnostics dataset. The dataset only includes AIC,
BIC and SSE*/
noconstant
method=ml;
FORECAST
LEAD=&n /*Specify number of quarters need to be forecasted*/
id=qtrdate
interval=qtr
OUT=&for&no&var&model&pq;/*Specify name of the output dataset*/
run;
```