Confronting Theory with Data: the Case of DSGE Modeling

Niraj Poudyal

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics, Science

Aris Spanos, Chair
Richard Ashley
Hans Haller
Kwok Ping Tsang

November 7, 2012
Blacksburg, Virginia

Keywords: DSGE model, statistical adequacy, VAR, Student’s t VAR, misspecification testing
Copyright 2012, Niraj Poudyal
Confronting Theory with Data: the Case of DSGE Modeling

Niraj Poudyal

ABSTRACT

The primary objective of this is to confront the DSGE model (Ireland, 2011) with data in an attempt to evaluate its empirical adequacy. The perspective used for this evaluation is based on unveiling the statistical model (structural VAR) behind the DSGE model, with a view to test its probabilistic assumptions vis-a-vis the data. It is shown that the implicit statistical model is seriously misspecified and the information from mis-specification (M-S) testing is then used to respecify the original structural VAR in an attempt to achieve statistical adequacy. The latter provides a precondition for the reliability of any inference based on the statistical model. Once the statistical adequacy of the respecified model is secured through thorough M-S testing, inferences like the likelihood-ratio test for the overidentifying restrictions, forecasting, impulse response analysis are applied to the original DSGE model to evaluate its empirical adequacy. At the end, the same inferential procedure is applied to the CAPM model.
Acknowledgments

I want to gratefully express thanks to my advisor, Aris Spanos, for encouraging me to pursue a Ph.D. and helping me to sail over the wide field of Econometrics. His encouraging teaching and priceless advice have been absolutely essential for the development of this research. My special thanks also go to Byron Tsang for his constructive suggestions that enabled me to complete my dissertation. I would like to thank my other committee members, Richard Ashley and Hans Haller for their valuable suggestions and interest in my work. I am also thankful to the staff members of our department: Amy Standford and Sara Tickle for their assistance during my years in the Department. Special thanks go to my friend, Nabin Baral, for his endless suggestions and support to improve my writing skills and critical thinking. To my father Dhruba Poudyal, who always gave me inspiration to work hard and to my mother Mira Poudyal who always encouraged me to pursue higher goals and happiness in life. Finally and most importantly, it would have been impossible to accomplish such a job without the kind of logistic and emotional support my wife, Gita Bhattarai, provided.
Dedication

I would like to dedicate this dissertation to all those who strive for human dignity and reason.
# Contents

List of Figures x

List of Tables xiii

1 Macroeconomic Modeling 1

1.1 Brief History of Macroeconomic Modeling 1

1.2 DSGE Models 4

1.3 Vector Autoregression (VAR) Models 9

1.3.1 DSGE as a Restricted VAR 10

1.4 Moving Forward in Macroeconomic Modeling 11

1.4.1 Non-Normality 11

1.4.2 Dynamic Heteroskedasticity 12

1.4.3 Modeling Heterogeneity 13

1.5 Reliability of Inference 14

1.6 Realisticness of DSGE Models 18

1.7 Bayesian Inference on DSGE Models 21

1.8 Concluding Remark 24

1.9 A Brief Overview of the Chapters 24

2 DSGE Modeling: A Probability Reduction Approach 26

2.1 Introduction 26

2.2 Ireland’s Structural Model 29
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.1 Representative Household</td>
<td>30</td>
</tr>
<tr>
<td>2.2.2 Finished Goods-producing Firm</td>
<td>31</td>
</tr>
<tr>
<td>2.2.3 Intermediate Goods-producing Firm</td>
<td>32</td>
</tr>
<tr>
<td>2.2.4 Central Bank</td>
<td>34</td>
</tr>
<tr>
<td>2.2.5 Efficient Allocation and Output Gap</td>
<td>34</td>
</tr>
<tr>
<td>2.2.6 Symmetric Equilibrium</td>
<td>36</td>
</tr>
<tr>
<td>2.2.7 Linearized Model</td>
<td>36</td>
</tr>
<tr>
<td>2.2.8 Identities of the Model</td>
<td>40</td>
</tr>
<tr>
<td>2.2.9 Solution of the Model</td>
<td>40</td>
</tr>
<tr>
<td>2.3 Statistical Model</td>
<td>43</td>
</tr>
<tr>
<td>2.3.1 Reduced Form</td>
<td>44</td>
</tr>
<tr>
<td>2.3.2 Identification and Estimation</td>
<td>45</td>
</tr>
<tr>
<td>2.4 Exploratory Data Analysis (EDA)</td>
<td>49</td>
</tr>
<tr>
<td>2.5 Probabilistic Reduction and VAR Model</td>
<td>53</td>
</tr>
<tr>
<td>2.6 Misspecification (M-S) Tests for VAR Model</td>
<td>56</td>
</tr>
<tr>
<td>2.6.1 Standardized Residuals</td>
<td>57</td>
</tr>
<tr>
<td>2.6.2 M-S Testing Auxiliary Regression for VAR Model</td>
<td>58</td>
</tr>
<tr>
<td>2.6.3 Interpretation of the M-S Testing Results</td>
<td>58</td>
</tr>
<tr>
<td>A Solution to the System of Expectational Difference Equations</td>
<td>61</td>
</tr>
</tbody>
</table>

3 Multivariate Student’s t Dynamic Models                               | 64   |
| 3.4 Introduction                                                       | 64   |
| 3.5 Multivariate Student’s t Distribution                              | 65   |
| 3.5.1 Joint Distribution                                               | 65   |
| 3.5.2 Special Cases of the Student’s t Distribution                    | 68   |
| 3.5.3 Moments                                                         | 68   |
| 3.5.4 Marginal Distribution                                            | 69   |
| 3.5.5 Conditional Distribution                                         | 70   |

3.4 Introduction

3.5 Multivariate Student’s t Distribution

3.5.1 Joint Distribution

3.5.2 Special Cases of the Student’s t Distribution

3.5.3 Moments

3.5.4 Marginal Distribution

3.5.5 Conditional Distribution
3.6 Matrix Variate Student’s $t$ Distribution ........................................... 71
  3.6.1 Conditional Distribution .............................................................. 73
3.7 Student’s $t$ VAR (St-VAR) Model ..................................................... 73
  3.7.1 Joint, Conditional and Marginal Distributions ................................. 75
  3.7.2 Special Cases of the St-VAR Model .............................................. 77
3.8 Stationary St-VAR Model ................................................................. 78
3.9 Heterogeneous St-VAR Model ............................................................ 79
3.10 Dynamic Multivariate Student’s $t$ Regression ................................. 80
3.11 Student’s $t$ Distribution Having Marginals with Different Degrees of Freedom 82
3.12 Asymmetric Student’s $t$ Distribution .............................................. 82
3.13 Estimation: Maximum Likelihood .................................................... 83
  3.13.1 Inference ....................................................................................... 83
4 Respecification: Student’s $t$ VAR (St-VAR) Model ............................. 85
  4.1 Introduction ....................................................................................... 85
  4.2 Why Not a GARCH Model? ............................................................... 86
  4.3 Respecification Strategy ................................................................. 87
  4.4 Estimation Results ............................................................................ 89
    4.4.1 Autoregressive Function .............................................................. 90
    4.4.2 Fitted Values ............................................................................... 92
    4.4.3 Forecasting: DSGE Model vs St-VAR Model ............................... 95
    4.4.4 Autoskedastic Function: Estimation ............................................ 98
    4.4.5 Autoskedastic Function: Interpretation ...................................... 100
  4.5 M-S Testing for Heterogeneous St-VAR Model .................................. 106
    4.5.1 Standardized Residuals ............................................................... 106
    4.5.2 M-S Testing Auxiliary Equations for St-VAR Model .................. 111
    4.5.3 M-S Testing Results for St-VAR Model ..................................... 112
  4.6 Impulse Response Function (IRF) Analysis ....................................... 114
    4.6.1 Mean Impulse Response Function ............................................. 115
5.12 Likelihood Ratio Tests .................................................. 163

6 Summary and Conclusion ................................................. 164
   6.1 Introduction ............................................................... 164
   6.2 Contributions ............................................................ 165
   6.3 Future Research .......................................................... 167

References ................................................................. 170

F M-S Tests for Distributional Assumptions ......................... 181
   F.1 Skewness Kurtosis (SK): Normality ............................... 181
   F.2 Skewness Kurtosis (SK): Student’s $t$ ............................ 181
   F.3 Kolmogorov-Smirnov (KS) ........................................... 182
   F.4 Anderson-Darling (AD) ................................................. 182

G R code for St-VAR Model .............................................. 183
List of Figures

2.1 Probability Reduction (PR) Approach ........................................... 28
2.2 Per Capita Real Gross Domestic Product (GDP) .......................... 51
2.3 Growth Rate of Per Capita Real GDP ........................................ 51
2.4 GDP Deflator ................................................................. 52
2.5 Inflation Rate (Log Difference of GDP Deflator) ......................... 52
2.6 Quarterly Gross Interest Rate (Three Months Treasury Bill) ........... 53
3.1 Bivariate Normal Density ($\mu = 0, \Sigma = I, \nu = 3$) .................... 66
3.2 Bivariate Student’s t Density ($\mu = 0, \Sigma = \frac{1}{3} I, \nu = 3$) .......... 66
3.3 Bivariate Normal Density Contour ($\mu = 0, \Sigma = I, \nu = 3$) ......... 67
3.4 Bivariate Student’s t Density Contour ($\mu = 0, \Sigma = \frac{1}{3} I, \nu = 3$) .. 67
3.5 Heteroskedastic Autoskedastic Function ($\mu = 0, \Sigma = I$) .......... 71
4.1 Fitted Values for Growth Rate of GDP Per Capita .......................... 94
4.2 Fitted Values for Inflation Rate ............................................. 94
4.3 Fitted Values for Interest Rate .............................................. 95
4.4 Prediction of Growth Rate of GDP Per Capita .............................. 97
4.5 Prediction of Inflation Rate .................................................. 97
4.6 Prediction of Interest Rate ................................................... 98
4.7 Fitted Conditional Variance for Growth Rate .............................. 103
4.8 Fitted Conditional Variance for Inflation Rate ............................. 103
4.9 Fitted Conditional Variance for Interest Rate ............................. 104
4.10 Fitted Conditional Covariance for Growth Rate and Inflation Rate . . . . . . 104
4.11 Fitted Conditional Covariance for Growth Rate and Interest Rate . . . . . . 105
4.12 Fitted Conditional Covariance for Inflation Rate and Interest Rate . . . . . . 105
4.13 Unstandardised Growth Rate Residual . . . . . . . . . . . . . . . . . . . . . 107
4.14 Standardised Growth Rate Residual . . . . . . . . . . . . . . . . . . . . . . . 107
4.15 Unstandardised Inflation Rate Residual . . . . . . . . . . . . . . . . . . . . . 108
4.16 Standardised Inflation Rate Residual . . . . . . . . . . . . . . . . . . . . . . 108
4.17 Unstandardised Interest Rate Residual . . . . . . . . . . . . . . . . . . . . . 109
4.18 Standardised Interest Rate Residual . . . . . . . . . . . . . . . . . . . . . . 109
4.19 Effect of 1% Rise in Growth Rate on Growth Rate . . . . . . . . . . . . . . 118
4.20 Effect of 1% Rise in Growth Rate on Inflation Rate . . . . . . . . . . . . . 118
4.21 Effect of 1% Rise in Growth Rate on Interest Rate . . . . . . . . . . . . . . 119
4.22 Effect of 1% Rise in Inflation Rate on Growth Rate . . . . . . . . . . . . . 119
4.23 Effect of 1% Rise in Inflation Rate on Inflation Rate . . . . . . . . . . . . . 120
4.24 Effect of 1% Rise in Inflation Rate on Interest Rate . . . . . . . . . . . . . 120
4.25 Effect of 1% Rise in Interest Rate on Growth Rate . . . . . . . . . . . . . . 121
4.26 Effect of 1% Rise in Interest Rate on Inflation Rate . . . . . . . . . . . . . 121
4.27 Effect of 1% Rise in Interest Rate on Interest Rate . . . . . . . . . . . . . . 122
4.28 Effect of 1% Rise in Growth Rate on Conditional Variance . . . . . . . . . . 123
4.29 Effect of 1% Rise in Inflation Rate on Conditional Variance . . . . . . . . . 123
4.30 Effect of 1% Rise in Interest Rate on Conditional Variance . . . . . . . . . 124

D.1 Identification of Ireland (2011) Model ($N = 1200$) . . . . . . . . . . . . . 137
D.2 Identification of Ireland (2011) Model ($N = 1200$) . . . . . . . . . . . . . 138
D.3 Identification of Ireland (2004) Model ($N = 955$) . . . . . . . . . . . . . 139

5.1 Three Month Treasury Bill Log Returns . . . . . . . . . . . . . . . . . . . . 151
5.2 CITI Log Returns . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 151
5.3 PFE Log Returns .................................................. 152
5.4 GM Log Returns .................................................. 152
5.5 SP500 Log Returns ............................................... 153
5.6 Fitted Values for CITI Log Returns ......................... 160
5.7 Fitted Values for PFE Log Returns ......................... 161
5.8 Fitted Values for GM Log Returns ......................... 161
5.9 Fitted Conditional Variance for CITI Log Returns .... 162
List of Tables

2.1 Dynamic Stochastic General Equilibrium (DSGE) Model . . . . . . . . . . . 39
2.2 Estimates of Structural Parameters . . . . . . . . . . . . . . . . . . . . . . . 47
2.3 Specification of VAR(2) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55
2.4 Normal VAR(2) Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 56
2.5 M-S Tests for VAR(2) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 59
2.6 M-S Test Results for VAR(2) . . . . . . . . . . . . . . . . . . . . . . . . . . 59
2.7 Normality Tests for VAR(2) . . . . . . . . . . . . . . . . . . . . . . . . . . . 59

3.1 Probabilistic Reduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 74
3.2 Student’s $t$ VAR($l;\nu$) . . . . . . . . . . . . . . . . . . . . . . . . . . . . 79
3.3 Student’s $t$ VAR($l = 3;\nu$) . . . . . . . . . . . . . . . . . . . . . . . . . . 79
3.4 Heterogeneous Student’s $t$ VAR($3;\nu$) . . . . . . . . . . . . . . . . . . . . 81

4.1 Respecification Strategy . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 89
4.2 Estimation Result of St-VAR and VAR . . . . . . . . . . . . . . . . . . . . . . 91
4.3 MSEP for 12 Periods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 96
4.4 Normal VAR(2): Conditional Covariance . . . . . . . . . . . . . . . . . . . . 99
4.5 M-S Tests for Heterogeneous St-VAR($3;\nu = 3$) . . . . . . . . . . . . . . . 113
4.6 M-S Tests for St-VAR Model . . . . . . . . . . . . . . . . . . . . . . . . . . . 113
4.7 Skewness-Kurtosis and Kolmogorov-Smirnov Tests . . . . . . . . . . . . . . 113
4.8 Likelihood Ratio Test . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 125

B.1 DSGE Model: 2011 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 132
B.2 DSGE Model: 2004 ................................................................. 133

C.1 Likelihood Ratio Test for Ireland (2004) Model . . . . . . . . . . . . . . . . . 135

E.1 Conditional Variance of \( p_t \): \( VAR(p_t|Z_{t-1}^0) \) ........................................... 142
E.2 Conditional Variance of \( r_t \): \( VAR(r_t|Z_{t-1}^0) \) ........................................... 143
E.3 Conditional Covariance of \( \{y_t, p_t\} \): \( VAR(y_t, p_t|Z_{t-1}^0) \) ................. 144
E.4 Conditional Covariance of \( \{y_t, r_t\} \): \( VAR(y_t, r_t|Z_{t-1}^0) \) ................. 145
E.5 Conditional Covariance of \( \{p_t, r_t\} \): \( VAR(p_t, r_t|Z_{t-1}^0) \) ................. 146

5.1 The Multivariate Normal Regression Model ........................................... 149
5.2 Unrestricted Static CAPM Model ......................................................... 153
5.3 M-S Tests for Unrestricted CAPM ......................................................... 155
5.4 M-S Tests on Unrestricted Static CAPM ................................................. 156
5.5 M-S Tests for St-VAR(1; \( \nu = 3 \)) ......................................................... 157
5.6 Kolmogorov-Smirnov Test ................................................................. 158
5.7 Dynamic CAPM for Period 1 ............................................................. 158
5.8 Dynamic CAPM for Period 2 ............................................................. 159
5.9 Dynamic CAPM for Period 3 ............................................................. 159
5.10 Likelihood Ratio Test ........................................................................ 163
Chapter 1

Macroeconomic Modeling

1.1 Brief History of Macroeconomic Modeling

In general, macroeconomists are interested in modeling cycles and trends of major macroeconomic series such as growth rate of output, interest rate, inflation, consumption and employment. The volatility and dynamics of such individual macro series and the relationships among these series have become a critical research area in macroeconomics.

Modern macroeconometric modeling began in the mid 1930s after Tinbergen (1937) built a system of equations into an econometric model of business cycle for the Netherlands and United States. He used economic theory to derive a model with behaviourally motivated dynamic equations and statistical methods to test the model against data (see Bårdsgaard et al. (2005) for detail). In his attempt to model business cycles, the role of the probabilistic assumptions underlying an econometric model was not fully recognized. The emphasis was almost exclusively on estimation, with very little references to tests of significance. He evaluated his models using “goodness-of-fit” and “economic validity” criteria, which are neither necessary nor sufficient condition to ensure the reliability of inference. As a result, specification of the model relied totally on economic theory without giving due recognition to the role of data.

Haavelmo (1944), who introduced the simultaneous-equations model into macroeconomic modeling, pointed out that failure to assure the validity of the probabilistic assumptions will
invariably give rise to highly misleading inferences. He argued that the joint distribution of all observable variables for the whole sample period provides the most general framework for statistical inference. This applies to specification, identification, estimation and hypothesis testing. It was the beginning of the probabilistic approach to econometrics. The issues raised by Haavelmo (1944) on empirical modeling have been neglected for many decades. Even today, very little emphasis is given to the validity of the implicit probabilistic (statistical) assumptions behind any macroeconomic model. This research attempts to bring the aforementioned issue in the forefront of current macroeconomic modeling with a particular focus on the dynamic stochastic general equilibrium (DSGE) model and capital asset pricing model (CAPM).

Building up on the tradition of Tinbergen (1937) and Haavelmo (1943, 1944), specification of a macroeconomic model has been influenced by either economic theory, the econometric analysis of historical data or a combination of both. The contentious issue among the modelers is whether to emphasize economic theory or information from data for modeling purpose (Pagan, 2003). In the second half of the twentieth century, these two distinct approaches to modeling macroeconomic data have been further developed and followed extensively by academics and policy makers alike. Consequently, these two approaches of modeling gave rise to two basic types of models in macroeconomics:

(a) Dynamic stochastic general equilibrium (DSGE) models

(b) Vector autoregressive (VAR) models

The first type of models are driven mainly by theory dubbed as Pre-Eminence of Theory (PET) perspective by Spanos (2009). These models will be called the PET models and the approach will be called the PET approach for the rest of this dissertation. The simultaneous-equation models and the DSGE models are some of the examples of the PET models. The
Cowles commission laid the foundation for the development of the simultaneous-equation models in macroeconomics emphasizing the issue of identification - the link between the statistical and the structural model. The DSGE models are also dominated by identification issues. Economic theory plays an important role in such identification procedure (see Klein (1988) for detail survey). Issues related to the validity of probabilistic assumptions (i.e. statistical adequacy) never became an integral part of the PET approach.

The PET approach in macroeconomic modeling has been criticized from two fronts.

(a) Lucas (1976) and Sims (1980)

(b) London School of Economics (LSE)

On the one hand, critiques such as Lucas (1976) and Sims (1980) argued that there is a need of even stronger and realistic theoretical basis for these models. Lucas (1976) claimed that there is a need to model the expectations without which the conditional models failed to explain and forecast, because of regime shifts caused by policy changes and the shift in expectations. But Sims (1980) pointed out that the problem in the PET models is due to very strict identifying restrictions.

The second type of macroeconomic models known as Vector Autoregression (VAR), whose specification is much less dependent on economic theory, arose in macroeconomics from Sims’s criticisms. Box and Jenkins (1970) also belongs to this second category of macroeconomic modeling approach. The VAR approach, as recognised by Haavelmo, is based on the central idea of viewing all the inferences from the perspective of the joint distribution of observables. Structural VAR (SVAR) is a VAR in which parameter restrictions are determined by theory and the rest of the modeling work such as lag length is determined by the data.

The second line of critique of the PET modeling approach came from the LSE methodology. This methodology largely ascribes the failure of those early models to missing dynamics or
model misspecification (violation of implicit probabilistic assumptions of a model for the data used), and hence emphasizes the importance of testing and evaluating econometric models (Ericsson, 2005; Hendry, 1993, 1995; Mizon, 1995). The VAR models, although much more data based, are also vulnerable to this line of criticism, because the probabilistic assumptions underlying any given VAR can also be violated for the given data.

Bårdsen et al. (2005), closely following what the LSE methodology stood for, put forward a methodology which stands somewhere in between the two approaches. They propose high emphasis on testing the statistical assumptions of the model before using it for inference.

What is important to emphasize at the outset is that these different approaches are not incompatible. Indeed, one of the objectives of this research is to show how the two approaches can be viewed as complementary. Because the DSGE and the VAR models are a critical part of this research, these modeling approaches are briefly described and critiqued in the following paragraphs.

1.2 DSGE Models

The roots of the dynamic stochastic general equilibrium (DSGE) models can be traced back to the Walrasian research program and the work of Keynes (1930). The DSGE models aim to explain the key macroeconomic phenomena such as business cycle and economic growth in terms of individual intertemporal decision making facing stochastic shocks. Basically, the DSGE models begin with rational agents such as consumers, firms and a central planner. They maximize their life time objective function (such as utility, profit and welfare) with respect to variables such as consumption, leisure and input subject to some constraints (such as budget constraint and technology constraint). Most often, the agents are supposed to
have infinite lifetime. This assumption is invoked because finite life creates computational complexity for the optimization problem. The objective functions are expressed in terms of the expected discounted present value. The first order condition of this dynamic optimization problem usually gives rise to the following behavioural relationships among the major macroeconomic variables (Ireland, 2004, 2011)

(a) Expectational IS curve as a relationship between interest rate and output (Kerr and King, 1996).

(b) Phillips curve as a relationship between inflation and output (Calvo, 1983).

(c) Taylor policy rule as a relationship between interest rate, output and inflation (Taylor, 1993).

There are many different versions of each of these relationships in the literature, each of them supporting a particular macroeconomic theory.

Besides these relationships, there are some major extensions. King and Plosser (1984), building on the works of Tobin (1963) developed a model that can account for the correlation between endogenous money and business cycles with special focus on the banking system. Their empirical strategy is aimed at ‘delivering a ‘good fit’ and identification.

Hansen (1985) tried to model the indivisibility of labor in the dynamic stochastic growth model. Before computing the statistics for the US economy, the time series was logged and detrended. This, indirectly, tries to address the problem of misspecification that the structural model might have due to the trending nature of macroeconomic data. His empirical strategy is calibration, which is equally vulnerable to model misspecification.

Barro and Sala-i Martin (1992) extended the DSGE models to include tax-financed government services that affect production and utility. His theory has implications for relations between the size of government, output growth rate and savings. Barro (1992) evaluates his
empirical results against the theoretical meaningfulness of the signs and the magnitude of
the estimated parameters.

Backus et al. (1992) add foreign sector to the DSGE framework via exchange rates. They
found that their model is not able to reproduce some of the properties of real data such as
strong autocorrelation. They use the generalized method of moments (GMM) to estimate
the model and they evaluate their model using prediction and fit.

As monopolistic competition is the major component of these models, staggered price has
been extensively used in the literature (Christiano et al., 2005; Gali and Gertler, 1999).
Mankiw and Reis (2001) introduced the idea of sticky information in the literature. Farmer
(1999) introduced the concept of overlapping generation to display multiple rational expecta-
tion equilibria. Long and Long and Plosser (1983) proposed a multi-sector model empha-
sizing the co-movements of the different sectors of the economy. Lucas (1978) discussed the
links between finance and macroeconomics using the DSGE framework. Asset pricing theory
has been incorporated by Christiano et al. (2005).

Kydland and Prescott (1982) started the real business cycle (RBC) revolution on the basis
of the DSGE macroeconomic models. For the first time in history, they tried to quantify
the general equilibrium model using post-war period US data. They fitted the post war
US quarterly data on real output, employment, investment, inventory, capital stock, hours,
productivity, real interest rate using calibration techniques. Calibration is a method of sim-
ulating data using the prior information on the parameters of the model and comparing the
statistical properties of the simulated data and the real data. Their approach is also aimed
at achieving the best possible fit. They tried to improve the realisticness of the model by
introducing multiple period production processes.

Kydland and Prescott (1982, 1990) find that the simulated data from their model show the
same patterns of volatility, persistence, and comovement as are present in the U.S. data.
When comparing the simulated data and the real data, no emphasis has been given on the
choice of statistically meaningful statistics for the data used. For example, they compute autocorrelations and correlations whose meaningfulness relies on the validity of underlying statistical premises. If the model selected is statistically misspecified, the whole comparison becomes unreliable.

The overall trend in this field of research is an attempt to improve the theoretical realistic-ness (the validity of substantive assumptions) of the DSGE models by incorporating more variables and sectors of the economy. The criteria used for model evaluation is fit and theoretical meaningfulness, which are neither necessary nor sufficient condition for securing the reliability of inference. While focusing on the realisticness issue, the reliability of inference (validity of the implicit probabilistic assumptions) has been completely ignored.

In all these DSGE models, the typical first order condition has the following four characteristics:

(a) Non-linear relationship
(b) Expectational relationship
(c) Dynamic relationship
(d) Normality of errors/shocks

While the dynamic nature of the first order condition does not pose any problem for inference, non-linearity and expectations do. The dynamic nature of the first order condition helps modeling persistent macroeconomic data. Most of the DSGE models embed autoregressive and moving average shocks in order to model the highly persistence nature of the macroeconomic data. The non-linearity is dealt with linearizing around some pre-defined steady state using methods such as Taylor’s expansion. But, the current literature in macroeconomics lacks a proper definition of the steady state. Since the models are based on the
assumption of stationary data, the steady state is supposed to be constant, which will be
challenged later in this research. Once the linearisation is complete, the model is in the form
of linear expectational dynamic stochastic system of equations.
Use of expectational relationship has its own problems. Hendry and Mizon (2010) argue that
the distribution of economic variables shifts over time making the present treatment of ex-
pectations in economic theories of inter-temporal optimization inappropriate. They further
show that the conditional expectations are neither unbiased nor minimum mean-squared
error predictors when unanticipated location shifts or breaks occur. Consequently, the law
of iterated expectations fails inter-temporally unless all the distributional shifts are perfectly
anticipated by all economic agents, a possibility contradicted by the recent financial crisis.
The stochastic linear dynamic expectational system can be solved using different methods
developed by Klein (2000) and Blanchard and Kahn (1980). The solution of the system is in
the state-space form. If the solution has some unobservable variables, it is estimated using a
combination of methods like Kalman filter (Kalman, 1960) and maximum likelihood, assum-
ing that the shocks are normally distributed. The problem of non-Normality is dealt with
by using some other filters such as Particle filter (Ristic et al., 2004). But these other filters
are highly unreliable for finite samples and they are based on mathematical approximation
theories.
The widely held belief that the DSGE modeling framework provides natural links to data
via the simulated trajectories of the variables involved, correlations among observables and
impulse response analysis, on closer examination constitutes a vice not virtue, because such
links are anaemic and lack credibility without statistical adequacy. The statistical adequacy
is guaranteed by testing the probabilistic assumptions of the macroeconomic model for the
data used. This will be the main theme of this dissertation.
The DSGE model by Ireland (2011) will be used in this research. (Ireland (2004) uses a
similar model and the same data to estimate it. The observation period of data is different
in the 2004 paper.) as an example to address the issues raised above. The choice of this particular model has two main objectives. First, it stands in the core of current DSGE modeling in a sense that it contains all the major components of a typical DSGE model. Second, this model, whether statistically adequate or not, is getting widespread attention in academia and policy circles. Using statistically inadequate models for policy simulations and decision making can be costly for the economy because the statistical results are unreliable.

1.3 Vector Autoregression (VAR) Models

The VAR model is a multivariate extension of the Box-Jenkins univariate autoregressive (AR) model (Box and Jenkins, 1970). It gives a wider role to play for data, because inference from the VAR can be made free from any theoretical (structural) restrictions. But it is argued to be lacking any microeconomic foundation and hence useless for policy makers. To address this criticism in part, a newer version of the VAR, known as structural VAR (SVAR) has received much more attention. In the SVAR, explicit identifying restrictions are imposed on some of the VAR parameters (Bernanke, 1986; Blanchard and Watson, 1986). The SVAR has been used to show the effect of money on output (Sims and Zha, 2005), the relative importance of supply and demand in business cycles (Blanchard and Quah, 1989), the effects of fiscal policy (Blanchard and Perotti, 2002) and so on.

Watson and Stock (2001) discuss extensively the relative potential of the VAR in terms of four major objectives of macroeconometric modeling: data description, forecasting, structural inference, and policy analysis. They argue that the VAR is good at the former two and lags behind at the later two. The most important feature of the VAR research is that the attention has stayed on cointegration issues (Johansen, 1988; Jusélius, 2006), which is testing within a prespecified model.
This approach does not address the broader issue of misspecification sufficiently. Their testing results become unreliable if the alternative model is misspecified for the given data (violation of probabilistic assumption by the data). The most significant drawback of the VAR models is the lack of statistical adequacy for given data. In this research the issue of statistical adequacy of the VAR models is raised and potentially statistically adequate models for the US data are proposed.

1.3.1 DSGE as a Restricted VAR

Consolo et al. (2009) show that the statistical model behind DSGE type model is a restricted VAR of some order. They try to address the issue of misspecification from the Bayesian perspective. Instead, my research will be focusing on how to do that in the classical (frequentist) framework.

Ravenna (2007) discusses the assumptions needed for a finite order VAR representation of any subset of a DSGE model variables to exist. Fernandez-Villaverde and Rubio-Ramirez (2007) discuss the invertibility problem and provide examples of well-specified DSGE models that lack a VAR representation. Even if it exists, the VAR representation of a DSGE model may require an infinite number of lags which is practically useless for a finite data set in macroeconomics. When one has an infinite order VAR representation as a statistical model for the given DSGE model, Chari et al. (2008) show that for a standard parametrization the coefficients in the VAR representation converge to zero extremely slowly, making a finite order VAR approximation unsuitable.

This research focuses on the DSGE models which have a finite order VAR representation. It will be shown in this research that a small order VAR is sufficient to capture the dynamics of the data. One has to remember that the issue of identification does not arise before achieving the statistical adequacy of the reduced form model for a given data set (Spanos, 1990).
Even without an exact identification, reliable inference can be drawn from the statistically adequate reduced form model. But the issue of statistical adequacy is equally important even if the VAR representation does not exist, for the reduced form might have some other form of statistical model.

1.4 Moving Forward in Macroeconomic Modeling

It is important to note that neither the DSGE models nor the VAR models address the problem of modeling the non-Normal, dynamic heteroskedastic and heterogeneous nature of macroeconomic data. Ignoring these crucial probabilistic structures of the data leads to statistically inadequate macroeconomic models making the whole inference drawn unreliable. An attempt will be made to address each of these issues in this research.

(a) Non-Normality

(b) Dynamic heteroskedasticity

(c) Heterogeneity

The significance of each of the aforementioned issues is discussed in the following paragraphs before discussing their effects on reliability of inference.

1.4.1 Non-Normality

One of the important contributions of this research is to take macroeconomic modeling beyond the conventional paradigm of Normality. This strategy has some very important implications for modeling macroeconomic data. Most often the non-Normality leads to either
non-linearity of mean, heteroskedasticity of variance or both. Rather than considering these features of the non-Normal model as burden, they can be useful in deciphering important structures of macroeconomic data. For example, misspecification tests in Chapter 2 show that the assumptions of homoskedasticity and Normality are severely violated by some important U.S. macroeconomic data. It will be shown that the macroeconomic data depicts the features of fat-tail distributions. The Student’s $t$ distribution will be proposed as an alternative modeling framework. The assumption of Student’s $t$ distribution will give rise to a statistically adequate model for the data.

1.4.2 Dynamic Heteroskedasticity

As pointed out earlier, most of the macroeconometric models impose Normality, linearity and homoskedasticity assumptions at the outset. Although heteroskedasticity and second order dependence have been the dominant feature of macroeconomic data, neither the DSGE models nor the VAR models have been able to model these features properly. Modeling these features requires the specification of the second moment of the distribution. This is not possible with the focus on first order linearized conditions and Normality assumptions, because they make the variance-covariance matrix to be constant over time and regressors. The second contribution of this research is to introduce the models based on the Student’s $t$ distributions, which allows to construct second order dependence and heteroskedasticity as an integral part of the model.

To model dynamic heteroskedasticity, the obvious choice for modelers is to use generalized autoregressive conditional heteroskedastic (GARCH) type models. The GARCH family of models have their origin in a paper by Engle (1982). The GARCH approach will not be used in this research due to some of its major drawbacks, which will be discussed in detail in Chapter 4. Spanos (2002) proposes the Student’s $t$ autoregressive model as an alternative to
the GARCH models. Heracleous (2003) proposes the Student’s $t$ VAR (St-VAR) for modeling dynamic heteroskedasticity. But her model imposes very stringent parametric restrictions to achieve fewer number of parameters. How these restrictions make her St-VAR model practically useless will be shown in Chapter 3. Building upon the works of Spanos (2002) and Heracleous (2003), this research will introduce the unrestricted St-VAR model.

### 1.4.3 Modeling Heterogeneity

Most of the macroeconomic models, particularly the DSGE models, assume homogeneity – constancy of parameters across time. Not enough attention has been paid in macroeconomics and finance to model heterogeneity in the data except using some form of the random walk process in the model. In Chapter 2, it will be shown that the assumption of homogeneity is severely violated by macroeconomic data even when the DSGE model includes random walk processes. This research will develop a heterogeneous St-VAR model to do the job. It will be shown that how a naive assumption on mean of the joint distribution can give rise to highly non-linear heterogeneity in the mean and the variance of the conditional distribution. This allows to model a complex heterogeneous structure of macroeconomic data within a framework of a single model.

Ignoring these three important features of macroeconomic data results in model misspecification, which compromises the reliability of inferences. In the next section, the reliability of inferences concerning the structural models such as DSGE models is discussed.
1.5 Reliability of Inference

The principal criteria for selecting one model (DSGE or VAR) over the other have been goodness-of-fit, economic meaningfulness (magnitudes and signs of the parameter estimates) and simplicity of the model. But these criteria are neither necessary nor sufficient conditions for securing the reliability of inference (Spanos, 2010). The reliability of estimation and inference relies on the nature of substantive and statistical assumptions imposed on the structure of the first order condition of the model (Johansen, 2006; Spanos, 2011). Spanos (2011) discusses the two sets of premises involved in modeling, and the importance of distinguishing them first for the reliability of inference. This research focuses on separating these two sets of assumptions and checking the validity of statistical assumptions first. The validity of the probabilistic assumptions is a necessary condition for assessing the validity of the substantive assumptions.

Spanos (2009) talks about the reliability of inference on models stemming from the PET perspective such as DSGE models. He further argues that any assessment concerning the sign, magnitude and significance of estimated coefficients, however informal, constitutes an inference whose credibility is completely undermined when the estimated model is statistically misspecified. With a special focus on DSGE models, the consequences of invoking invalid assumptions on the inferences are discussed elsewhere (Campos et al., 2005; Franchi and Jusélius, 2007; Johansen, 2006). Johansen (2006) argues that for the maximum likelihood approach to give reliable inference, the macroeconomic model has to satisfy two sets of assumptions: substantive and likelihood.

the U.S. data. Although some dynamics can be found in such DSGE models, they will be proved to be insufficient to cater to the second order dynamics and heterogeneity present in the data.

Heterogeneity results in unstable parameters that change with time, making the model worthless for inference purpose. Meeks (2005) shows that macroeconomic data severely violate the stationarity assumption imposed by the structural DSGE model developed by Ireland (2004). Through recursive estimation, he shows that the parameters of the model change across time, invalidating all the inferences drawn thereafter. Ireland (2004) also shows the rejection of $t$-invariance of parameters (homogeneity), although the independence assumption is found to be fairly good. Even with time varying parameters, he continues to make inferences from his model.

Johansen (2006) uses the Ireland (2004) model to illustrate the implication of incorrect specification on inference. He shows that U.S. data severely violates the assumption to $t$-invariance of parameters in Ireland’s (2004) model. He raises two important questions regarding the reliability of inference:

(a) Choice of a statistical model to achieve valid inference for the economic model,

(b) Given a correctly chosen statistical model, how reliable are the asymptotic results found in the statistical literature for the analysis of data at hand?

His response to the first problem is to choose a statistical model that describes the data adequately. In response to the second problems, he argues that the asymptotic sampling distribution under unit roots is unreliable even with a statistically adequate model. Testing under unit roots is testing within a pre-specified model, which is irrelevant when the statistical model is misspecified.

In the next few paragraphs, a new methodological framework proposed by Spanos (1986) to
address the issues related to the reliability of inference is discussed. This approach is supposed to bridge the gap between theory and data using a sequence of interconnected models with a view to delineate and probe for the potential errors at different stages of modeling. The emphasis is on separating the substantive assumptions and the statistical assumptions at the very beginning of modeling (Mayo, 1996; Spanos, 1986). The substantive assumptions pertain to the realisticness issue, while the statistical assumptions pertain to the (statistical) reliability of inference.

They further clarify the meaning of the reliability of inference by claiming that the unreliability of inference appears in the difference between the nominal error probability (error probability under the chosen model) and the actual error probabilities (error probability under the correctly specified model). In other words, symptom of the unreliability of inference is seen in the error probabilities of the estimates of the parameter and not necessarily in the estimates themselves.

Spanos (2009) argues that the problem arises when one passes off the substantive information on data by estimating the structural model directly. His suggested approach is to separate the structural model and the statistical model so that at the end one can choose the parametrization to render the structural model a special case of (restriction on) the statistical model. This strategy allows one to evaluate the theory by testing the overidentifying restrictions. Hendry and Mizon (1990) argue that one can evaluate dynamic econometric models by encompassing the VAR. Franchi and Jusélius (2007) propose the cointegrated VAR (CVAR) as a way to matching the assumptions of the structural model with the statistical properties of the data.

In order to test the reliability of PET models such as DSGE model, Spanos (2009) proposes to derive the implicit statistical model (reduced form) of the original structural model and explicitly state its own implicit probabilistic assumptions. If the assumptions are valid for the given data, the model is statistically adequate and further inference procedures like
testing, prediction and goodness-of-fit can be considered reliable. If not, then one needs to respecify (select a new model) with an appropriate probabilistic structure that would render the given data a typical realization thereof. The respecification has to be done to the whole not the part of the model because different assumptions are interrelated. If one of the assumptions is invalid, the whole model is rejected and a completely new model has to be respecified. Spanos (2009) remarks that this process of getting a statistically adequate model might require several iterations before such a model is reached. Spanos (1986, 1989) calls it the Probabilistic Reduction (PR) approach.

The PR approach is completely different from the ‘error-fixing’ approach such as generalized least squares (GLS). Spanos and McGuirk (2001) strongly oppose traditional error-fixing strategies, such as error-autocorrelation correction and heteroskedasticity/autocorrelation consistent standard error, because these strategies make statistical unreliability worse, not better. For example, under the GLS approach, if the assumption of homoskedasticity is found invalid, the solution targets only this particular problem as if the assumption of homoskedasticity is independent of the other set of assumptions. Spanos (1995) argues that the ‘error-fixing’ strategy ignores the potential misspecification in the systematic component of the model and it ignores alternative theories which might fit the same data equally well or even better.

In Chapter 2, the PR approach is discussed in detail and used in the context of Ireland (2011) DSGE macroeconomic model. Its statistical model is revealed first and its statistical adequacy is checked for the given data. Then in Chapter 4, a new model is respecified in the light of the misspecification tests and the exploratory data analysis (EDA) done.

Apart from DSGE models, in Chapter 5, the capital asset pricing model (CAPM) is embedded in the multivariate dynamic Student’s $t$ regression model, which is a reparameterization of the Student’s $t$ VAR model. It will be shown that the CAPM model is highly misspecified with respect to financial data on asset prices of major companies. But the respecified mul-
tivariate dynamic Student’s $t$ regression model turns out to be statistically adequate with respect to the data.

In the next section, the realisticness of the DSGE models are reviewed briefly. Although it is inappropriate to evaluate the realisticness issue unless the reliability of statistical inference is guaranteed by postulating a statistically adequate model, it is important to evaluate the validity of substantive assumptions of the DSGE models.

1.6 Realisticness of DSGE Models

Most of the literature since Kydland and Prescott (1982) has been focused on improving the realisticness of a macroeconomic model – which is the most difficult part in the PR approach – while ignoring the reliability issue completely. As a result, most of the highly unrealistic substantive assumptions have remained unchecked. The main substantive assumptions of the DSGE models and their realisticness are discussed below.

The DSGE model is the latest culmination of the developments in the Walrasian tradition (Colander, 2006). According to Colander (2006), the most critical question that the Walrasians attempt to answer is: how the market coordinates agents with high level information processing capabilities living in information rich environments. Hoover (2006) draws the similar conclusion by saying that the Walrasian attitude is: one must know everything to know anything. To depict the limitation of the DSGE modeling approach in particular and the Walrasian approach in general, both of the critics highlight how unrealistic the macroeconomic environment should be for the DSGE models to be used to explain the real economic situation.

In order to achieve the unique (or a minimum number of) stable equilibrium for the dynamic
optimization problem, the DSGE models invoke the following basic substantive assumptions:

(a) Individual agents, with an unlimited life span, have all the information and process it instantaneously and accurately.

(b) There exists a representative agent who represents all the individuals in the macroeconomic system.

(c) Individual agents (economic decision makers such as consumers and producers) maximize their objective function subject to technological and resource constraints of various forms.

(d) The fluctuations in the economy are interpreted as equilibrium outcomes.

There are three vital problems with the assumptions (a) and (b) for macroeconomic analysis. First, the representative agent assumption does not allow for interaction among agents making the agents independent. If this assumption of independent representative agents is invalid, then the whole aggregation argument fails. Talking about the lack of agent interdependence in these types of models, Aoki’s (2006) argues that the DSGE models are still too simplistic to be taken seriously for policy use.

Second, the result of aggregation becomes even more misleading, if one considers the heterogeneity of agents. Third, the individual agents do not have perfect information and live a finite life in an uncertain environment. This argument is also against the rational expectation hypothesis.

The problem with the assumption (c) is that the objective function maximized by individual agent is in terms of the conditional expectation. Moreover, it is also assumed that the conditional expectations are unbiased and minimize the forecast mean-squared error, which is disproved by Hendry and Mizon (2010). Whenever there is a structural change there is a problem for economic theory-led models, such as DSGE models (Hall, 1978; Hendry and
Mizon, 2010). They further argue that the DSGE models cannot be truly structural on theoretical grounds, despite their micro-foundations. Moreover, one can easily question the validity of the optimizing behaviour of the agents in real life circumstances. Assumption (d) has been challenged by a research program on integrated Keynesian disequilibrium dynamics (Chiarella and Flaschel, 2000). This research program has focused on behavioural equations and dynamic adjustment mechanism along Keynesian lines.

Dutt and Skott (2008) argue that the large part of what has happened in macroeconomics since 1960s has been a wasteful detour. Macroeconomists have developed the sophisticated tools, but the usefulness of these tools is questionable. Moreover, they argue that a great deal of damage may have already been done when the tools are used at the policy-making level. They point to the following four arguments against macroeconomics based on the optimization:

(a) The cognitive limitations and bounded rationality of all real world decision makers do not support the idea of optimizing individual behaviour. Kahneman (2000) and Camerer et al. (2004) have documented the existence of systematic departures from optimizing behaviour.

(b) Even if all individual agents were fully rational and maximized a well-behaved utility function subject to the standard constraints, the aggregate variables do not behave as if determined by an optimizing representative agent (Kirman, 1992).

(c) The existence of social norms and conventions provides a further reason to abstain from the mechanical application of the optimization method based on exogenously given and constant preferences. Norms and conventions change over time, both endogenously and as a result of exogenous shocks.

(d) The gains from the explicit optimization problem are often less than the costs of the
required simplifications in other areas of modeling. Macroeconomist have struggled to solve the problems of intertemporal optimization, which grossly simplify real world decision problems. At the same time, it is implicitly assumed that the agents in the real world are capable of solving these much more complex decisions.

1.7 Bayesian Inference on DSGE Models

The Bayesian approach has been widely used to estimate the DSGE macroeconomic models, because it is supposed to enhance both the realisticness of the model and the reliability of inference. In this section, the Bayesian approach is critically reviewed in brief. More and more economists prefer the Bayesian approach to the classical approach to estimate a DSGE model for various reasons. Fernández-Villaverde (2010) lists the arguments in favor of the Bayesian approach, which are summarized below:

(a) The use of the classical approach to a relatively small sample size (which is almost always the case for macroeconomic time series data) often gives model results that are inconsistent with the macroeconomists’ view of the functioning of the economy.

(b) In the classical approach one has to maximize a complicated and highly dimensional function such as the likelihood of a DSGE model. In contrast, the Bayesian approach relies on the integration, which is easier than the maximization.

(c) The classical approach needs specific methods to address the issues such as non-stationarity. But this is not the case in the Bayesian approach (Sims and Uhlig, 1991). In other words, the Bayesian approach provides better results even under misspecification.

(d) The pre-sample information is often very rich and considerably useful. The classical approach of inference does not have a mechanism to take the pre-sample information
properly. But in the Bayesian approach, the pre-sample information can be integrated easily.

(e) The classical approach has a difficult time moving from point estimates to whole distributions of policy-relevant parameters. The Bayesian approach provides the whole distribution of the parameter estimates.

(f) The Bayesian approach provides a middle ground between the classical approach and the calibration approach.

Now each of these arguments in favor of the Bayesian approach is discussed critically in the following paragraphs. Argument made in (a) relates to lack of sufficient data information. In the classical approach, even relatively small sample size can give relatively precise inferences if the model is correctly specified, especially for independently and identically distributed data. An attempt has been made to argue that the views of macroeconomists are correct a priori and the data must support it. This argument is completely against the norm of science, where views are updated on the basis of new evidence or observation.

Argument in (b) is very weak because easiness and simplicity do not have any relation with the reliability of inference. Sacrificing the reliability for easiness is costly in terms of drawing reliable inference.

Argument in (c) ignores the fact that the likelihood function in both the classical and Bayesian approaches is heavily affected by assumptions such as stationarity and independence. However, the Bayesian approach is vulnerable to misspecification from one more dimension – misspecification of the prior.

Argument (d) assumes that the pre-sample (mostly theory determined) information is true by definition and has to be incorporated before the data has been used. But this leads to unreliable inference as there is no room left for testing the prior information. Moreover,
the Bayesian approach requires one to know the prior information in terms of the whole distribution of the parameters, which is rarely available in practice. But, on the other hand the classical approach can handle these issues using proper Neyman-Pearson testing within a statistically adequate model.

The argument made in (f) is the weakest of all. The main difference between the classical approach and the Bayesian approach is not methodological and technical, but philosophical. There can be no cases where the Bayesian approach is a special case of the classical approach, the calibration technique or a combination of them.

Hence, choosing the Bayesian approach in place of the classical approach does not ameliorate the problem of the reliability of inference. Using the Bayesian approach to statistically analyse the DSGE models adds significantly to the untrustworthiness of the resulting evidence by glossing over the statistical misspecification. Even by ignoring the philosophical and methodological issues of the Bayesian approach, the reliability of inference crucially relies on the adequacy of the statistical model chosen. The posterior distribution is nothing but a product of the likelihood function and the prior. If the chosen model is misspecified, the likelihood is going to be wrong and the whole posterior will be invalid. Using the Bayesian approach on the ground of simplicity and computational advantage cannot be considered a valid reason.

If the primary aim of macroeconomic modeling is to learn from the data about the economic phenomena of interest, statistical adequacy is the necessary precondition for securing the reliability of the statistical inference procedures, irrespective of whether one uses the classical or the Bayesian approach, because both depend on the validity of the statistical model in question. That is, for reliable probing of substantive questions of interest, statistical adequacy is a must!
1.8 Concluding Remark

If the statistical model does not allow for the restrictions on the parameter space defined by
the economic model, either the economic model is inadequate (misspecified) for explaining
features of data or the quality, quantity or both of the data itself is not sufficient enough.
If the purpose of modeling is not inference on the model equations and the parameters, but
to use the model (e.g. for prediction or policy simulation) that we have a firm faith on, then
the consistency of the estimators is not guaranteed and the standard errors of the estima-
tors might be statistically meaningless. This approach impedes further development in the
field because one does not need to face any counter evidence and hence it treats inference as
not falsifiable. Development of new models in the light of new evidence becomes unnecessary.

1.9 A Brief Overview of the Chapters

In Chapter 2, Ireland’s (2011) structural model is briefly explained. The ‘reduced form’ in
the form of a VAR model is derived as a statistical generating mechanism of the statistical
model. The statistical model is estimated without any parameter restrictions. A thorough
misspecification testing is conducted to find possible misspecification.
In Chapter 3, the multivariate and the matrix variate Student’s t distributions are discussed
in detail to show that the latter is a special case of the former. Various models emerging from
the multivariate Student’s t distribution are specified for both stationary and non-stationary
processes.
In Chapter 4, the Student’s t VAR model is used as a respecified model and subjected to the
M-S tests to explore its statistical adequacy for the data. The results of the Student’s t VAR
model is compared with that of the Normal VAR model in terms of coefficients, significance, sign, impulse response function (IRF), fitted values and residuals.

In Chapter 5, the CAPM model is revisited from the PR perspective. In Chapter 6, concluding remarks are made and the implications of the research for future projects are discussed.
Chapter 2

DSGE Modeling: A Probability Reduction Approach

2.1 Introduction

The most critical step to ensure the reliability of inference of a DSGE model is to test its statistical adequacy for the data being used to estimate the model. In this chapter, Ireland’s (2011) theoretical DSGE model is first converted into a structural model with a view to render it estimable in light of the available data. The structural model is then embedded into a statistical model by making the implicit probabilistic assumptions concerning the observable processes involved explicit. In doing so, the probability reduction (PR) approach, as suggested by Spanos (1986), is followed. In this empirical modeling approach, observed data $Z_0$ is viewed as a typical realization of the stochastic process $\{Z_t, t \in \mathbb{N}\}; \mathbb{N} = \{1, 2, 3, \ldots \}$, whose probabilistic structure is modeled in terms of the joint distribution of observables. The statistical model in question is viewed as a parametrization of the underlying process $\{Z_t, t \in \mathbb{N}\}$.

A key methodological debate in current macroeconometric modeling is between “theory first” and “data first” perspectives; see Spanos (2010) for extensive discussion on comparative analysis of two perspectives in macroeconometric modeling. Spanos (1986) proposed an all encompassing methodological framework for empirical modeling (Figure 2.1). He argues
that this framework is devised to enable the modeler to bridge the gap between theory and
data using a sequence of interconnected models with a view to delineate and probe for the
potential errors at different stages of modeling.

Spanos (2010) argues that estimating the structural model $\mathcal{M}_\psi(z)$ directly imposes the
substantive information on data and leaves no room for testing it against the data because the
estimated structural model $\hat{\mathcal{M}}_\psi(z)$ is almost never substantively or statistically adequate.
The traditional approach of estimating the structural model directly does not allow one to
delineate between two sources of error: is the theory wrong or are the probabilistic assump-
tions implicitly imposed in the form of a statistical model invalid for the given data $Z_0$? To
solve the problem, one approach is to separate the substantive and the statistical information
once the data $Z_0$ is chosen (Spanos, 1986). This enables one to delineate and probe for the
potential errors at different stages of modeling. This way of bridging theory and data can
be viewed as based on the following two pillars of the PR approach

(a) From the theory side, a theory model is transformed into a structural model $\mathcal{M}_\psi(z)$ to
    render it estimable with the data $Z_0$.

(b) From the data side, the statistical information is carried by a statistical model $\mathcal{M}_\theta(z)$
    whose parametrization is chosen to nest $\mathcal{M}_\psi(z)$ parametrically in the context of the
    statistical model.

Once a statistically adequate model $\mathcal{M}_\theta(z)$ is established, it can be related to the struc-
tural model $\mathcal{M}_\psi(z)$ via an implicit function $G(\psi, \theta) = 0$, where $\psi \in \Phi$ and $\theta \in \Theta$ are
the structural and the statistical parameters respectively. Often the statistical model has
more parameters than the structural model, which enables one to test the overidentifying
restrictions

$$H_0 : G(\psi, \theta) = 0, \text{ vs. } H_1 : G(\psi, \theta) \neq 0$$
Figure 2.1: Probability Reduction (PR) Approach
where, $H_0$ and $H_1$ stand for the null hypothesis and the alternative hypothesis respectively in the Neyman-Pearson setting.

The rejection of the $H_0$ provides evidence against the empirical adequacy of the structural model for the given data $Z_0$. This way of identification enables one to carry out the statistical analysis without any reference to the structural model once a statistically adequate model is reached (Spanos, 1990). This identification, known as statistical identification, differs from the traditional notion of identification (structural identification), where it is defined as pertaining to the conditions ensuring solving the mapping $G(\psi, \theta) = 0$, uniquely for $\psi$, taking the statistical parameters $\theta$ as given. Consolo, Favero and Paccagnini (2009) discuss extensively about the difference between the structural identification and the statistical identification with special attention to the DSGE models. The structural identification ignores the fact that $\theta$ is well-defined only when the hypothetical reduced form is statistically adequate. The PR approach requires a sound link between the data and the statistical model by securing the statistical adequacy of the implicit statistical model before solving the implicit function $G(\psi, \theta) = 0$ to identify $\psi$.

There are many versions of the DSGE macroeconomic models in the literature. But Ireland’s (2011) model is a typical example. His model is used to explain how the PR approach can be applied. A brief review of his structural model is presented below before deriving its implicit statistical model.

### 2.2 Ireland’s Structural Model

In this model, there are three economic agents and one central bank in the economy. The three agents are (a) a representative household, (b) the intermediate goods-producing firms and (c) a representative final goods-producing firm. The household maximizes its expected
life time utility subject to its intertemporal budget constraints. The firms maximize their expected life time profit subject to the technology constraint they face.

2.2.1 Representative Household

The representative household enters each time period $t = 0, 1, 2, 3...$ with a total bond $B_{t-1}$, money $M_{t-1}$ and a lump sum monetary transfer of $T_t$ from the central bank. $B_t/r_t$ is the price of new bond where $r_t$ is the gross interest rate between $t$ and $t+1$. In the period $t$, the household supplies $l_t$ units of labor to different intermediate goods-producing firms thereby earning $W_t l_t$, where $W_t$ is the nominal wage rate for the time period. The household purchases $C_t$ units of the finished goods at a price $P_t$ from the representative final goods-producing firm. At the end of the period $t$, the household also receives a nominal profit share

$$D_t = \int_0^1 D_t(i)di$$

from the intermediate goods producing firms $i\in[0, 1]$, where $D_t(i)$ is the profit earned from the firm $i$ at the time period $t$. Hence the intertemporal budget constraint faced by the representative household for the periods $t = 0, 1, 2, 3,...$ is

$$M_{t-1} + B_{t-1} + T_t + W_t l_t + D_t \geq P_t C_t + B_t/r_t + M_t$$

(2.1)

Subject to the budget constraint (2.1), the representative household maximizes the life time infinite horizon expected utility function

$$E_0 \sum_{t=0}^{\infty} \beta^t [a_t \ln(C_t - \gamma C_{t-1}) + \ln(M_t/P_t) - l_t]$$

(2.2)
where $0 < \beta < 1$, $0 \leq \gamma < 1$ and the preference shock $a_t$ follows

\[
\ln(a_t) = \rho_a \ln(a_{t-1}) + \varepsilon_{at}
\]  

(2.3)

with $0 \leq \rho_a < 1$. Now the household chooses $C_t, l_t, B_t$ and $M_t$, \( \forall \ t = 0, 1, 2, ... \). The first order condition of the optimization problem (2.1)-(2.3) is given by the budget constraint (2.1) itself with an equality sign and the following equations

\[
\Lambda_t = \frac{a_t}{C_t - \gamma C_{t-1}} - \beta \gamma E_t \left( \frac{a_{t+1}}{C_{t+1} - \gamma C_t} \right)
\]  

(2.4)

\[
a_t = \Lambda_t (W_t / P_t)
\]  

(2.5)

\[
\Lambda_t = \beta r_t E_t (\Lambda_{t+1} / \pi_{t+1})
\]  

(2.6)

\[
\frac{M_t}{P_t} = \left( \frac{a_t}{\Lambda_t} \right) [r_t / (r_t - 1)]
\]  

(2.7)

where $\Lambda_t \geq 0$ is the Lagrange multiplier and $p_t = P_t / P_{t-1}$ is the gross inflation rate from time $t - 1$ to $t$. The optimality condition of money holding (2.7) is dropped by Ireland (2011), because this condition serves only to determine how much money the central bank needs to supply to clear the markets given its interest rate target $r_t$.

### 2.2.2 Finished Goods-producing Firm

The finished goods-producing firm buys $Y_t(i)$, $i \in [0, 1]$ units of the intermediate good at the price $P_t(i)$ to produce $Y_t$ units of the finished goods using the constant-returns-to-scale technology described by

\[
\left[ \int_0^1 Y_t(i)^{(\theta_t-1)/\theta_t} \, di \right]^{\theta_t/(1-\theta_t)} \geq Y_t
\]  

(2.8)
where $\theta_t$ is the elasticity of demand for the intermediate goods, which follows the autoregressive process

$$\ln(\theta_t) = (1 - \rho_{\theta})\ln(\overline{\theta}) + \rho_{\theta}\ln(\theta_{t-1}) + \varepsilon_{\theta t}$$

(2.9)

where $\overline{\theta} > 1$ and $1 > \rho_{\theta} \geq 0$.

The profit maximization results into the input-price combination of

$$Y_t(i) = \left[ P_t(i)/P_t \right]^{-\theta_t} Y_t$$

(2.10)

$$P_t = \left[ \int_0^1 P_t(i)^{1-\theta_t} di \right]$$

(2.11)

for all $t = 0, 1, 2, ...$.

### 2.2.3 Intermediate Goods-producing Firm

The representative intermediate goods-producing firm $i$ hires $l_t(i)$ units of labor from the representative household to manufacture $Y_t(i)$ units of the intermediate good $i$ according to the constant-returns-to-scale technology

$$H_t l_t(i) \geq Y_t(i)$$

(2.12)

where $H_t$ is a technology shock which follows the random walk with drift process

$$\ln(H_t) = \ln(\overline{h}) + \ln(H_{t-1}) + \varepsilon_{ht},$$

(2.13)

with $\overline{h} > 1$. The intermediate goods-producing firms sell their product in a monopolistically competitive market where they face the cost of nominal price adjustment in terms of the
finished goods as follows (Rotemberg, 1982)

$$\phi \left[ \frac{P_t(i)}{p_{t-1}^{1-\alpha}P_{t-1}(i)} - 1 \right]^2 Y_t, \quad (2.14)$$

where $\phi \leq 0$ measures the magnitude of the price adjustment cost and the gross steady state inflation is $\bar{p} \geq 1$. This quadratic cost function makes the intermediate goods-producing firms' problem dynamic. If $\alpha = 0$, the firms are fully forward looking (Ireland, 2004). If $\alpha = 1$, the firms are fully backward looking. Now, the firms choose $P_t(i)$ so as to maximize

$$E_0 \sum_{t=0}^{\infty} \beta^t \Lambda_t [D_t(i)/P_t], \quad (2.15)$$

where $D_t(i)/P_t$ is the real profit defined as

$$\frac{D_t(i)}{P_t} = \left[ \frac{P_t(i)}{P_t} \right]^{1-\theta_t} Y_t - \left[ \frac{P_t(i)}{P_t} \right]^{-\theta_t} \left( \frac{W_t}{P_t} \right) \left( \frac{Y_t}{Z_t} \right) - \frac{\phi}{2} \left[ \frac{P_t(i)}{p_{t-1}^{1-\alpha}P_{t-1}(i)} - 1 \right]^2 Y_t \quad (2.16)$$

The first order condition of this problem is given by

$$\left( \theta_t - 1 \right) \left[ \frac{P_t(i)}{P_t} \right]^{-\theta_t} = \theta_t \left[ \frac{P_t(i)}{P_t} \right]^{1-\theta_t} \left( \frac{W_t}{P_tZ_t} \right) - \phi \left[ \frac{P_t(i)}{p_{t-1}^{1-\alpha}P_{t-1}(i)} - 1 \right] \left[ \frac{P_t(i)}{p_{t-1}^{1-\alpha}P_{t-1}(i)} \right]$$

$$+ \beta \phi E_t \left\{ \left( \frac{\Lambda_{t+1}}{\Lambda_t} \right) \left[ \frac{P_{t+1}(i)}{p_{t}^{1-\alpha}P_t(i)} - 1 \right] \left[ \frac{P_tY_{t+1}}{Y_tP_t(i)} \right] \left[ \frac{P_{t+1}(i)}{p_{t}^{1-\alpha}P_t(i)} \right] \right\} \quad (2.17)$$

and (2.12) with equality for all $t = 0, 1, 2, ...$.
2.2.4 Central Bank

The central bank conducts the monetary policy following a variant of the Taylor (1993) rule

\[\ln r_t - \ln r_{t-1} = \rho_p \ln(p_t/\bar{p}) + \rho_y \ln(y_t/\bar{y}) + \varepsilon_{rt}\] (2.18)

where \(\rho_p, \rho_y \in \mathbb{R}\), \(\bar{p}\) and \(\bar{y}\) are steady state (average) values of \(p_t\) and \(y_t\), respectively. It is important to note that this rule is derived under the assumption of constant variance of inflation rate and growth rate. In this chapter, this assumption is proved to be invalid for the US data.

2.2.5 Efficient Allocation and Output Gap

The efficient allocation of resources can be thought from the perspective of a benevolent social planner who allocates \(n_t(i)\) unit of the household labor to produce the efficient level of output \(Q_t\) and allocates the production of each intermediate good \(i \in [0, 1]\) all according to the constant-returns-to-scale technology. Hence the social planner faces the problem of maximizing the representative household’s welfare

\[E \sum_{t=0}^{\infty} \beta^t \left\{ a_t \ln(Q_t - \gamma Q_{t-1}) - \int_0^1 n_t(i) di \right\} \] (2.20)

subject to the feasibility constraints (based on (2.8) and (2.12))

\[H_t \left[ \int_0^1 n_t(i)^{(\theta_t-1)/\theta_t} \right]^{\theta_t/(\theta_t-1)} \geq Q_t\] (2.21)
The first order condition to this problem define the efficient level of output $Q_t$ as

$$\Xi_t = \frac{a_t}{Q_t - \gamma Q_t} - \beta \gamma E_t \left( \frac{a_{t+1}}{Q_{t+1} - \gamma Q_t} \right), \quad (2.22)$$

the efficient level of labor $n_t(i)$ as

$$a_t = \Xi_t H_t (Q_t/H_t)^{1/\theta_t} n_t(i)^{-1/\theta_t} \quad (2.23)$$

for $i \in [0, 1]$ and (2.21) with an equality sign for all $t = 0, 1, 2, ...$. Here $\Xi_t$ is a non negative Lagrange multiplier on the aggregate feasibility constraint for the period $t$. Equation (2.23) implies that $n_t(i) = n_t$ for all $i \in [0, 1]$ and $t = 0, 1, 2, ...$ giving rise to

$$n_t = (\Xi_t/a_t)^{\theta_t} H_t^{\theta_t} (Q_t/H_t) \quad (2.24)$$

Substituting this equation (2.24) in (2.21), we get

$$\Xi_t = a_t/H_t \quad (2.25)$$

for all $t = 0, 1, 2, ...$

Using (2.22), we get

$$\frac{1}{H_t} = \frac{1}{Q_t - \gamma Q_{t-1}} - \beta \gamma E_t \left[ \left( \frac{a_{t+1}}{a_t} \right) \left( \frac{1}{Q_{t+1} - \gamma Q_t} \right) \right] \quad (2.26)$$

for all $t = 0, 1, 2, ...$. Now we can define the output gap as a ratio between the actual output and the efficient output level

$$g_t = \frac{Y_t}{Q_t} \quad (2.27)$$

for $t = 0, 1, 2, ...$
2.2.6 Symmetric Equilibrium

In the symmetric equilibrium, all the intermediate goods-producing firms make identical decisions so that $Y_t(i) = Y_t, l_t(i) = l_t, P_t(i) = P_t, D_t(i) = D_t$ for all $i \in [0, 1]$ and $t = 0, 1, 2, ...$.

Market clearing requires $M_t = M_{t-1} + T_t$ for the money and $B_t = B_{t-1} = 0$ for the bonds for all $t = 0, 1, 2, ...$. With this, equations (2.5), (2.7), (2.12) and (2.16) can be solved to get $W_t, M_t, l_t$ and $D_t$. Here, the representative household’s budget constraint will turn into

$$Y_t = C_t + \frac{\phi}{2} \left( \frac{P_t}{\bar{p}} - 1 \right)^2 Y_t$$  \hspace{1cm} (2.28)

2.2.7 Linearized Model

Now we have 11 equations (2.1), (2.3), (2.4), (2.6), (2.9), (2.13), (2.17)-(2.19), (2.26) and (2.27) in 11 variables $Y_t, C_t, p_t, r_t, y_t, Q_t, g_t, \Lambda_t, a_t, \theta_t$ and $H_t$. Among these 11 variables the last three are shocks and the rest are endogenous variables. Although $Y_t$ and $C_t$ both inherit a root process at equilibrium, the stochastically detrended variables $y_{ht} = Y_t/H_t, c_t = C_t/H_t, q_t = Q_t/H_t, \lambda_t = H_t\Lambda_t, h_t = H_t/H_{t-1}$ remain stationary. In the steady state, the stationary variables are constant over time in the absence of any shocks

$$y_t = \bar{y}, \hspace{0.5cm} c_t = \bar{c}, \hspace{0.5cm} p_t = \bar{p}, \hspace{0.5cm} r_t = \bar{r}, \hspace{0.5cm} y_{ht} = \bar{y}_h, \hspace{0.5cm} g_t = \bar{g}, \hspace{0.5cm} \lambda_t = \bar{\lambda}, \hspace{0.5cm} q_t = \bar{q}, \hspace{0.5cm} a_t = 1, \hspace{0.5cm} \theta_t = \bar{\theta}, \hspace{0.5cm} h_t = \bar{h}.$$

Let, $\hat{y}_{ht} = \ln(y_{ht}/\bar{y}_h), \hat{c}_t = \ln(c_t/\bar{c}), \hat{p}_t = \ln(p_t/\bar{p}), \hat{r}_t = \ln(r_t/\bar{r}), \hat{g}_t = \ln(g_t/\bar{g}), \hat{y}_t = \ln(y_t/\bar{y}), \hat{\lambda}_t = \ln(\lambda_t/\bar{\lambda}), \hat{q}_t = \ln(q_t/\bar{q}), \hat{a}_t = \ln(a_t), \hat{\theta}_t = \ln(\theta_t/\bar{\theta})$, and $\hat{h}_t = \ln(h_t/\bar{h})$ represent the percentage deviation of each variable from its steady state level. The first-order Taylor expansion to the budget constraint (2.28) implies

$$\hat{c}_t = \hat{y}_{ht}$$  \hspace{1cm} (2.29)
and the rest of the model becomes

\[ \hat{a}_t = \rho_a \hat{a}_{t-1} + \varepsilon_{at} \]  

\[ (\bar{h} - \beta \gamma)(\bar{h} - \gamma) \hat{\lambda}_t = \gamma \bar{h} \tilde{y}_{ht-1} - (\bar{h}^2 + \beta \gamma^2) \tilde{y}_{ht} + \beta \gamma \bar{h} E_t \tilde{y}_{ht+1} + (\bar{h} - \beta \gamma \rho_a) (\bar{h} - \gamma) \hat{a}_t - \gamma \bar{h} \tilde{h}_t \]  

\[ \hat{\lambda}_t = \hat{r}_t + E_t \hat{\lambda}_t - E_t \hat{p}_{t+1} \]  

\[ \hat{e}_t = \rho_e \hat{e}_{t-1} + \varepsilon_{et} \]  

\[ \hat{h}_t = \varepsilon_{ht} \]  

\[ (1 + \beta \alpha) \hat{p}_t = \alpha \hat{p}_{t-1} + \beta E_t \hat{p}_{t+1} - \psi \hat{\lambda}_t + \psi \hat{a}_t + \hat{e}_t \]  

\[ \hat{r}_t - \hat{r}_{t-1} = \rho_p \hat{p}_t + \rho_y \hat{y}_t + \varepsilon_{rt} \]  

\[ \hat{y}_t = \hat{y}_{ht} - \hat{y}_{ht-1} + \hat{h}_t, \]  

\[ 0 = \gamma \bar{h} \hat{q}_{t-1} - (\bar{h}^2 + \beta \gamma^2) \hat{q}_t + \beta \gamma \bar{h} E_t \hat{q}_{t+1} + \beta \gamma (\bar{h} - \gamma) (1 - \rho_a) \hat{a}_t - \gamma \bar{h} \tilde{h}_t \]  

and

\[ \hat{q}_t = \hat{y}_{ht} - \hat{g}_t \]  

for all \( t = 0, 1, 2, \ldots \) where \( \psi = (\bar{\theta} - 1)/\phi \) and \( \hat{e}_t = -(1/\phi) \hat{\theta}_t, \rho_e = \rho_a, \text{ var}(\varepsilon_{et}) = (1/\phi)^2 \sigma^2_{\theta}. \)

By transforming \( \hat{\theta}_t \) into \( \hat{e}_t \), estimation of \( \theta \) is avoided before estimating the whole model.

Finally, the error terms \( \varepsilon_{at}, \varepsilon_{\theta t}, \varepsilon_{ht}, \varepsilon_{rt} \) are assumed to be jointly Normal, independently and
identically distributed

\[
\begin{pmatrix}
\varepsilon_{at} \\
\varepsilon_{\theta t} \\
\varepsilon_{ht} \\
\varepsilon_{rt}
\end{pmatrix}
\sim NIID
\begin{pmatrix}
0 \\
0 \\
0 \\
0
\end{pmatrix}
\begin{pmatrix}
\sigma^2_a & 0 & 0 & 0 \\
0 & \sigma^2_{\theta} & 0 & 0 \\
0 & 0 & \sigma^2_h & 0 \\
0 & 0 & 0 & \sigma^2_r
\end{pmatrix}
\]  

(2.40)

For the reasons that are going to be apparent later, this assumption is crucial for being able to estimate the model using the maximum likelihood method. Later in this chapter, this assumption will be challenged by the M-S tests for the data used for estimation. Violation of this assumption by the data severely undermines the reliability of inference to be drawn from the model.

After linearization, Ireland (2011) DSGE theory-model \( M_\psi(z; \xi; \epsilon) \), is specified in Table 2.1 in terms of

(a) observables: \((Y_t, P_t, r_t)\), \(Y_t\)-production, \(P_t\)-price level, and \(r_t\)-gross interest rate,

(b) latent variables: \( \xi_t = (Q_t, g_t, \Lambda_t) \), \(Q_t\)-efficient output, \(g_t = Y_t / Q_t\)-output gap, \(\Lambda_t\)-Lagrange multiplier,

(c) latent shocks: \( \epsilon_t = (a_t, \theta_t, H_t) \), \(a_t\)-preference, \(\theta_t\)-demand, \(H_t\)-technology.

Now a DSGE model can be defined as follows. DSGE models aim to describe the behavior of the economy in an equilibrium steady state stemming from optimal microeconomic decisions associated with several agents (households, firms, governments, central banks). It is essentially deterministic theory-model in the form of a system of first order difference equations, but driven by latent stochastic (autocorrelated) shocks; see Canova (2007).
Table 2.1: Dynamic Stochastic General Equilibrium (DSGE) Model

Behavioral equations:
(i) \( \ln\left(\frac{Y_t}{Q_t}\right) = \ln\left(\frac{Y_t}{H_t}\right) - \ln\left(\frac{Q_t}{H_t}\right) \)
(ii) \( \ln\left(\frac{P_t}{P_{t-1}}\right) = \alpha + \beta \ln\left(\frac{P_{t-1}}{P_{t-2}}\right) + \gamma \ln\left(\frac{P_{t-1}}{P_{t-2}}\right) E_t \ln\left(\frac{P_{t+1}}{P_t}\right) - \frac{\psi}{1+\beta} \ln\left(\frac{H_t}{Y_t}\right) + \frac{\psi}{1+\beta} \alpha t + \frac{1}{1+\beta} \beta t \)
(iii) \( \ln\left(\frac{r_t}{r_{t-1}}\right) = \rho_p \ln\left(\frac{r_t}{r_{t-1}}\right) + \rho_y \ln\left(\frac{Y_t}{Y_{t-1}}\right) + \varepsilon_{rt} \)
(iv) \( \ln\left(\frac{H_t}{H_{t-1}}\right) = \ln\left(\frac{H_t}{Y_t}\right) - \ln\left(\frac{Y_t}{Y_{t-1}}\right) + \ln\left(\frac{H_t}{H_{t-1}}\right) \)
(v) \( \ln\left(\frac{Y_t}{Y_{t-1}}\right) = \ln\left(\frac{Y_t}{H_t}\right) - \ln\left(\frac{Y_{t-1}}{H_{t-1}}\right) + \ln\left(\frac{Y_t}{Y_{t-1}}\right) \)

Shocks:
(viii) \( \ln a_t = \rho a_t + \ln a_{t-1} + \varepsilon_{at} \)  
(ix) \( \ln\left(\frac{\theta_t}{\varphi_t}\right) = \ln\left(\frac{\theta_t}{\varphi_t}\right) + \ln\left(\frac{\theta_{t-1}}{\varphi_{t-1}}\right) \)

Parameters: \( \varphi := (\alpha, \beta, \gamma, \psi, \rho_p, \rho_o, \rho_y, \rho_{\theta}, \sigma^2_{\alpha}, \sigma^2_{\beta}, \sigma^2_{\gamma}, \sigma^2_{\psi}) \)
2.2.8 Identities of the Model

As pointed out by Spanos (2012), this model has a high proportion of identities (2.29, 2.34, 2.37 and 2.39) in comparison to the genuine behavioural equations. The identity (2.29) is the household’s budget constraint and the identity (2.39) is derived from the definition of the output gap. These identities are void of any substantive information. Spanos (2012) suggests to eliminate the accounting identities before estimating the structural model, and replace them with the genuine behavioural equations for the variables in question. He further suggests to use the adjustment equations (in the form of difference equations) in place of the equilibrium conditions such as (2.29). It should be noted that the solution of such adjustment equations in the form of the difference equations are functions of time. This can help to address the lack of heterogeneity in the DSGE models. This also raises an important question pertaining to the realisticness of the model. For the current research, however, we ignore this aspect of the macroeconomic modeling so that we can focus on assessing the reliability of this particular DSGE model.

2.2.9 Solution of the Model

The total number of the endogenous variables is seven: ($\hat{y}_{ht}$, $\hat{y}_t$, $\hat{p}_t$, $\hat{r}_t$, $\hat{q}_t$, $\hat{\lambda}_t$, $\hat{g}_t$). Here $\hat{g}_t$, $\hat{\lambda}_t$, $\hat{q}_t$ and $\hat{y}_{ht}$ are the unobserved variables. Using (2.29),(2.31), (2.32) and (2.35)-(2.39), we get the linear expectational difference equation system as

$$AE_t s_t^0 = Bs_t^0 + Cv_t,$$  \hspace{1cm} (2.41)

with (2.30),(2.33) and (2.34) giving rise to

$$v_t = Pv_{t-1} + \epsilon_t$$  \hspace{1cm} (2.42)
where

$$s_t^0 = \begin{bmatrix} \hat{y}_{t-1} & \hat{p}_{t-1} & \hat{r}_{t-1} & \hat{y}_{ht-1} & \hat{\lambda}_{t-1} & \hat{g}_{t-1} & \hat{p}_t & \hat{\lambda}_t & \hat{y}_{ht} & \hat{g}_t \end{bmatrix}^T,$$

$$v_t = \begin{bmatrix} \hat{a}_t & \hat{e}_t & \hat{h}_t & \varepsilon_{rt} \end{bmatrix}^T, \quad \varepsilon_t = \begin{bmatrix} \varepsilon_{at} & \varepsilon_{et} & \varepsilon_{ht} & \varepsilon_{rt} \end{bmatrix}^T \sim \text{NIID} (0, V),$$

$$A = \begin{bmatrix} 0 & (1 + \beta \alpha) & 0 & 0 & \psi & 0 & -\beta & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & \bar{h}^2 + \beta \gamma^2 & (\bar{h} - \beta \gamma)(\bar{h} - \gamma) & 0 & 0 & 0 & -\beta \gamma z & 0 \\ \rho_y & \rho_p & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\bar{h}^2 - \beta \gamma^2 & 0 & \bar{h}^2 + \beta \gamma^2 & 0 & 0 & \beta \gamma \bar{h} & -\beta \gamma \bar{h} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & \alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma \bar{h}, & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \gamma \bar{h}, & 0 & \gamma \bar{h} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} \psi & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ (\bar{h} - \beta \gamma \rho_a)(\bar{h} - \gamma) & 0 & -\gamma \bar{h} \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ -\beta \gamma (\bar{h} - \gamma)(1 - \rho_a) & 0 & \gamma \bar{h} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$
There are a number of solution methods for the above linearized DSGE model; see Blanchard and Kahn (1980) and Klein (2000) among others. Ireland (2004, 2007, 2011) uses the method developed by Klein (2000) to solve the system (See Appendix A at the end of this chapter for the detail derivation of the solution using the method of Klein (2000)). This method requires the following two assumptions to be valid:

(a) the shock vector $\varepsilon_t$ is an independent and an identical multivariate Normally distributed random vector.

(b) the number of the generalized eigenvalues of the matrix couple $\mathbf{A}$ and $\mathbf{B}$ lying outside the unit circle must be equal to the number of the predetermined variables. Solving the equation $|\mathbf{A} - \lambda \mathbf{B}| = 0$ for $\lambda$ gives the generalized eigenvalues.

Validity of the first assumption is questionable for a typical macroeconomic data. So it will be tested formally later in this chapter. The second assumption relies totally on the parameter restrictions imposed by the structural model. The parameter restrictions imposed will be proved to be invalid in Chapter 4 through the likelihood ratio test, once we have a statistically adequate model which can nest the above structural model as a special case.

Using the method of Klein (2000), the solution of the system (2.41) can be written by combining with (2.42) in the state space form as

$$
\mathbf{s}_{t+1} = \Pi \mathbf{s}_t + \mathbf{W} \varepsilon_{t+1} \quad (2.43)
$$
where, \( s_t = \begin{bmatrix} \hat{y}_{t-1} & \hat{p}_{t-1} & \hat{r}_{t-1} & \hat{a}_{ht-1} & \hat{a}_t & \hat{e}_t & \hat{h}_t & \varepsilon_{rt} \end{bmatrix}^\top \)

\[
\Pi_{(10 \times 10)} = \begin{bmatrix} \Pi_1(6 \times 6) & \Pi_2(6 \times 4) \\ 0_{(4 \times 6)} & P_{(4 \times 4)} \end{bmatrix}
\]

\[
W = \begin{bmatrix} 0_{6 \times 4} \\ I_{(4 \times 4)} \end{bmatrix}
\]

where, \( \Pi_1 \) and \( \Pi_2 \) are functions of \( A, B, C, P \).

### 2.3 Statistical Model

As explained in the introduction of the chapter, the reliability of any inference to be drawn from the directly estimated structural model is unknown until the statistical model underlying the structural model has been proved statistically adequate for the data chosen. The first step is to unveil the statistical model behind the structural model in the form of a complete list of the probabilistic assumptions imposed in the process. Without doing so, testing these assumption against the data becomes a hopeless task. So the reduced form of the structural model (2.43) is derived before turning it into a well-defined statistical model specified in terms of the probabilistic assumptions pertaining to the observable variables involved. The reduced form, here, is defined as a model with all the observable endogenous variables on the right hand side of the system and the rest of the variables on the left hand side. It should be noticed that the reduced form still inherits the structural restrictions imposed by the theory on its parameters. Moving from the reduced form to the statistical model involves the relaxation of these structural parameter restrictions so as to reveal the full statistical parameter space associated with the statistical model.
2.3.1 Reduced Form

The system of equations (2.43) can be decomposed into two subsystems by defining a vector of the observable variables \( d_t = \begin{bmatrix} \hat{y}_t & \hat{p}_t & \hat{r}_t \end{bmatrix}^\top \) and a vector of the unobservable variables \( Y_t = \begin{bmatrix} \hat{y}_{ht} & \hat{\lambda}_t & \hat{g}_t \end{bmatrix}^\top \) as follows

\[
\begin{align*}
    d_t &= DY_{t-1} + Ed_{t-1} + Fv_t \quad (2.44) \\
    Y_t &= GY_{t-1} + Hd_{t-1} + Kv_t \quad (2.45)
\end{align*}
\]

where \( v_t = Pv_{t-1} + \varepsilon_t \) and \( D, E, F, G, H, K \) are formed by the partitioning of \( \Pi \) as follows

\[
\Pi_{(10\times10)} = \begin{bmatrix}
    E_{(3\times3)} & D & F \\
    H & G_{(3\times3)} & K \\
    0 & 0 & 0_{4\times4}
\end{bmatrix}
\]

Assuming that \( D^{-1} \) exists, elimination of \( Y_t \) from (2.44) and (2.45) yields (whenever the usual inverse of the matrix does not exist, the generalized inverse is used following Rao and Mitra (1972). The generalized inverse is the same as the regular inverse when the inverse of the matrix exists)

\[
\begin{align*}
    d_t &= [DG^{-1} + E] d_{t-1} + D(H - GD^{-1}E)d_{t-2} + e_t \quad (2.46) \\
    e_t &= Fv_t + D(K - GD^{-1}F)v_{t-1} \quad (2.47) \\
    v_t &= Pv_{t-1} + \varepsilon_t \quad (2.48)
\end{align*}
\]
Using (2.46), (2.47) and (2.48), we can eliminate \(v_t\) to get

\[
d_t = \Psi_1 d_{t-1} + \Psi_2 d_{t-2} + u_t \tag{2.49}
\]

where,

\[
u_t = \Psi_3 \varepsilon_t
\]

\[
\Psi_1 = DGD^{-1} + E, \quad \Psi_2 = D(H - GD^{-1}E), \quad \Psi_{3(3 \times 4)} = [\Lambda - I]^{-1} \Lambda F;
\]

\[
A_{3 \times 3} = FP(FP + D(K - GD^{-1}F))^{-1} + D(K - GD^{-1}F)(FP + D(K - GD^{-1}F))^{-1}
\]

\[
(u_t|d_{t-1}, d_{t-2}) \sim N(0, \Psi_3 V \Psi_3^T).
\]

We can see that the error vector \(u_t\) is just a linear combination of the various structural shocks of the model. The structural model imposes the restrictions on the coefficients and the covariance matrix of the model (2.49). The issue of the structural identification does not arise at this stage because we have not achieved the statistical identification (Spanos, 1990) and statistical adequacy yet. Spanos (2012) argues that the same structural model can be identified or unidentified depending upon the quality of data. Here, all the unobserved variables are eliminated from the model because the statistical adequacy of a model can be tested only against the observed data. The minimum condition for the DSGE model to be meaningful is that its statistical model in terms of the observables should be statistically adequate via the data on the observables we have.

### 2.3.2 Identification and Estimation

The structural model consisting of equations (2.29)-(2.39) or (2.43) has 12 parameters namely

\[
\psi = \{\alpha, \beta, \psi, \gamma, \rho_y, \rho_p, \rho_a, \rho_e, \sigma_a, \sigma_e, \sigma_h, \sigma_r\}
\]
The parameter space defined by the structural model is

\[ \alpha, \beta \in (0, 1), \rho_p, \rho_y \in \mathbb{R} \]
\[ \phi \in [0, \infty), \theta \in (1, \infty), \gamma, \rho_e = \rho_\theta \in [0, 1) \]
\[ \sigma_a, \sigma_e, \sigma_h, \sigma_r \in \mathbb{R}^+, \sigma_e^2 = (1/\phi)^2 \sigma_y^2, \psi = (\theta - 1)/\phi \in [0, \infty) \]

It should be noted that all these constraints imposed on the structural parameter space can turn out to be not only very restrictive but also impossible (when the structural parameter restrictions violate the parameter space defined by the statistical model), which makes all the inference useless. Moreover, by imposing them in the structural model itself, we miss the opportunity to test them against the data in the context of the statistical model. Once we derive the statistical model behind this structural model we would have relaxed all the parameter restrictions imposed by the theory and allow the parameters to have their own voice determined by the data within the statistical parameter space. To conclude, instead of helping identifying the parameters, the structural restrictions actually hinders proper inference.

Ireland (2011) assumes that \( \psi = 0.1 \) which reflects that the firms reset their price every 3.74 quarters (Gali and Gertler, 1999). He further computes the steady state values as

\[ \bar{h} = \bar{y} = \frac{1}{T} \sum (Y_t/Y_{t-1}) = 1.0048 \quad \bar{p} = \frac{1}{p} \sum (P_t/P_{t-1}) = 1.0086 \quad \bar{r} = \frac{1}{T} \sum r_t = 1.0127 \]

where \( T \) is the sample size. Ireland computes these values from the data ignoring the fact that these means make sense only when the data mimics the Normal, independent and identically distributed process. Ireland then estimates the structural model but gets \( \beta > 1 \), which is a violation of the structural parameter restriction. To solve this problem, he fixes \( \beta = 0.99 \) using the steady state relation \( \bar{r} = \frac{\bar{p}}{\bar{h}^\beta} \). The estimation results of Ireland (2011) model is reproduced in Table 2.2. In his estimation, the data used ranges from the quarter 1983:1
through 2009:4 for all the observables $y_t, p_t$ and $r_t$. One important point to be noticed in his estimation results is that out of 1000 bootstrap simulations, the distribution of $\rho_e$ and $\alpha$ are degenerate at lower bound of the structural parameter space leaving no room to compute their standard errors. This can be suspected as an effect of some structural restrictions causing estimates to disobey the statistical parameters space.

### Table 2.2: Estimates of Structural Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>0.3904</td>
<td>0.0685</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0000</td>
<td>-</td>
</tr>
<tr>
<td>$\rho_p$</td>
<td>0.4153</td>
<td>0.0430</td>
</tr>
<tr>
<td>$\rho_y$</td>
<td>0.1270</td>
<td>0.0278</td>
</tr>
<tr>
<td>$\rho_a$</td>
<td>0.9797</td>
<td>0.0116</td>
</tr>
<tr>
<td>$\rho_e$</td>
<td>0.0000</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_a$</td>
<td>0.0868</td>
<td>0.0497</td>
</tr>
<tr>
<td>$\sigma_e$</td>
<td>0.0017</td>
<td>0.0003</td>
</tr>
<tr>
<td>$\sigma_h$</td>
<td>0.0095</td>
<td>0.0013</td>
</tr>
<tr>
<td>$\sigma_r$</td>
<td>0.0014</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

This kind of identification strategy is typical in the DSGE macroeconometric literature, which raises the following methodological problems:

(a) The structural identification strategy used by Ireland is ad-hoc. $\beta = 0.99$ and $\psi=0.1$ are examples of such identification strategies, because they are neither tested against nor identified using the data.

(b) This strategy does not use the full information in the data and the structure of the model. The data on the variables $y_t, p_t$ and $r_t$ are non-Normal, non-identical and non-independent. In such a situation, the steady state values cannot be considered constant as suggested in the aforementioned identification strategy. The steady states are at least trending.
(c) This strategy gives the structural parameter space which restricts and sometimes violates the statistical parameter space. Although the parameter restrictions are plausible from the substantive point of view, statistically they can be very restrictive and even impossible. Rather one can always allow the data to speak for itself and test one’s substantive theory without any such prior within the boundary of the statistical parameter space.

If we assume Normality, Markov(2) and homogeneity (identically distributed), one straightforward way to identify the steady state values using the statistical model is explained as follows. The operational definition of the observables in terms of the data is given by:

\[ \hat{y}_t = \ln(Y_t) - \ln(Y_{t-1}) - \ln(\bar{y}) \]
\[ \hat{p}_t = \ln(P_t) - \ln(P_{t-1}) - \ln(\bar{p}) \]
\[ \hat{r}_t = \ln(r_t) - \ln(\bar{r}) \]

where \( Y_t \) is the quarterly per capita real GDP of the US economy, \( P_t \) is the quarterly GDP deflator, \( r_t \) is the quarterly gross interest rate on 90 days treasury bill. \( \bar{y}, \bar{r} \) and \( \bar{p} \) denote the arithmetic mean of \( \frac{Y_t}{Y_{t-1}}, r_t \) and \( \frac{P_t}{P_{t-1}} \) respectively. Since, \( d_t = Z_t - z \), where \( Z_t = [\ln(y_t) \ \ln(p_t) \ \ln(r_t)]^\top \) and \( z = [\ln(\bar{y}) \ \ln(\bar{p}) \ \ln(\bar{r})]^\top \), equation (2.49) can be written as:

\[ Z_t = \Psi_0 + \Psi_1 Z_{t-1} + \Psi_2 Z_{t-2} + u_t \quad (2.50) \]

where, the condition \( \Psi_0 = (I - \Psi_1 - \Psi_2)z \) can be used to identify \( z \) once (2.50) is estimated. With the structural parameter restrictions imposed, (2.50) in the estimated form is given below
\[
\begin{bmatrix}
\hat{\ln(y_t)} \\
\hat{\ln(p_t)} \\
\hat{\ln(r_t)}
\end{bmatrix} = 
\begin{bmatrix}
3.157 \\
1.544 \\
0.532
\end{bmatrix} + 
\begin{bmatrix}
0.598 & 0.275 & -1.330 \\
-0.047 & 0.325 & -0.765 \\
0.033 & 0.057 & 0.498
\end{bmatrix} 
\begin{bmatrix}
\ln(y_{t-1}) \\
\ln(p_{t-1}) \\
\ln(r_{t-1})
\end{bmatrix} + 
\begin{bmatrix}
0 & 0 & 1.664 \\
0 & 0 & 0.043 \\
0 & 0 & 0.114
\end{bmatrix} 
\begin{bmatrix}
\ln(y_{t-2}) \\
\ln(p_{t-2}) \\
\ln(r_{t-2})
\end{bmatrix}
\]

(2.51)

The number of zeros in (2.51) hints the amount of restrictions imposed by the structural model on the structural VAR(2) (2.50). The estimated result of the unrestricted VAR(2) is presented as the last three columns of Table 4.2 in Chapter 4. Once we achieve a statistically adequate model in Chapter 4, we will be in a position to do a likelihood ratio test to see if the estimated structural VAR(2) model (2.51) is approved by the data.

### 2.4 Exploratory Data Analysis (EDA)

The Probabilistic Reduction (PR) approach renders the use of graphical techniques and the EDA an integral part of statistical modeling (Spanos, 2006). The EDA can provide very useful information concerning the statistical information in the data for the specification and respecification of the model. Before going into modeling the data set \( Z_0 \) as defined in the identification subsection, the visual analysis is presented to identify the probabilistic structure of the data which can also provide potential hints about respecifying the model. Ireland (2011) uses data from the quarter 1983:1 through 2009:4. Ireland (2004), in exactly the same model with minor changes, uses data from the quarter 1948:1 through 2003:1. In this research, the data is updated to include quarters from 1948:1 through 2010:3. Using a larger sample size substantially increases the power of the M-S tests and the standard error of the estimates are significantly improved. Figures 2.2, 2.4 and 2.6 show the quarterly t-plots of the data for 100 times per-capita real GDP \((100Y_t)\), GDP deflator \((P_t)\) and gross
interest rate \( (r_t) \). These t-plots in level show that there are clear non-linear trends in all the series. Hence a model without a trend in mean, variance or both is likely to be misspecified. The positive dependence is clearly seen in \( Y_t \) and \( r_t \). The interest rate seems to be very volatile and its dependence is smoother and stronger.

The t-plots in the log-difference (Figures 2.3 and 2.5) show the clear presence of changing variance across time. Figures 2.3 and 2.5 show the t-plots of the data on the growth rate \( y_t \) and inflation \( p_t \) series respectively. These plots show a break in the volatility around 1982. For \( P_t \), even the log differenced data \( p_t \) is trending in a non linear fashion.

One very important point to be noted is that if a variable in level has trend, then the log difference will never get rid of the trend. This means we cannot exclude the trends in our model even if the data might look stationary once log-differenced. We can show this using an example as follows: Let \( X_t = \delta_0 + \delta_1 t \), then \( x_t = \ln(X_t/X_{t-1}) \) is always a function of \( t \). Moreover the linear trend in the original level form now reappears non-linearly in the log-differenced form.

For now, we ignore all the exhibited chance regularity patterns pertaining to the probabilistic structure of the data for the purpose of showing how it affects the inference when these patterns are ignored as done in Ireland (2011) model.
Figure 2.2: Per Capita Real Gross Domestic Product (GDP)

Figure 2.3: Growth Rate of Per Capita Real GDP
Figure 2.4: GDP Deflator

Figure 2.5: Inflation Rate (Log Difference of GDP Deflator)
2.5 Probabilistic Reduction (PR) and VAR Model

It can be noticed from (2.50) and (2.51) that the parameter space of $\Psi_0$, $\Psi_1$ and $\Psi_2$ are determined by the structural model which can be statistically very restrictive. Now, we can view the model (2.50) as a structural VAR with two lags, whose restrictions are derived from the structural model. It can be embedded in the following statistical generating mechanism (GM), which is free from any structural restrictions

$$Z_t = a_0 + A_1^T Z_{t-1} + A_2^T Z_{t-2} + u_t, u_t \sim N(0, \Omega), t \in \mathbb{N}$$  \hspace{1cm} (2.52)

For the statistical purpose, all the restrictions from the structural model (2.50) are relaxed so that the data can speak for itself and we can test the theory against the independent voice of the data. In other words, we can always test the hypothesis
once we have the statistically adequate model, which gives (2.52) as a statistical GM. Ignoring
the structural restrictions for the statistical purpose allows us to delineate the substantive
information from the statistical ones so that the implicit statistical assumptions behind (2.52)
can be tested independent of any substantive assumptions.

The transformation from the original structural model (2.43) to the statistical model (2.52)
can be viewed as a reparameterization of the original structural parameters in different stages
as follows

$$\psi_1 = \{A, B, C, P, V\} \rightarrow \psi_2 = \{\Pi\} \rightarrow \theta_1 = \{\Psi_0, \Psi_1, \Psi_2, V\} \rightarrow \theta_2 = \{a_0, A_1, A_2, \Omega\}$$

Here, $\psi_1$ has 12 structural parameters and $\theta_2$ has 27 statistical parameters. In (2.50), which
is a structural VAR (SVAR), the coefficients $\Psi_0, \Psi_1$ and $\Psi_2$ are related and restricted by
the structure of the Ireland’s model, which can be tested once the statistical adequacy of
(2.52), which is a Normal VAR(2) , is guaranteed. We can go back to the structural model
through the overidentifying restrictions $G(\psi_1, \theta_2) = 0$. A formal likelihood ratio test of this
restriction is done in Chapter 4 after a statistically adequate model is secured for the data
chosen.

Now, the reliability of inference results based on the structural model (2.43) relies on the
statistical adequacy of the statistical model (2.52). To assess the statistical adequacy of
(2.52), it will be viewed as a statistical GM which arises from a set of restrictions on the joint
distribution of the observations on $Z_t = [\ln(y_t) \ \ln(p_t) \ \ln(r_t)]^T$. In other words, equation
(2.52) can be viewed as stemming from a parametrization of the stochastic process \{Z_t, t∈\mathbb{N}\}
whose probabilistic structure is given by the joint distribution $D(Z_1, Z_2, ..., Z_T; \Theta)$, where
$T$ is the sample size. To turn any stochastic process into a statistical model, three sets of
assumptions are required (Spanos, 1990): dependence (I), heterogeneity (H) and distribution
To render the VAR(2) in (2.52) as a statistical GM of a statistical model, the three sets of restrictions are imposed on the stochastic process \( \{Z_t, t \in \mathbb{N}\} \) as shown in Table 2.3.

<table>
<thead>
<tr>
<th>Distribution (D)</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependence (I)</td>
<td>Markov(2)</td>
</tr>
<tr>
<td>Heterogeneity (H)</td>
<td>Homogeneous</td>
</tr>
</tbody>
</table>

With the Markov(2) and the homogeneity (stationarity) restrictions, focus can be explicitly made on the lower dimensional joint distribution \( D(Z_t, Z_{t-1}, Z_{t-2}; \Theta), t \in \mathbb{N} \), which can be decomposed into a product of the conditional distribution and the marginal distribution as follows

\[
D(Z_t, Z_{t-1}, Z_{t-2}; \Theta) = D(Z_t|Z_{t-1}, Z_{t-2}; \Theta_1)D(Z_{t-1}, Z_{t-2}; \Theta_2), t \in \mathbb{N}
\]  

(2.53)

where \( \Theta, \Theta_1 \) and \( \Theta_2 \) are the parameters of the joint distribution, the conditional distribution and the marginal distribution respectively. The Normality assumption implies that

\[
\begin{bmatrix}
Z_t \\
Z_{t-1} \\
Z_{t-2}
\end{bmatrix}
\sim 
\mathcal{N}
\begin{bmatrix}
\mu \\
\mu \\
\mu
\end{bmatrix},
\begin{bmatrix}
\Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\
\Sigma_{12}^\top & \Sigma_{11} & \Sigma_{12} \\
\Sigma_{13}^\top & \Sigma_{12}^\top & \Sigma_{11}
\end{bmatrix}
\]  

(2.54)

where \( \mu(k \times 1) \) and \( \Sigma_{ij}(k \times k) \) for \( i, j = 1, 2, 3 \), partition of \( \Theta \), are the first moment (mean) and the second moment (covariance) of the joint distribution respectively and \( k \) is the number of variables in \( Z_t \). The Normality assumption ensures the Normality of the conditional distribution as well as the marginal distribution. Using (2.53) and (2.54), the exhaustive list of the probabilistic assumptions behind the VAR(2) model is given in Table 2.4. Here, the parameters of the model (2.52) can be viewed in terms of the parameters of the joint distribution (2.54)(Spanos, 2006b).

The parameters of the VAR(2) model (2.52) can be derived directly from the joint distribu-
Table 2.4: Normal VAR(2) Model

<table>
<thead>
<tr>
<th>Condition</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Normality:</td>
<td>$D(Z_t, Z_{t-1}, \ldots, Z_1; \theta)$ is Normal</td>
</tr>
<tr>
<td>2. Linearity:</td>
<td>$E(Z_t</td>
</tr>
<tr>
<td>3. Homosked.:</td>
<td>$Var(Z_t</td>
</tr>
<tr>
<td>4. Markov:</td>
<td>${Z_t, t \in \mathbb{N}}$ is a Markov(2) process</td>
</tr>
<tr>
<td>5. $t$-invariance:</td>
<td>$\Theta_1 := (a_0, A_1, A_2, \Omega)$ are $t$-invariant for all $t \in \mathbb{N}$</td>
</tr>
</tbody>
</table>

\[
a_0 = [\mu - A_1 \mu - A_2 \mu]
\]
\[
A_1 = \Sigma_{12} (\Sigma_{11} - \Sigma_{12} \Sigma_{11}^{-1} \Sigma_{12})^{-1} - \Sigma_{13} \Sigma_{11} \Sigma_{12} \Sigma_{11}^{-1} \Sigma_{12}^{-1} \Sigma_{12}
\]
\[
A_2 = -\Sigma_{12} \Sigma_{11} \Sigma_{12} (\Sigma_{11} - \Sigma_{12} \Sigma_{11}^{-1} \Sigma_{12})^{-1} + \Sigma_{13} (\Sigma_{11}) - \Sigma_{12} \Sigma_{11}^{-1} \Sigma_{12}^{-1} \Sigma_{12}
\]
\[
\Omega = \Sigma_{11} - \left[ \begin{array}{cc} \Sigma_{12} & \Sigma_{13} \\ \Sigma_{12}^T & \Sigma_{11} \end{array} \right]^{-1} \left[ \begin{array}{c} \Sigma_{12} \\ \Sigma_{13} \end{array} \right]
\]

2.6 Misspecification (M-S) Tests for VAR Model

The estimation result of the Normal VAR(2), alongside a statistically adequate model, is presented and discussed in Chapter 4. At this point, it is checked if the model assumptions of the VAR(2) are valid for the data. The way to ensure the statistical adequacy of the model in question (VAR(2) in our case) is to apply a thorough M-S testing. Spanos and McGuirk (2001) suggest the M-S tests for the Normal linear regression model based on the F-tests on the significance of certain coefficients of the auxiliary regressions.

The regressions with standardized residuals as the regressand and different functions of the regressors of the null model as regressors are defined as the auxiliary regressions. Instead of using all the regressors of the null model as the regressors of the auxiliary equation, Spanos and McGuirk (2001) suggest to use the fitted values of the null model as the regressors of
the auxiliary regression because the fitted value represents the linear combination of the regressors of the null model. This strategy is appropriate when the number of regressors is large (in the VAR(2), with 3 variables and 2 lags there are 6 regressors). Moreover, using the polynomials and the extra lags in the auxiliary regressions reduces the degrees of freedom significantly. So the fitted values of the original regression is used as a proxy for the right hand side variables of the auxiliary regressions.

Since this is a M-S test, low power is not an issue unless we accept the null model. Moreover, different forms of the auxiliary equations can be used to probe for potential departures from the null model assumptions in different directions. Trying the different functional forms also ensures the thoroughness of the test. Mayo and Spanos (2004) extensively discuss about how one can ensure the thoroughness and the reliability of the misspecification testing.

2.6.1 Standardized Residuals

The standardized residuals used for the M-S tests are defined as follows

\[ \hat{u}_t = \begin{bmatrix} \hat{u}_{yt} \\ \hat{u}_{pt} \\ \hat{u}_{rt} \end{bmatrix} = L^{-1}(Z_t - \hat{Z}) = L^{-1}(Z_t - \hat{a}_0 - \hat{A}_1 Z_{t-1} - \hat{A}_2 Z_{t-2}) \quad (2.55) \]

where, \( \hat{\Omega} = LL^T \), \( \hat{Z}_t = [\hat{y}_t \ \hat{p}_t \ \hat{r}_t]^T \) is the fitted vector. It should be noted that the conditional variance-covariance matrix \( \Omega \) depends neither on \( t \) nor on the regressors \( Z^o_{t-1} \). In other words, \( \Omega \) is both homogeneous and homoskedastic.
2.6.2 M-S Testing Auxiliary Regression for VAR Model

The auxiliary regression is a regression of the residual on the right hand side terms in the conditional mean (autoregressive function) of the null model (i.e. Normal VAR(2)) and some extra terms, which have the potential to pick up the departures from the null. For example, the auxiliary regressions to test the assumptions [2]-[5] of Table 2.4 are written in terms of the standardized residual of the growth rate equation (\( \hat{u}_{yt} \)) of the VAR(2)

\[
\hat{u}_{yt} = a_0 + a_1 \hat{u}_{yt-1} + a_2 \hat{u}_{yt-2} + v_t
\]  

(2.56)

\[
\hat{u}_{yt} = b_0 + b_1 \hat{y}_t + b_2 \hat{y}_t^2 + b_3 t + b_4 t^2 + v_t
\]  

(2.57)

\[
\hat{u}_{yt}^2 = c_0 + c_1 \hat{y}_t + c_2 \hat{y}_t^2 + c_3 \hat{y}_{t-1}^2 + c_4 \hat{y}_{t-2}^2 + c_5 t + c_6 t^2 + v_t
\]  

(2.58)

These auxiliary regressions can be used in many other different forms for thorough M-S testing. Many different forms were tried before realizing that the results are not significantly different. Hence, one of the F-test results is presented in this chapter. Table 2.5 summarizes the hypotheses tested and the corresponding degrees of freedom inside parenthesis. Table 2.6 is the result of the M-S tests shown in Table 2.5.

The distributional assumption [1] is assessed via the standardized residual. For the assumption of Normality, three different tests are done as shown in Table 2.7. Details of these three tests are given in Appendix F.

2.6.3 Interpretation of the M-S Testing Results

It can be seen that most of the assumptions are severely violated. The low p-values inside the parentheses in Tables 2.6 and 2.7 indicate serious misspecification. Violation of the
### Table 2.5: M-S Tests for VAR(2)

<table>
<thead>
<tr>
<th>Null Hypotheses</th>
<th>Auxiliary Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linearity F(242,1)</td>
<td>$H_0 : b_2 = 0$ (2.57)</td>
</tr>
<tr>
<td>$t$-invariance F(242,2)</td>
<td>$H_0 : b_3 = b_4 = 0$ (2.57)</td>
</tr>
<tr>
<td>Independence F(242,2)</td>
<td>$H_0 : a_1 = a_2 = 0$ (2.56)</td>
</tr>
<tr>
<td>Heteroskedasticity F(238,2)</td>
<td>$H_0 : c_1 = c_2 = 0$ (2.58)</td>
</tr>
<tr>
<td>2nd order Independence F(238,2)</td>
<td>$H_0 : c_3 = c_4 = 0$ (2.58)</td>
</tr>
<tr>
<td>2nd order $t$-invariance F(238,2)</td>
<td>$H_0 : c_5 = c_6 = 0$ (2.58)</td>
</tr>
</tbody>
</table>

### Table 2.6: M-S Test Results for VAR(2)

<table>
<thead>
<tr>
<th>Growth Rate ($y_t$)</th>
<th>Inflation Rate ($p_t$)</th>
<th>Interest Rate ($r_t$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linearity</td>
<td>1.44 [0.232]</td>
<td>0.607 [0.437]</td>
</tr>
<tr>
<td>Homoskedasticity</td>
<td>5.299 [0.006]</td>
<td>37.285 [0.000]</td>
</tr>
<tr>
<td>1st Independence</td>
<td>0.348 [0.706]</td>
<td>0.013 [0.987]</td>
</tr>
<tr>
<td>2nd Independence</td>
<td>0.234 [0.701]</td>
<td>3.488 [0.032]</td>
</tr>
<tr>
<td>1st $t$-invariance</td>
<td>1.233 [0.294]</td>
<td>3.509 [0.032]</td>
</tr>
<tr>
<td>2nd $t$-invariance</td>
<td>12.008 [0.000]</td>
<td>50.542 [0.000]</td>
</tr>
</tbody>
</table>

### Table 2.7: Normality Tests for VAR(2)

<table>
<thead>
<tr>
<th>Tests</th>
<th>$y_t$</th>
<th>$p_t$</th>
<th>$r_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson-Darling (AD)</td>
<td>0.722 [0.058]</td>
<td>4.359 [0.000]</td>
<td>8.501 [0.000]</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov (KS)</td>
<td>0.054 [0.120]</td>
<td>0.095 [0.000]</td>
<td>0.142 [0.000]</td>
</tr>
<tr>
<td>Skewness-Kurtosis (SK(2))</td>
<td>21.895 [0.000]</td>
<td>300.14 [0.000]</td>
<td>1348.8 [0.000]</td>
</tr>
</tbody>
</table>
invariance assumption is particularly acute problem for the skedastic function of the inflation and the growth rate equations. The only assumption that seems to be accepted by the data set we have is the linearity assumption. The heteroskedasticity is severe for all the variables. Lingering temporal dependence seems to be a problem for at least the interest rate autoregression. The interest rate equation also shows the signs of second order dependence. The most serious departures from the null model assumptions appear to be the departures from the assumptions of homogeneity, Normality and homoskedasticity. The validity of the linearity assumptions directs the respecification towards other elliptically symmetric family of distributions because homoskedasticity characterizes the Normal distribution within this family; see Spanos (1994). In the light of the lessons learned from these M-S tests, the respecification is done in Chapter 4.
Klein (2000) proposes a solution method that is a hybrid of the solutions proposed by Blanchard and Kahn (1980) and Sims (2002). His method is applied to a system in the form

\[ A \mathbf{E}_{t} \mathbf{s}_{t+1} = B \mathbf{s}_{t} + C \mathbf{v}_{t}, \]  

\[ \mathbf{v}_{t} = \mathbf{P} \mathbf{v}_{t-1} + \mathbf{\epsilon}_{t} \]  

Here, \( \mathbf{P} \) is an autocorrelation matrix, \( \mathbf{\epsilon}_{t} \) is an independently and identically distributed innovation process with dimension \((h \times 1)\). Let us assume that out of \( p \) variables, there are \( n \) predetermined variables in \( \mathbf{s}_{t} \) so that \( \mathbf{A}, \mathbf{B} \) and \( \mathbf{C} \) are of dimensions \((p \times p)\), \((p \times p)\) and \((p \times h)\) respectively. An important advantage of Klein’s method over other methods is that \( \mathbf{A} \) is allowed to be singular. His method relies on the complex generalized Schur decomposition of the matrix couple \( \mathbf{A} \) and \( \mathbf{B} \), which identifies the unitary matrices \( \mathbf{Q} \) and \( \mathbf{Z} \) such that

\[ \mathbf{QAZ} = \mathbf{S} \]  

\[ \mathbf{QBZ} = \mathbf{T} \]
where $S$ and $T$ are both upper triangular. The generalized eigenvalues of the matrix couple $A$ and $B$ can be written as the ratios of the diagonal elements of $T$ and $S$ as follows:

$$\lambda(B, A) = \{t_{ii}/s_{ii} : i = 1, 2, 3...\}$$

The matrices $Q, Z, S$ and $T$ can always be arranged such that the generalized eigenvalues are in an ascending order from left to right. Let there are $m$ generalized eigenvalues lying outside the unit circle. If $n = m$, which is satisfied by Ireland (2011) estimated parameter values for the US data, there exists a unique solution to the system A.1-A.2. If $n > m$, there exist multiple solutions. If $m > n$, then there does not exist any solution to the system.

The matrices $Q, Z, S$ and $T$ can be partitioned as

$$Q = \begin{bmatrix} Q_1(n \times p) \\ Q_2(n \times p) \end{bmatrix}, Z = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}, S = \begin{bmatrix} S_{11} & S_{12} \\ 0_{(n \times p-2)} & S_{22} \end{bmatrix}, T = \begin{bmatrix} T_{11} & T_{12} \\ 0_{(n \times p-2)} & T_{22} \end{bmatrix}$$

Here $Z_{11}$ corresponds to the non-explosive eigenvalues of the system. Now we triangularize the system A.1 by defining a new vector $z_t$ of the auxiliary variables are defined as

$$z_t = Z's_t^0$$  \hspace{1cm} (A.5)$$

This new vector $z_t$ can be further partitioned into a stable $s_t$ and an unstable $u_t$ components as follows

$$z_t = \begin{bmatrix} s_t \\ u_t \end{bmatrix}$$

---

1Generalized eigenvalue of matrix couple $A$ and $B$ is defined as the roots of matrix equation $|A - \lambda B| = 0$
Using (A.3) and (A.4), the system A.1 can be written as

\[
\begin{bmatrix}
S_{11} & S_{12} \\
0 & S_{22}
\end{bmatrix} E_t \begin{bmatrix}
s_{t+1} \\
u_{t+1}
\end{bmatrix} = \begin{bmatrix}
T_{11} & T_{12} \\
0 & T_{22}
\end{bmatrix} \begin{bmatrix}
s_t \\
u_t
\end{bmatrix} + \begin{bmatrix}
Q_1 \\
Q_2
\end{bmatrix} C v_t
\]

(A.6)

Lower portion of A.6 contains the unstable components of the system.

\[
S_{22} E_t u_{t+1} = T_{22} u_t + Q_2 C v_t
\]

(A.7)

or,

\[
u_t = T^{-1}_{22} S_{22} E_t u_{t+1} - T^{-1}_{22} Q_2 C v_t
\]

(A.8)

Now this component can be solved by forward iteration to get

\[
u_t = [P'S_{22} - T_{22}]^{-1} Q_2 C v_t
\]

(A.9)

since

\[
\lim_{t \to \infty} \left(T^{-1}_{22} S_{22}\right)^t u_t = 0
\]

(A.10)

The solution for this unstable component can now be used to solve for the stable part, yielding

\[
s_{t+1} = S^{-1}_{11} T_{11} s_t + S^{-1}_{11} \{T_{12} M - S_{12} MP + Q_1 C\} v_t - Z^{-1}_{11} Z_{12} M \varepsilon_{t+1}
\]

(A.11)
Chapter 3

Multivariate Student’s $t$ Dynamic Models

3.4 Introduction

The literature in the Student’s $t$ distribution consists of two major branches

(a) Multivariate Student’s $t$ distribution (Kotz and Nadarajah, 2004)

(b) Matrix variate Student’s $t$ distribution (Gupta and Nagar, 2000)

In this chapter, the multivariate and the matrix variate Student’s $t$ distributions are briefly introduced and compared, showing that the later is a special case of the former. Based on the multivariate Student’s $t$ distribution, the Student’s $t$ VAR (St-VAR) model and the multivariate Student’s $t$ dynamic linear regression model will be derived under both the homogeneity (stationarity) and the heterogeneity assumptions using the Probabilistic Reduction (PR) approach. Heracleous (2003) derives the St-VAR model using the matrix variate Student’s $t$ distribution. It will be shown latter that her St-VAR imposes very strict parameter restrictions, making it unrealistic for practical purposes. The St-VAR model derived in this research will embed her St-VAR as a special case.
3.5 Multivariate Student’s $t$ Distribution

3.5.1 Joint Distribution

Let $X := (X_1, X_2, \ldots, X_p)^\top$ be a $p$ dimensional random vector. $X$ is said to have the $p$-variate Student’s $t$ distribution with the degrees of freedom $\nu$, a location vector $\mu$ and a scaling matrix $\Sigma$ if its joint probability density function (PDF) is given by

$$f(x; \theta) = \frac{\Gamma((\nu + p)/2)}{(\pi\nu)^{p/2}\Gamma(\nu/2)|\Sigma|^{1/2}} \left[ 1 + \frac{1}{\nu}(x - \mu)^\top \Sigma^{-1}(x - \mu) \right]^{-(\nu+p)/2}, \quad \theta = (\mu, \Sigma, \nu) \quad (3.1)$$

In what follows equation (3.1) is denoted by

$$X \sim St(\mu, \Sigma; \nu) \quad (3.2)$$

The degrees of freedom parameter $\nu$ is also called the shape parameter. An increase in $\nu$ reduces the peakedness of the distribution. As $\nu$ decreases towards 1, the marginal distributions have increasingly heavy tails. If $\mu = 0$, the distribution is said to be central; otherwise it is said to be non-central. Figures 3.1 and 3.2 show the density of the bivariate Normal distribution and the bivariate Student’s $t$ distribution with the degrees of freedom $\nu = 3$ respectively. Figures 3.3 and 3.4 show their corresponding contour plots.
Figure 3.1: Bivariate Normal Density \((\mu = 0, \Sigma = I, \nu = 3)\)

Figure 3.2: Bivariate Student’s \(t\) Density \((\mu = 0, \Sigma = \frac{1}{3}I, \nu = 3)\)
Figure 3.3: Bivariate Normal Density Contour ($\mu = 0, \Sigma = I, \nu = 3$)

Figure 3.4: Bivariate Student’s $t$ Density Contour ($\mu = 0, \Sigma = \frac{1}{3}I, \nu = 3$)
3.5.2 Special Cases of the Student’s $t$ Distribution

The Student’s $t$ distribution can give rise to many other distributions as its special cases. Some of them are as follows.

(a) If $\nu \to \infty$, the density will become that of the $p$-variate Normal with the mean $\mu$ and the covariance matrix $\Sigma$.

(b) If $\nu = 1$, the density reduces to the $p$-variate Cauchy density. For the Cauchy distribution, none of the moments exist.

(c) If $(\nu + p)/2 = m$ is an integer, then the density reduces to the $p$-variate Pearson type VII distribution.

(d) $Y = \ln(X)$ has the log-Student’s $t$ distribution.

3.5.3 Moments

There always exist the first $\nu - 1$ moments for the Student’s $t$ distribution. For example, if $\nu = 2$, only the first moment, mean, exists.

$$E(X) = \mu$$

(3.3)

If $\nu = 3$, both the mean and the variance exist. But no other higher moments exist.

$$Var(X) = \frac{\nu}{\nu - 2} \Sigma$$

(3.4)

If $\nu = 4$, the skewness exists and is equal to 0. If $\nu \geq 5$, the excess kurtosis is $6/(\nu - 4)$. The median and the mode of the distribution always exist and equal to $\mu$. 
3.5.4 Marginal Distribution

Let $\mathbf{X} \sim St(\boldsymbol{\mu}, \Sigma; \nu)$. Consider the partition of $\mathbf{X}$, $\boldsymbol{\mu}$ and $\Sigma$ as follows

$$
\mathbf{X} = \begin{pmatrix}
\mathbf{X}_1(p_1 \times 1) \\
\mathbf{X}_2(p_2 \times 1)
\end{pmatrix}
$$

(3.5)

$$
\boldsymbol{\mu} = \begin{pmatrix}
\boldsymbol{\mu}_1(p_1 \times 1) \\
\boldsymbol{\mu}_2(p_2 \times 1)
\end{pmatrix}
$$

(3.6)

$$
\Sigma = \begin{pmatrix}
\Sigma_{11}(p_1 \times p_1) & \Sigma_{12}(p_1 \times p_2) \\
\Sigma_{21}(p_2 \times p_1) & \Sigma_{22}(p_2 \times p_2)
\end{pmatrix}; p_1 + p_2 = p
$$

(3.7)

so that the following can be written

$$
\begin{pmatrix}
\mathbf{X}_1 \\
\mathbf{X}_2
\end{pmatrix} \sim St \left( \begin{pmatrix}
\boldsymbol{\mu}_1 \\
\boldsymbol{\mu}_2
\end{pmatrix}, \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{pmatrix}; \nu \right)
$$

(3.8)

where $\mathbf{X}_1$ and $\boldsymbol{\mu}_1$ are the $p_1$-variate vectors ($p_1 < p$) and $\Sigma_{11}$ is a positive definite $p_1 \times p_1$ matrix. Consequently, $\mathbf{X}_1$ is a $p_1$-variate Student's $t$ distributed random vector with $\nu$ degrees of freedom with the location parameter $\boldsymbol{\mu}_1$ and the scaling matrix $\Sigma_{11}$

$$
f(\mathbf{x}_1) = \frac{\Gamma((\nu + p_1)/2)}{\pi^{p_1/2}(\nu/2)^{p_1/2}|\Sigma_{11}|} \left[ 1 + \frac{1}{\nu}(\mathbf{x}_1 - \boldsymbol{\mu}_1)^\top \Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1) \right]^{-(\nu+p_1)/2}
$$

(3.9)

and is denoted by

$$
\mathbf{X}_1 \sim St(\boldsymbol{\mu}_1, \Sigma_{11}; \nu).
$$

(3.10)

Using the similar argument

$$
\mathbf{X}_2 \sim St(\boldsymbol{\mu}_2, \Sigma_{22}; \nu).
$$

(3.11)
3.5.5 Conditional Distribution

Spanos (1994) provides a lemma which shows that

\[
[X_1 | X_2] \sim St \left( \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} [X_2 - \mu_2], [\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}] q(X_2); \nu + p_2 \right) \tag{3.12}
\]

where

\[
q(X_2) = \frac{\nu}{\nu + p_2} \left[ 1 + \frac{1}{\nu} (X_2 - \mu_2)^\top \Sigma_{22}^{-1} (X_2 - \mu_2) \right] \tag{3.13}
\]

As one can see the conditional mean is linear as in the case of the Normal distribution. But unlike the constant conditional variance-covariance (homoskedastic) in the Normal distribution case, it is heteroskedastic in the Student’s t distribution case; the conditional variance-covariance (autskedastic function) changes with \(X_2\). If \(\nu \to \infty\), then \(q(X_2) = 1\), and the conditional distribution (3.12) reduces to the Normal conditional distribution with the homoskedastic autoskedastic functions. Moreover, the autoskedastic function depends on the way the moment \(\mu\) is specified. As opposed, in the joint Normal distribution case, the conditional variance-covariance is independent of its mean \((\mu)\). Figure 3.5 clearly shows how the curvature of the autoskedastic functions depends on the value of \(\nu\). Lower the value of \(\nu\), higher the curvature or higher the peakedness of the joint density or fatter the tail.

Another point to note about the conditional distribution is that the degrees of freedom increases by the dimension of the conditioning vector \(X_2\). So even if \(\nu = 1\), at least the mean of the conditional distribution exists. If \(p_2 > 1\), the first two moments (mean and variance) of the conditional distribution exist.
### 3.6 Matrix Variate Student’s $t$ Distribution

As it will turn out, the matrix variate Student’s $t$ distribution is the special case of the multivariate Student’s $t$ distribution. A random matrix $Z(T \times m) = \{X_1(m \times 1), X_2(m \times 1), \ldots, X_T(m \times 1)\}$ is said to have a matrix variate Student’s $t$-distribution with the parameters $\mu, V, \Omega$ and the degrees of freedom $\nu$ if its probability density function (PDF) is given by

$$f(z) = \frac{\Gamma((\nu + m + T - 1)/2)}{\pi^{T/2}m\Gamma((\nu + T - 1)/2)}|\nu V|^{-\frac{1}{2}m}|\Omega|^{-\frac{1}{2}T}|I_T + \frac{1}{\nu}V^{-1}(z - \mu)^{\top}\Omega^{-1}(z - \mu)|^{-\frac{1}{2}(\nu+m+T-1)}$$

(3.14)

In short this can be denoted as

$$Z \sim St_{T,m}(\mu, \nu V, \Omega; \nu)$$

(3.15)
If \( \nu > 2 \), then its first and second moments are given respectively by

\[
E(Z) = \mu \\
Cov(vec(Z)) = \frac{\nu}{\nu - 2} V \otimes \Omega
\]

This distribution belongs to the class of the matrix variate elliptically contoured distributions. If \( \nu = 1 \), it reduces to the Cauchy distribution. We can also express equation (3.14) as a multivariate Student’s \( t \) distribution.

\[
vec(Z)_{(T \times T \times 1)} = \begin{pmatrix} 
  X_1 \\
  X_2 \\
  \vdots \\
  X_T 
\end{pmatrix} \sim St_{T,m} \left( \begin{pmatrix} 
  \mu_1 \\
  \mu_2 \\
  \vdots \\
  \mu_T 
\end{pmatrix} , \begin{pmatrix} 
  \nu v_{11} \Omega & \nu v_{12} \Omega & \cdots & \nu v_{1T} \Omega \\
  \nu v_{21} \Omega & \nu v_{22} \Omega & \cdots & \nu v_{2T} \Omega \\
  \vdots & \vdots & \ddots & \vdots \\
  \nu v_{T1} \Omega & \nu v_{T2} \Omega & \cdots & \nu v_{TT} \Omega 
\end{pmatrix} ; \nu \right)
\]

(3.18)

where \( X_t(m \times 1), \mu_t(m \times 1), t = 1, 2, \ldots, T \) and \( \Omega(m \times m) \). To show that this is a special case of the multivariate Student’s \( t \) distribution, let us suppose \( T = 2 \) so that

\[
\begin{pmatrix} 
  X_1 \\
  X_2 
\end{pmatrix} \sim St_{T,m} \left( \begin{pmatrix} 
  \mu_1 \\
  \mu_2 
\end{pmatrix} , \begin{pmatrix} 
  \nu v_{11} \Omega & \nu v_{12} \Omega \\
  \nu v_{21} \Omega & \nu v_{22} \Omega 
\end{pmatrix} ; \nu \right)
\]

(3.19)

This distribution is exactly the same as (3.8) except the following restrictions on the covariance matrix

\[
\Sigma_{ij} = v_{ij} \Omega, \ i, j = 1, 2
\]

(3.20)
This yields

\[ \Sigma_{12} \Sigma_{22}^{-1} = \frac{v_{12}}{v_{22}} I \]  

(3.21)

\[ \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} = \nu \left( v_{11} - \frac{v_{12}v_{21}}{v_{22}} \right) \Omega \]  

(3.22)

### 3.6.1 Conditional Distribution

With the restrictions in (3.20), the conditional distribution becomes

\[ [X_1|X_2] \sim St \left( \mu_1 + \frac{v_{12}}{v_{22}} (X_2 - \mu_2), \nu \left( v_{11} - \frac{v_{12}v_{21}}{v_{22}} \right) \Omega q(X_2); \nu + m \right) \]  

(3.23)

We can easily see that the regression coefficient matrix \( \Sigma_{12} \Sigma_{22}^{-1} \) is a diagonal matrix with a same element along the diagonal when \( X_1 \) and \( X_2 \) are both of the same dimension \( m \). Similar conclusions can be drawn about the autoskedastic functions. This is a very strict restriction on the statistical GM. This is the reason for adopting the multivariate Student’s \( t \) distribution instead of the matrix variate Student’s \( t \) distribution.

In the next few sections, the dynamic multivariate regression models based on the multivariate Student’s \( t \) distribution are derived.

### 3.7 Student’s \( t \) VAR (St-VAR) Model

The St-VAR model using the random matrices based on Gupta and Nagar (1999) is proposed by Heracleous (2003). She uses the matrix variate Student’s \( t \) distribution which, by definition, imposes too many strict restrictions on the variance-covariance matrix, making it of little value for practical modeling purpose. For example with Markov(1), only non zero coefficient in the conditional mean equation will be the coefficient of its own lag and they will
be same for each equation. In this research, all those restrictions are relaxed and a St-VAR is developed based on the multivariate Student’s $t$ distribution.

Let $\{Z_t, t = 1, 2, \ldots \}$ be a vector stochastic process. For the PR approach to be used as a modeling framework, one needs to impose the three sets of restrictions on the stochastic process so that the process can be summarized by a joint distribution. The restrictions are imposed in way shown in Table 3.1.

<table>
<thead>
<tr>
<th>Table 3.1: Probabilistic Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution(D):</td>
</tr>
<tr>
<td>Dependence(I):</td>
</tr>
<tr>
<td>Heterogeneity(H):</td>
</tr>
</tbody>
</table>

With this set of restrictions in Table 3.1, the joint distribution of $\{Z_t, Z_{t-1}, \ldots, Z_{t-l}\}$ can be written as

$$
X_t = \begin{bmatrix}
    Z_t \\
    Z_{t-1} \\
    Z_{t-2} \\
    \vdots \\
    Z_{t-l}
\end{bmatrix} \sim St
\begin{bmatrix}
    \mu_z \\
    \mu_z \\
    \mu_z \\
    \vdots \\
    \mu_z
\end{bmatrix},
\begin{bmatrix}
    \Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \ldots & \Sigma_{1l+1} \\
    \Sigma_{12}^T & \Sigma_{11} & \Sigma_{12} & \ldots & \Sigma_{1l} \\
    \Sigma_{13}^T & \Sigma_{12}^T & \Sigma_{11} & \ldots & \Sigma_{1l-1} \\
    \vdots & \vdots & \vdots & \ldots & \vdots \\
    \Sigma_{1l+1}^T & \Sigma_{1l}^T & \Sigma_{1l-1}^T & \ldots & \Sigma_{11}
\end{bmatrix}; \nu
$$

or,

$$
X_t \sim St(\mu, \Sigma; \nu)
$$

where, $\nu$ is the degrees of freedom parameter,

$$
X_t = \begin{bmatrix} Z_t & Z_{t-1} & Z_{t-2} & \ldots & Z_{t-l} \end{bmatrix}^T
\quad \mu = \begin{bmatrix} \mu_z & \mu_z & \mu_z & \ldots & \mu_z \end{bmatrix}^T
$$
\[
\Sigma = \begin{bmatrix}
\Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \ldots & \Sigma_{1l+1} \\
\Sigma_{12}^T & \Sigma_{11} & \Sigma_{12} & \ldots & \Sigma_{1l} \\
\Sigma_{13}^T & \Sigma_{12}^T & \Sigma_{11} & \ldots & \Sigma_{1l-1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\Sigma_{l+1}^T & \Sigma_{l}^T & \Sigma_{l-1}^T & \ldots & \Sigma_{11} \\
\end{bmatrix}
\]

The dimensions of the vectors and the matrices used above are as follows

\[Z_t: (k \times 1), \Sigma: (k \times k), \mu_z: (k \times 1)\]
\[\mu: (p \times 1), p = (l + 1)k\]
\[\Sigma: (p \times p)\]

The probability density function (PDF) of the Student’s t \(X_t\) is

\[
D(X_t; \Theta) = \frac{\Gamma(\nu + p)/2}{(\pi \nu)^{p/2} \Gamma(\nu/2)|\Sigma|^{1/2}} \left[1 + \frac{1}{\nu}(X_t - \mu)\Sigma^{-1}(X_t - \mu)\right]^{-(\nu+p)/2}
\]

(3.26)

where, \(p = (l + 1)k\) is number of variables in \(X_t\), \(k\) is number of variables in \(Z_t\), \(l\) is number of lags. Note that unlike in the Normal distribution, \(\text{Var}(X_t) = \frac{\nu}{\nu-2} \Sigma\). This makes a huge difference for relatively small values of \(\nu\), because the distribution of the estimators depends on it.

### 3.7.1 Joint, Conditional and Marginal Distributions

Let the vectors \(X_t\) and \(\mu\), and the matrix \(\Sigma\) are partitioned as follows

\[
X_t = \begin{bmatrix}
Z_t(k \times 1) \\
Z_{t-1}^0(lk \times 1)
\end{bmatrix}, \quad \mu = \begin{bmatrix}
\mu_z(k \times 1) \\
\mu_{lk}(lk \times 1)
\end{bmatrix}, \quad \Sigma = \begin{bmatrix}
\Sigma_{11}(k \times k) & \Sigma_{12}(k \times lk) \\
\Sigma_{12}^T(lk \times k) & Q(lk \times lk)
\end{bmatrix}
\]

Here, \(\mu_{lk}(lk \times 1)\) is a vector of \(lk\) \(\mu_z\)’s. Now, the joint, the conditional and the marginal distributions for all \(t \in \mathbb{N}\) can be derived as
\[
D(Z_t, Z^0_{t-1}; \Theta) = D(Z_t|Z^0_{t-1}; \Theta_1)D(Z^0_{t-1}; \Theta_2) \sim St(\mu, \Sigma; \nu) \tag{3.27}
\]

\[
D(Z_t|Z^0_{t-1}; \Theta_1) \sim St(a_0 + A^\top Z^0_{t-1}, \Omega q(Z^0_{t-1}); \nu + lk) \tag{3.28}
\]

\[
D(Z^0_{t-1}; \Theta_2) \sim St(\mu_{lk}, Q; \nu) \tag{3.29}
\]

where

\[
A^\top = \Sigma_{12} Q^{-1}
\]

\[
a_0 = \mu_z - A^\top \mu_{lk},
\]

\[
\Omega = \Sigma_{11} - \Sigma_{12} Q^{-1} \Sigma_{12}^\top
\]

\[
q(Z^0_{t-1}) = \left[1 + \frac{1}{\nu}(Z^0_{t-1} - \mu_{lk})^\top Q^{-1}(Z^0_{t-1} - \mu_{lk})\right]
\]

\[
\Theta_1 = \{a_0, A, \Omega, Q, \mu\}, \quad \Theta_2 = \{\mu, Q\}, \quad \Theta = \Theta_1 \cup \Theta_2
\]

The lack of variation freeness between \(\Theta_1\) and \(\Theta_2\) can be clearly seen as both \(\Theta_1\) and \(\Theta_2\) have the same elements \(\mu\) and \(Q\). If we denote the elements of the matrices and the vectors as follows

\[
Q^{-1} = \{q_{ij}; i, j = 1, 2, \ldots, lk\},
\]

\[
\Omega = \{\omega_{ij}; i, j = 1, 2, \ldots, lk\},
\]

\[
Z^0_{t-1} = \{z_i; i = 1, 2, \ldots, lk\},
\]

then the conditional variance-covariance can be written as follows

\[
\text{Var}(Z_t|\sigma(Z^0_{t-1})) = \frac{\nu}{\nu + lk - 2} \begin{bmatrix}
\omega_{11} & \omega_{12} & \cdots & \omega_{1k} \\
\omega_{12} & \omega_{22} & \cdots & \omega_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
\omega_{1k} & \omega_{2k} & \cdots & \omega_{kk}
\end{bmatrix}
\begin{bmatrix}
1 + \frac{1}{\nu} \left(\sum_{i=1}^{lk} \sum_{j=1}^{lk} q_{ij}(z_i - \mu_i)(z_j - \mu_j)\right)
\end{bmatrix}
\tag{3.30}
\]
We can see how this conditional covariance differs from that of the Normal VAR model. This is a quadratic function of the conditioning variables \((Z_{t-1}^0)\) making the conditional variance (autoskedastic function) heteroskedastic and dynamic at the same time. In order to model the mean heterogeneity, if we assume that \(\mu = \mu(t)\), then the autoskedastic functions will also inherit the heterogeneity. If \(\mu(t)\) is a polynomial of degree \(m\), the autoskedastic functions will be polynomial of degree \(2m\) in \(t\). If one had to model the heterogeneous conditional variance-covariance in the Normal VAR model, this strategy would not work. The conditional variance-covariance in the Normal VAR has to be modelled separately for its heterogeneity. This makes the St-VAR model a richer and easier modeling framework.

### 3.7.2 Special Cases of the St-VAR Model

The Student’s \(t\) distribution is a very general distribution in the sense that it allows one to choose the density with different shapes for the different values of \(\nu > 0\). Following are the two cases where the distribution takes two extreme forms.

(a) When \(\nu = 1\), the joint distribution \(D(X_t; \Theta)\) in (3.27) becomes the Cauchy distribution which has none of the moments. \(\mu\) and \(\Sigma\) simply become the location and the scaling parameters. But the conditional distribution \(D(Z_t|Z_{t-1}^0; \Theta_1)\) will have \(v + lk\) degrees of freedom meaning all the first \(v + lk - 1\) moments exist allowing us to continue to use the same formula for the regression and the skedastic functions. This means even with the Cauchy joint distribution, we still have the same Student’s \(t\) VAR. Note that the marginal distribution \(D(Z_{t-1}^0; \Theta_2)\) is still a Cauchy distribution without any moments.

(b) When \(\nu \to \infty\), the distribution turns into the Normal distribution which has all the moments, and we are back to the Normal VAR model.
3.8 Stationary St-VAR Model

The Student’s $t$ VAR (St-VAR($l;\nu$)) model is presented in Table 3.2 with a complete list of the probabilistic assumptions involved. This specification differs from the Normal VAR model in some important ways.

(a) The statistical GM in the St-VAR model now has the conditional variance-covariance as a quadratic function of $Z_{t-1}^0$. This adds to the dynamism of the model through dynamic autoskedastic functions (conditional variance-covariance). Note that, in the Normal VAR model, dynamic and/or heteroskedastic autoskedastic function is not possible. But when $\nu \to \infty$, $Cov(Z_t|\sigma(Z_{t-1}^0)) = \Omega$ and the St-VAR model reduces back to the Normal VAR with the constant conditional variance-covariance.

(b) If we impose some heterogeneity restriction on $\mu_t$, not only the autoregressive function, but also the autoskedastic functions of the St-VAR model will be heterogeneous. For example, if we impose a polynomial of degree $m$ in $t$ for $\mu_t$, the autoregressive function of the St-VAR model remains a polynomial of degree $m$ in $t$, but its autoskedastic function will become a polynomial of degree $2m$ in $t$. The heterogeneous autoskedastic function in the Normal VAR can be obtained only by specifying the variance-covariance $\Sigma$ of the joint distribution as heterogeneous or function of $t$.

Table 3.2 shows the stationary St-VAR($l;\nu$) model and Table 3.3 shows its special case for $l = 3$. 
Table 3.2: Student’s t VAR($l;\nu$)

| Statistical GM: $Z_t = a_0 + A_t^\top Z_{t-1} + u_t, Var(Z_t|\sigma(Z_{t-1})) = \frac{\nu}{\nu+3k-2} Q_0^t(Z_{t-1}), t \in \mathbb{N}$ |
|---|
| [1] Distribution D($Z_t, Z_{t-1}^0; \Theta$) is Student’s t with $\nu$ d.f. |
| [2] Linearity $E(Z_t|\sigma(Z_{t-1})) = a_0 + A_t^\top Z_{t-1}^0$ is linear in $Z_{t-1}^0 := (Z_{t-1}, ... Z_1)$ |
| [3] Heteroskedasticity $Var(Z_t|\sigma(Z_{t-1})) = \frac{\nu}{\nu+3k-2} Q_0^t(Z_{t-1})$ |
| [4] Markov: $\{Z_t, t \in \mathbb{N}\}$ is a Markov($l$) process |
| [5] $t$-invariance $\Theta = \{a_0, A, \mu, \Omega, Q\}$ are $t$-invariant for all $t \in \mathbb{N}$. |

Table 3.3: Student’s t VAR($l = 3;\nu$)

<table>
<thead>
<tr>
<th>Statistical GM: $Z_t = a_0 + A_t^\top Z_{t-1} + A_t^2 Z_{t-2} + A_t^3 Z_{t-3} + u_t, t \in \mathbb{N}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1] Student’s t: $D(Z_t, Z_{t-1}, ..., Z_1; \nu, \Theta)$, Student’s t with $\nu$ d.f.</td>
</tr>
<tr>
<td>[2] Linearity: $E(Z_t</td>
</tr>
<tr>
<td>[3] Heteroskedastic: $Var(Z_t</td>
</tr>
<tr>
<td>$q(Z_{t-1}^0) = \left{1 + \frac{1}{\nu} \begin{bmatrix} Z_{t-1} - \mu_z \ Z_{t-2} - \mu_z \ Z_{t-3} - \mu_z \end{bmatrix}^T Q_0^t(Z_{t-1}) \begin{bmatrix} Z_{t-1} - \mu_z \ Z_{t-2} - \mu_z \ Z_{t-3} - \mu_z \end{bmatrix} \right}^{-1}$</td>
</tr>
<tr>
<td>[4] Markov: ${Z_t, t \in \mathbb{N}}$ is a Markov(3) process</td>
</tr>
<tr>
<td>[5] $t$-invariance: $\Theta := (a_0, \mu, A_1, A_2, A_3, \Omega, Q)$ are $t$-invariant for all $t \in \mathbb{N}$.</td>
</tr>
</tbody>
</table>

3.9 Heterogeneous St-VAR Model

The stationary models eliminate the possibility of modeling time varying volatility in time series, which is a very important probabilistic structure of any macroeconomic data. The EDA and the M-S tests in Chapter 2 clearly indicated $t$-variance problem for the data on growth rate, inflation rate and interest rate. So a heterogeneous St-VAR model is needed.

Here, a non-stationary St-VAR($l = 3;\nu$) is specified. Let us assume that

$$\mu_z(t) = \mu_0 + \mu_1 t + \mu_2 t^2$$  \hspace{1cm} (3.31)
where, $\mu_0, \mu_1, \mu_2$ are all $k \times 1$ vectors. This makes the autoregressive function a quadratic function of $t$, say

$$a_0 = \mu_z(t) - A_1^T \mu_z(t - 1) - A_2^T \mu_z(t - 2) - A_3^T \mu_z(t - 3) = \delta_0 + \delta_1 t + \delta_2 t^2$$

where

$$\delta_0 = (I - A_1^T - A_2^T - A_3^T) \mu_0 + (A_1^T + 2A_2^T + 3A_3^T) \mu_1 - (A_1^T + 4A_2^T + 9A_3^T) \mu_2$$
$$\delta_1 = (I - A_1^T - A_2^T - A_3^T) \mu_1 + (2A_1^T + 4A_2^T + 6A_3^T) \mu_2$$
$$\delta_2 = (I - A_1^T - A_2^T - A_3^T) \mu_2$$

One important aspect of this model is that although heterogeneity is imposed only in mean of the joint distribution, both mean and variance-covariance of the conditional distribution are heterogeneous (i.e. functions of $t$). Table 3.4 is the specification of heterogeneous St-VAR($l = 3; \nu$) model. The functional form of $\mu(t)$ can be changed according to the kind of heterogeneity present in the data. If the data is not deseasonalized, which is the most preferred form of the data to be modeled, the mean can be specified using various trigonometric functions of $t$.

### 3.10 Dynamic Multivariate Student’s t Regression

The dynamic linear regression model for the Normal distribution has been derived in Spanos (1986). Similarly, Spanos (2006a) has derived Student’s $t$ dynamic linear regression model. In this section, we extend the model for multivariate regressand case. For the sake of simplicity, we develop a multivariate Student’s $t$ dynamic linear regression model with one lags ($l = 1$).
Let us consider a partition of $X_t$. The parametrization of the conditional distribution is similar to that of the St-VAR model.

**Linearity:** $E(Z_t|\sigma(Z_t^{(i)})) = \delta_0 + \delta_1 t + \delta_2 t^2 + A_1^T Z_{t-1} + A_2^T Z_{t-2} + A_3^T Z_{t-3}$

**Heterosked.:** $\text{Var}(Z_t|\sigma(Z_t^{(i)})) = (\frac{\nu}{\nu + 3k-2}) \cdot \Omega q(Z_t^{(i)})$ is free of $Z_t^{(i):}=(Z_{t-1},...Z_1)$,

$$q(Z_t^{(i)}) = \left\{ 1 + \frac{1}{\nu} \begin{bmatrix} Z_{t-1} - \mu_z(t) \\ Z_{t-2} - \mu_z(t) \\ Z_{t-3} - \mu_z(t) \end{bmatrix}^T \Omega^{-1} \begin{bmatrix} Z_{t-1} - \mu_z(t) \\ Z_{t-2} - \mu_z(t) \\ Z_{t-3} - \mu_z(t) \end{bmatrix} \right\}$$

**Markov:** $\{Z_t, \ t \in \mathbb{N}\}$ is a Markov(3) process

**$t$-invariance:** $\Theta := (\delta_0, \delta_1, \delta_2, \mu_0, \mu_1, \mu_2, A_1, A_2, A_3, \Omega, Q)$ are $t$-invariant for all $t \in \mathbb{N}$.

Let us consider a partition of $X_t$, $\mu$ and $\Sigma$ as follows

$$X_t = \begin{bmatrix} Z_t \\ Z_{t-1} \end{bmatrix} = \begin{bmatrix} y_t(m_1 \times 1) \\ X_t(m_2 \times 1) \\ Z_{t-1}(k \times 1) \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{12}^T & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{13}^T & \Sigma_{23}^T & \Sigma_{33} \end{bmatrix}$$

Now, we have the following decomposition of the joint distribution

$$D(y_t, X_t, Z_{t-1}; \Theta) = D(y_t|X_t, Z_{t-1}; \Theta_1)D(X_t, Z_{t-1}; \Theta_2) \sim St(\mu, \Sigma; \nu) \quad (3.34)$$

$$(y_t|X_t, Z_{t-1}) \sim St(a_0 + A_1^T X_t + A_2^T Z_{t-1}, \Omega q(X_t, Z_{t-1}); \nu + m_2 + k) \quad (3.35)$$

$$(X_t, Z_{t-1}) \sim St \left( \begin{bmatrix} \mu_2 \\ \mu_3 \end{bmatrix}, \begin{bmatrix} \Sigma_{22} & \Sigma_{23} \\ \Sigma_{23}^T & \Sigma_{33} \end{bmatrix}; \nu \right) \quad (3.36)$$

The parametrization of the conditional distribution is similar to that of the St-VAR model.
3.11 Student’s $t$ Distribution Having Marginals with Different Degrees of Freedom

One potential problem with the multivariate Student’s $t$ distribution we are using is that all the marginal distributions have the same degrees of freedom. This feature works as a restriction that all the marginal distributions have the same degrees of freedom so that the vector stochastic process $Z_t$ is assumed to have all the elements with same kurtosis. Jones (2002) generalized the bivariate Student’s $t$ distribution to distributions having marginals with different degrees of freedom. Further research is needed to turn it into a general multivariate distribution for developing VAR like models. The data used in this dissertation depict slightly different kurtosis. For example, the growth rate ($g_t$) has much lower kurtosis than the interest rate ($r_t$). Much better and general results can be expected by using different degrees of freedom for different data series. But this dissertation will be limited to models with the Student’s $t$ distribution having single degrees of freedom.

3.12 Asymmetric Student’s $t$ Distribution

Jones (2001a), Branco and Dey (2001), Azzalini and Capitanio (2003) and Kim and Mallick (2003) constructed a bivariate skew $t$-distribution. This type of distribution is especially interesting for modeling non-symmetric data (which is not the problem for the data set we are using in this dissertation). But the non-symmetric data is a very plausible scenario. Under such data circumstances modeling mean and variance might not be sufficient. Most of the information about the data may lie in the third moment function. This can give rise to a whole new family of asymmetric regression models.
3.13 Estimation: Maximum Likelihood

The lack of variation freeness between $\Theta_1$ and $\Theta_2$ does not allow to estimate $\Theta_1$ by ignoring the marginal distribution; see Spanos (1994). In other words, the likelihood function does not have enough information without the marginal part. Therefore, the likelihood function must be based on the whole joint distribution $f(X_t; \Theta)$, where

$$\ln f(\Theta; X_t) \propto -\frac{1}{2} \ln |\Sigma| - \frac{1}{2}(\nu + p) \ln \left[ 1 + \frac{1}{\nu}(X_t - \mu)\Sigma^{-1}(X_t - \mu) \right]$$

(3.37)

where, $c = \ln \left[ \Gamma \left\{ \frac{(\nu + p)}{2} \right\} \right] - \ln \left[ \Gamma \left( \frac{\nu}{2} \right) \right] - \frac{p}{2} \ln (\pi \nu)$. With the assumption of identical distribution, the log likelihood function for $T$ observations is given by

$$l_T(\Theta) = \ln L(\Theta; X_1, X_2, \ldots, X_T)$$

$$\propto Tc - \frac{T}{2} \ln |\Sigma| - \frac{1}{2}(\nu + p) \sum_{t=1}^{T} \ln \left[ 1 + \frac{1}{\nu}(X_t - \mu)\Sigma^{-1}(X_t - \mu) \right]$$

(3.39)

The maximum likelihood estimates (MLE) of the parameters $\mu$ and $\Sigma$ are obtained by maximizing the log-likelihood function (3.39). Since there does not exist explicit solution to the first order condition of the above maximization problem, one has to use numerical methods. Once $\hat{\mu}$ and $\hat{\Sigma}$ are obtained, (3.27)-(3.29) can be used to get the estimates $\hat{\Theta}_1 = \{\hat{a}_0, \hat{A}, \hat{\Omega}, \hat{\mu}\}$ for the parameters of the conditional distribution representing the Student’s $t$ VAR model.

3.13.1 Inference

If the model is correctly specified, the inference procedures such as hypothesis testing and prediction based on the maximum likelihood estimators have following optimal properties.
(a) Consistency: \( p \lim_{T \to \infty} \hat{\Theta} = \Theta \)

(b) Asymptotic Normality and efficiency: \( \hat{\Theta} \sim N(\Theta, I^{-1}(\Theta)) \), where \( I(\Theta) = \frac{\partial^2 \ln l_T}{\partial \Theta \partial \Theta^\top} \) is the Fisher information matrix. In practice, the observed information matrix \( I(\hat{\Theta}) \) can be used.

Once the information matrix is estimated from the Hessian matrix based on the likelihood function (3.39), they can be transformed to get the standard errors of the parameters of the conditional distribution \( \hat{\Theta}_1 = \{\hat{a}_0, \hat{A}, \hat{\Omega}, \hat{\mu}\} \) using the delta method (Greene, 2003). The delta method is explained in brief in the following paragraph.

Let \( \beta \) and \( \theta \) are two parameters linked by a function \( \beta = F(\theta) \). If \( \hat{\theta} \) is an MLE of \( \hat{\theta} \) then \( \hat{\theta} \sim N(\theta, \text{var}(\hat{\theta})) \). Given this, the distribution of \( \hat{\beta} \) is given by

\[
\hat{\beta} \sim N(\beta, \nabla F(\hat{\theta})^\top \text{var}(\hat{\theta}) \nabla F(\hat{\theta}))
\] (3.40)

where \( \nabla \) stands for gradient (multivariate equivalent of derivative). The derivative (or the Jacobian matrix) for the delta method can be obtained by using the numerical derivative techniques (Fornberg and Sloan, 1994; Lindfield and Penny, 1989), in R. One disadvantage of this method is that it is based on the first order Taylor expansion of the function \( F \), which can be very crude for non-linear cases. If \( F \) is linear, then the delta method gives the exact variance of \( \hat{\beta} \). However, one can easily go for the higher order Taylor expansion (Spanos (1999) pp: 493-494 for the second order Taylor expansion) to increase the accuracy.
Chapter 4

Respecification: Student’s $t$ VAR (St-VAR) Model

4.1 Introduction

Empirical modeling involves several stages such as specification, misspecification (M-S) testing, respecification and theory testing, and a number of interrelated models, including the theory model, the structural model, the statistical model and the empirical model. Focusing on the relationship between a statistical and a structural model, it has been argued that for every structural model there is an underlying statistical model which is often specified implicitly via the error term(s). In Chapter 2, it was shown that the statistical model behind the DSGE model is a stationary Normal VAR(2) model, which was found to be misspecified. In this chapter, the problem of respecification is considered. In general, this problem is addressed by using the M-S testing results, in conjunction with graphical techniques, to narrow down the set of possible models. The heterogeneous St-VAR ($3; \nu=3$) model is proposed as the respecified model to replace the stationary Normal VAR(2) model. The ultimate arbiter of whether one’s choice of an alternative (respecified) model is appropriate is thorough M-S testing to establish its statistical adequacy.
4.2 Why Not a GARCH Model?

The respecification of the stationary Normal VAR model aims to take into account all the systematic information in the data that has not been accounted for by the original model. Such information includes departures from Normality, homoskedasticity, independence (1st and 2nd order) as well as parameter $t$-invariance. In light of that, an obvious choice for most of the researchers is to estimate the generalized autoregressive conditional heteroskedastic (GARCH) model.

The main idea underlying the GARCH model is to supplement the standard autoregressive function with a conditional heteroskedastic autoskedastic function. The autoregressive function is estimated first to obtain the standardized residuals. The squared of these residuals are regressed on the regressors of the autoregressive function, which is deemed to be the heteroskedastic conditional variance-covariance equation (autoskedastic functions). In the case of the VAR model, such a conditional variance-covariance is also dynamic and is expected to capture the second order dependence as well, if present in the data.

Heracleous (2003) summarizes the following drawbacks of the GARCH models. First, the functional form of the conditional variance is ad hoc and subjective. Second, although both the conditional mean and the conditional variance come from the same joint distribution, there is no way in the GARCH models to trace back the parameters of the joint distribution from those of the conditional distribution because the two moments are modeled separately ignoring the potential relationship between the parameters of the conditional mean and the conditional variance (both sets of parameters must be the functions of the parameters of the joint distribution). Third, to make sure that the conditional variance is always positive, ad hoc restriction on the parameters of the conditional distribution are imposed, which can not be verified by the data. Moreover, these parameter restrictions can easily violate the
true parameter space defined within the statistical model in question. Finally and the most importantly, the assumption of Normality directly violates the concept of heteroskedastic covariance with linear mean. Therefore, the autoregressive function and the autoskedastic function have to be modeled together as a by-product of the same joint distribution with one to one correspondence between the parameters of the conditional and the joint distributions (Spanos, 2002). Bhattacharyya (1943) has suggested a non-Normal joint distribution, whose conditional distribution is non-linear in mean and heteroskedastic in variance.

4.3 Respecification Strategy

Respecification is done by taking into account the statistical information not accounted for by the stationary Normal VAR(2) model in Chapter 2. The stationary Normal VAR(2) model was derived as the implicit statistical model behind Ireland (2011) DSGE model. The obvious question that arises pertains to the probabilistic features of the data (growth rate of per-capita real GDP, inflation rate and interest rate) that the stationary Normal VAR(2) model could not account for.

In any misspecification (M-S) test, mere rejection of the null model does not suggest that the alternative model is statistically adequate. For example, the fact that the assumption of Markov (2) was rejected for the interest rate equation does not automatically mean that the VAR model with Markov (3) assumption is the statistically adequate model. One needs to start the respecification, in light of the M-S test results, from the very beginning by postulating a new model that might be potentially a statistically adequate model.

The M-S tests in Chapter 2 allow us to conclude the following about the stationary Normal VAR(2) model for the macroeconomic data we have chosen for this research.

(a) No departures from linearity is indicated.
(b) Departures from homoskedasticity is indicated in all three equations.

(c) Normality is strongly rejected. The presence of heteroskedasticity also points to this problem, because Normality (when combined with linearity) and heteroskedasticity are mutually exclusive within the elliptically symmetric family of distributions (Spanos, 1994).

(d) Lingering temporal dependence is indicated for the interest rate equation.

(e) Second order dependence is indicated in the inflation rate equation and the interest rate equation.

(f) Departures from the $t$-invariance show up in the autoregressive function of inflation rate and interest rate equations, and the autoskedastic functions of the growth rate and inflation rate equations.

The strategies adopted to tackle the aforementioned misspecification in the stationary Normal VAR(2) model are as follows

(a) To model heteroskedasticity, the heterogeneous St-VAR(3; $\nu=3$) is proposed in an attempt to preserve the linearity assumption. The choice of this particular distribution is also aimed at addressing the problem of non-Normality, especially indicated by high kurtosis values of the standardized residuals of the stationary Normal VAR(2) model.

(b) To model the first order and the second order temporal dependence more lags are introduced using Markov(3), and the heteroskedasticity of the autoskedastic functions of the Student’s $t$ distribution, respectively.

(c) To model the $t$-variance (heterogeneity), a polynomial of $t$ in the mean of the joint distribution is introduced. As the mean of the joint distribution is part of the conditional
mean as well as the conditional variance for the Student’s $t$ distribution, it is hoped to solve the issue of the departures from both the mean $t$-invariance and the variance $t$-invariance.

Table 4.1 summarizes the proposed respecification of the distribution, the dependence and the heterogeneity assumptions. These assumptions give rise to the heterogeneous St-VAR($3; \nu$) model. See Table 3.5 in Chapter 3 for its full specification.

<table>
<thead>
<tr>
<th>Original Model</th>
<th>M-S Test Results</th>
<th>Respecified Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAR(2)</td>
<td>Normal</td>
<td>St-VAR(3; $\nu = 3$)</td>
</tr>
<tr>
<td>Distribution (D)</td>
<td>Normal</td>
<td>Student’s $t$</td>
</tr>
<tr>
<td>Dependence (I)</td>
<td>Markov(2)</td>
<td>Markov(3)</td>
</tr>
<tr>
<td>Heterogeneity (H)</td>
<td>Homogeneous</td>
<td>$\mu_z(t) = \mu_0 + \mu_1 t + \mu_2 t^2$</td>
</tr>
</tbody>
</table>

### 4.4 Estimation Results

Note that $Z_t = [y_t \ p_t \ r_t]^T$ and $\mu_z(t) = [\mu_y(t) \ \mu_p(t) \ \mu_r(t)]^T$, where all $\mu_y(t), \mu_p(t)$ and $\mu_r(t)$ are specified to be second order polynomials in $t$. For estimation, the degrees of freedom is chosen to be $\nu=3$ on statistical adequacy grounds. In light of the fact that the low value of $\nu$ implies heteroskedastic autoskedastic functions, relatively low value of $\nu$ is used to capture the heteroskedasticity. Although $\nu = 1, 2$ is applicable, taking $\nu > 2$ implies that the first two moments of the joint distribution exist.

As there is no explicit solution to the first order condition of the maximization of the log-likelihood function in (3.39), a combination of the numerical optimization methods such as N-M (Nelder and Mead, 1965) and BFGS (Broyden, 1970) are used to ensure that the optimization procedure leads to the global optimum. The estimation result of the stationary
Normal VAR(2) model is presented side by side with the estimation result of the heterogeneous St-VAR \((l = 3; \nu = 3)\) model in Table 4.2. First, we compare the estimation results of the autoregressive function before comparing the autoskedastic functions.

### 4.4.1 Autoregressive Function

The most significant differences are indicated by boxes in Table 4.2. Although the autoregressive function of both the St-VAR model and the Normal VAR model are linear and have exactly the same parametrization, the difference in the estimation results (between the stationary Normal VAR(2) and the heterogeneous St-VAR\((3; \nu = 3)\)) is obvious. Differences include the signs and the magnitudes of the coefficients as well as their statistical significance, implying very different inferences and policy implications.

The first important distinct result of the heterogeneous St-VAR model is that the coefficients of \(t\) and \(t^2\) for the growth rate \((y_t)\) and interest rate \((r_t)\) equations are highly significant. This is expected as one can clearly see the trending inflation rate and interest rate in Figures 2.5 and 2.6. This implies that the non-stationary nature of data requires a different modeling strategy to avoid the misspecification. The heterogeneous feature of the data cannot be modelled adequately by the stationary VAR model even with higher number of lags. The departure from the \(t\)-invariance assumption does not go away even with 6 lags in the stationary VAR model for the data we have. This provides the evidence for the trending steady state values of \(y_t, r_t\) and \(p_t\).

Second, the interest rate equation has a statistically insignificant but positive intercept term in the stationary Normal VAR(2) model. But the same intercept is negative and statistically significant in the heterogeneous St-VAR\((3; \nu = 3)\) model.

Third, in the stationary Normal VAR(2) model, the coefficient of \(r_{t-1}\) is positive, small and
Table 4.2: Estimation Result of St-VAR and VAR

<table>
<thead>
<tr>
<th></th>
<th>St-VAR(3;(\nu=3))</th>
<th>VAR(2)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Growth((y_t))</td>
<td>Inflation((p_t))</td>
<td>Interest((r_t))</td>
<td>Growth((y_t))</td>
</tr>
<tr>
<td>1</td>
<td>0.493 [0.000]</td>
<td></td>
<td></td>
<td>0.526 [0.000]</td>
</tr>
<tr>
<td>(t)</td>
<td>-0.159 [0.000]</td>
<td>10.752 [0.206]</td>
<td>0.787 [0.000]</td>
<td></td>
</tr>
<tr>
<td>(t^2)</td>
<td>1.060 [0.000]</td>
<td>-46.848 [0.120]</td>
<td>-3.765 [0.000]</td>
<td></td>
</tr>
<tr>
<td>(y_{t-1})</td>
<td>0.285 [0.000]</td>
<td>0.022 [0.178]</td>
<td>0.023 [0.000]</td>
<td>0.305 [0.000]</td>
</tr>
<tr>
<td>(p_{t-1})</td>
<td>0.287 [0.023]</td>
<td>0.456 [0.000]</td>
<td>0.026 [0.146]</td>
<td>0.193 [0.191]</td>
</tr>
<tr>
<td>(r_{t-1})</td>
<td>-0.607 [0.061]</td>
<td>0.166 [0.187]</td>
<td>1.359 [0.000]</td>
<td>0.015 [0.360]</td>
</tr>
<tr>
<td>(y_{t-2})</td>
<td>0.110 [0.027]</td>
<td>0.004 [0.848]</td>
<td>0.005 [0.524]</td>
<td>0.096 [0.137]</td>
</tr>
<tr>
<td>(p_{t-2})</td>
<td>-0.238 [0.168]</td>
<td>0.176 [0.005]</td>
<td>0.012 [0.597]</td>
<td>-0.243 [0.104]</td>
</tr>
<tr>
<td>(r_{t-2})</td>
<td>0.273 [0.671]</td>
<td>-0.305 [0.199]</td>
<td>-0.680 [0.000]</td>
<td>-0.197 [0.582]</td>
</tr>
<tr>
<td>(y_{t-3})</td>
<td>-0.205 [0.001]</td>
<td>0.012 [0.585]</td>
<td>-0.003 [0.774]</td>
<td></td>
</tr>
<tr>
<td>(p_{t-3})</td>
<td>-0.446 [0.021]</td>
<td>0.222 [0.001]</td>
<td>-0.011 [0.672]</td>
<td></td>
</tr>
<tr>
<td>(r_{t-3})</td>
<td>0.215 [0.606]</td>
<td>0.167 [0.288]</td>
<td>0.265 [0.000]</td>
<td></td>
</tr>
</tbody>
</table>

Insensitive for the growth rate equation. But once one moves to the heterogeneous St-VAR (3;\(\nu=3\)) model, the coefficient of the same variable in the same equation becomes significant with high magnitude and a negative sign. This has a big policy implication. If a policy maker is to adopt the Normal VAR(2) model, which is statistically inadequate, the last quarter rise in the interest rate \((r_{t-1})\) by 1% will result into rise (if any) in the growth rate \((y_t)\) by 0.015%. On the other hand, if one were to adopt the heterogeneous St-VAR(3;\(\nu=3\)) model, if statistically adequate, the last quarter rise in the interest rate \((r_{t-1})\) by 1% leads to a fall in the growth rate \((y_t)\) by 0.61%.

Fourth, although the sign and the magnitudes of the coefficients of \(y_{t-2}\) and \(p_{t-1}\) in the growth rate equation do not differ noticeably between the two models, they are both statistically significant in the heterogeneous St-VAR(3;\(\nu=3\)) model but insignificant in the stationary Normal VAR(2) model.

Fifth, in the interest rate equation, the coefficient of \(y_{t-1}\) is same in both the models in terms of the sign and the magnitude. But much lower p-value in the heterogeneous St-
VAR(3;ν=3) model implies that the evidence for the non-zero coefficient is much stronger than in the stationary Normal VAR(2) model.

Sixth, in the interest rate equation, the coefficient of $r_{t-2}$ is significant in both the models with the same sign. But the magnitude of the coefficient in the heterogeneous St-VAR(3;ν=3) model is more than three times bigger than in the stationary Normal VAR(2) model.

Seventh, by adopting the statistically inadequate stationary Normal VAR(2) model, the chances of picking up the effects of the third lag coefficients are lost. Note that there are many significant coefficients of the third lag variables in the heterogeneous St-VAR(3;ν=3) model. For example, the heterogeneous St-VAR(3;ν=3) model indicates that $p_{t-3}$ has a significant negative effect on the growth rate of the current per-capita real GDP.

Last, but not the least, many other coefficients of the two models differ in the magnitude significantly. This can have significant impact on policy experiments such as the impulse response function (IRF) analysis.

It is important to note that unless we achieve the statistical adequacy of the heterogeneous St-VAR(3;ν=3) model separately, the inference results discussed above will still be unreliable. The complete M-S tests are done later in the chapter. The model passes all the tests. The fitted values and the standardized residuals of the heterogeneous St-VAR (3;ν=3) model are introduced in the next few sections because the M-S tests are based on them.

### 4.4.2 Fitted Values

Prediction in the Normal VAR model is limited to the estimated autoregression equation (conditional mean) as the autoskedastic function (conditional variance) is constant across time and regressors. The fitted value in the stationary Normal VAR(3) model would simply be

$$\tilde{Z}_t = \hat{A}_0 + \hat{A}_1Z_{t-1} + \hat{A}_2Z_{t-2} + \hat{A}_3Z_{t-3}$$

(4.1)
But the skedastic function in the heterogeneous St-VAR model changes with \( t \) and the conditioning variables (i.e. lagged terms). The prediction has to incorporate this feature of the model. In other words, the prediction has to be done not only for the mean, but also for the variance. The prediction inference will be incomplete without this new information about the heterogeneous and heteroskedastic autoskedastic functions. So a new definition of the fitted value is proposed as follows

\[
\hat{Z}_t = \hat{Z}_t + S_t, \quad S_t \sim St \left( 0, \frac{\nu + 3k - 2}{\nu + 3k} \text{Var}(Z_t|Z_{t-1}); \nu + lk \right)
\] (4.2)

where \( l \) is the number of lags and \( k \) is the number of variables in \( Z_t \).

The easiest way to see the difference between these two definitions of the fitted values (4.1 and 4.2) is to plot these two fitted values against the actual values (Figures 4.1-4.3). The bubbles are the actual values of the variable, the blue-dashed lines show the conventional fitted value \( \hat{Z}_t = [\hat{y}_t \hat{p}_t \hat{r}_t]^\top \), the red-solid lines show the newly defined fitted value, \( \hat{\hat{Z}}_t \), which takes the heterogeneous and heteroskedastic nature of the conditional variance into account.

The estimation results of the conditional variance-covariance (autoskedastic function) of the heterogeneous St-VAR(3;\( \nu = 3 \)) model is presented in the next subsection.

The difference between \( \hat{Z}_t \) and \( \hat{\hat{Z}}_t \) is most visible in growth rate plot (Figure 4.1) and inflation rate plot (Figure 4.2). \( \hat{\hat{Z}}_t \) is a lot more able to capture the volatility of the data. Because of the very low conditional variance for the interest rate, the difference between \( \hat{Z}_t \) and \( \hat{\hat{Z}}_t \) is not clear enough to be seen visually in the graph of the interest rate (Figure 4.3).
Figure 4.1: Fitted Values for Growth Rate of GDP Per Capita

Figure 4.2: Fitted Values for Inflation Rate
4.4.3 Forecasting: DSGE Model vs St-VAR Model

One way to evaluate a model is to perform forecasting and compare it with actual data. To compare the forecasting performance of the Ireland’s (2011) DSGE model and the heterogeneous St-VAR(3;\(\nu=3\)) model, both of them are estimated for the time period 1983:1 through 2008:4. And the \(s\)-time ahead forecasting is done for the period 2009:1-2012:1. The \(s\)-time ahead forecast is computed as follows

\[
\hat{Z}_{t+1} = \hat{A}_0 + \hat{A}_1Z_t + \hat{A}_2Z_{t-1} + \hat{A}_3Z_{t-2} + S_{t+1}
\]
\[
\hat{Z}_{t+2} = \hat{A}_0 + \hat{A}_1\hat{Z}_{t+1} + \hat{A}_2Z_t + \hat{A}_3Z_{t-1} + S_{t+2}
\]
\[
\hat{Z}_{t+3} = \hat{A}_0 + \hat{A}_1\hat{Z}_{t+2} + \hat{A}_2\hat{Z}_{t+1} + \hat{A}_3Z_t + S_{t+3}
\]
\[
\hat{Z}_{t+s} = \hat{A}_0 + \hat{A}_1 \hat{Z}_{t+s-1} + \hat{A}_2 \hat{Z}_{t+s-2} + \hat{A}_3 \hat{Z}_{t+s-3} + S_{t+s},
\]

where \( S_{t+s} \sim St(0, Var(Z_{t+s}|Z^0_{t+s-1}); \nu + lk) \). The forecasting performance of any statistical model can be judged in the following two steps.

(a) The forecasting errors have to be non-systematic (white-noise) in a statistical sense. If the forecasting errors exhibit the presence of systematic information, the forecasting is likely to be statistically unreliable.

(b) If the forecasting errors are non-systematic, then the mean squared prediction error (MSPE):

\[
MSEP = \frac{1}{s} \sum_{i=1}^{s} (\hat{Z}_{t+s} - Z_{t+s})^2
\]

(4.3)

can be used to compare the forecasting performance of a model. A model with lower MSEP can be deemed to be better forecaster than other models.

Figures 4.4-4.6 show the forecasting performance of the Ireland’s (2011) DSGE model (a misspecified model) and the heterogeneous St-VAR(3;\(\nu=3\)) model (a statistically adequate model). Forecasting for the DSGE model is done using the estimated reduced form (2.51). It can be easily seen that the forecasting performance of the statistically adequate model is a lot better than that of the DSGE model in both steps (a) and (b). The MSEP of the DSGE model and the heterogeneous St-VAR(3;\(\nu=3\)) model is shown in Table 4.3.

<table>
<thead>
<tr>
<th>Table 4.3: MSEP for 12 Periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth Rate</td>
</tr>
<tr>
<td>DSGE</td>
</tr>
<tr>
<td>St-VAR</td>
</tr>
</tbody>
</table>
Figure 4.4: Prediction of Growth Rate of GDP Per Capita

Figure 4.5: Prediction of Inflation Rate
The MSEP cannot be statistically meaningful unless the prediction errors are non-systematic. Hence the MSEP of the DSGE model cannot be interpreted meaningfully. The systematic prediction error of the DSGE model is one symptom of model misspecification. The prediction error of the St-VAR model is much more non-systematic. Still the improvement in the MSEP can be easily seen in Table 4.3.

4.4.4 Autoskedastic Function: Estimation

One important misspecification discovered in Chapter 2 was the violation of the second order $t$-invariance (second order homogeneity) assumption. It is expected to capture this heterogeneity without modeling the variance-covariance $\frac{\nu}{\nu-2} \Sigma$ of the joint distribution as a function of $t$. Although $\Sigma$ is constant across $t$ and $Z_{t-1}^0$, the conditional variance-covariance of the conditional distribution $D(Z_t|Z_{t-1}^0; \Theta_1)$ is a function of both $Z_{t-1}^0$ and $\mu_z(t)$. More-
over, the quadratic \( \mathbf{\mu}_z(t) \) implies that \( E(\mathbf{Z}_t|\sigma(\mathbf{Z}_{t-1}^0)) \) is quadratic in \( t \) and \( \text{Var}(\mathbf{Z}_t|\sigma(\mathbf{Z}_{t-1}^0)) \) is quartic in \( t \) — polynomial of degree 4 in \( t \). A polynomial of the fourth degree is hoped to be rich enough to capture the second order heterogeneity. The conditional covariance matrix is, by definition, heteroskedastic and dynamic in a St-VAR model, which implies that \( \text{Cov}(\mathbf{Z}_t|\sigma(\mathbf{Z}_{t-1}^0)) \) changes with \( \mathbf{Z}_{t-1}^0 \). This has important implications for macroeconomic modeling, for it allows the autoregressive function and the autoskedastic function to change not only with \( \mathbf{Z}_{t-1}^0 \) but also with \( t \). Using the notation of Chapter 3, the conditional covariance of \( y_t \) and \( r_t \) for the heterogeneous St-VAR(3;\( \nu=3 \)) model is presented as an example in equation (4.4)

\[
\text{Var}(y_t, r_t|\sigma(\mathbf{Z}_{t-1}^0)) = \frac{\nu}{\nu+1}(1 + 1)(q_{11}\tilde{y}_{t-1}^2 + q_{22}\tilde{p}_{t-1}^2 + q_{33}\tilde{r}_{t-1}^2 + q_{44}\tilde{y}_{t-2}^2 + q_{55}\tilde{p}_{t-2}^2 \nonumber \\
+ q_{66}\tilde{r}_{t-2}^2 + q_{77}\tilde{y}_{t-3}^2 + q_{88}\tilde{p}_{t-3}^2 + q_{99}\tilde{r}_{t-3}^2 + 2q_{12}\tilde{y}_{t-1}\tilde{p}_{t-1} + 2q_{13}\tilde{y}_{t-1}\tilde{r}_{t-1} + 2q_{14}\tilde{y}_{t-1}\tilde{y}_{t-2} \\
+ 2q_{15}\tilde{y}_{t-1}\tilde{r}_{t-2} + 2q_{16}\tilde{y}_{t-1}\tilde{y}_{t-2} + 2q_{17}\tilde{y}_{t-1}\tilde{y}_{t-3} + 2q_{18}\tilde{y}_{t-1}\tilde{p}_{t-3} + 2q_{19}\tilde{y}_{t-1}\tilde{r}_{t-3} + 2q_{23}\tilde{p}_{t-1}\tilde{r}_{t-1} \\
+ 2q_{24}\tilde{p}_{t-1}\tilde{r}_{t-2} + 2q_{25}\tilde{p}_{t-1}\tilde{p}_{t-2} + 2q_{26}\tilde{p}_{t-1}\tilde{y}_{t-3} + 2q_{27}\tilde{p}_{t-1}\tilde{r}_{t-3} + 2q_{28}\tilde{p}_{t-1}\tilde{p}_{t-3} + 2q_{29}\tilde{p}_{t-1}\tilde{r}_{t-3} \\
+ 2q_{33}\tilde{r}_{t-1}\tilde{y}_{t-2} + 2q_{34}\tilde{r}_{t-1}\tilde{p}_{t-2} + 2q_{35}\tilde{r}_{t-1}\tilde{r}_{t-2} + 2q_{36}\tilde{r}_{t-1}\tilde{r}_{t-3} + 2q_{37}\tilde{r}_{t-1}\tilde{y}_{t-3} + 2q_{38}\tilde{r}_{t-1}\tilde{p}_{t-3} + 2q_{39}\tilde{r}_{t-1}\tilde{r}_{t-3} \\
+ 2q_{45}\tilde{y}_{t-2}\tilde{p}_{t-2} + 2q_{46}\tilde{y}_{t-2}\tilde{r}_{t-2} + 2q_{47}\tilde{y}_{t-2}\tilde{y}_{t-3} + 2q_{48}\tilde{y}_{t-2}\tilde{p}_{t-3} + 2q_{49}\tilde{y}_{t-2}\tilde{r}_{t-3} + 2q_{56}\tilde{p}_{t-2}\tilde{r}_{t-2} \\
+ 2q_{57}\tilde{p}_{t-2}\tilde{r}_{t-3} + 2q_{58}\tilde{p}_{t-2}\tilde{p}_{t-3} + 2q_{59}\tilde{p}_{t-2}\tilde{y}_{t-3} + 2q_{60}\tilde{r}_{t-2}\tilde{p}_{t-3} + 2q_{61}\tilde{r}_{t-2}\tilde{r}_{t-3} + 2q_{62}\tilde{r}_{t-2}\tilde{r}_{t-3} \\
+ 2q_{63}\tilde{r}_{t-2}\tilde{r}_{t-3} + 2q_{64}\tilde{r}_{t-2}\tilde{r}_{t-3} + 2q_{65}\tilde{r}_{t-2}\tilde{r}_{t-3}) 
\]

(4.4)

where \( \tilde{y}_t = y_t - \mu_y(t) \), \( \tilde{p}_t = p_t - \mu_p(t) \), \( \tilde{r}_t = r_t - \mu_r(t) \) are the mean deviation of \( y_t, p_t, r_t \) from their respective means, all quadratic in \( t \). In Table 4.4, the maximum likelihood estimates of the conditional variance-covariances for the stationary Normal VAR(2) model is presented.

<table>
<thead>
<tr>
<th>Table 4.4: Normal VAR(2): Conditional Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{Var}(y_t</td>
</tr>
<tr>
<td>( \text{Var}(p_t</td>
</tr>
<tr>
<td>( \text{Var}(r_t</td>
</tr>
</tbody>
</table>
Equation (4.5) is the estimated autoskedastic function of \( y_t \) for the heterogeneous St-VAR(3; \( \nu = 3 \)) model. The numbers in the parenthesis below the estimated coefficients are p-values. The other estimated autoskedastic functions are placed in Appendix E. Note that the coefficients of different variances and covariances differ only in magnitude and sign because they all have the same quadratic form \( q(\mathbf{Z}_{t-1}) \) multiplied by different \( \omega \), an element of \( \Omega \).

\[
\begin{align*}
\widetilde{Var}(y_t | \mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \mathbf{Z}_{t-3}) &= 0.133 + 0.104\hat{y}_{t-1}^2 + 0.131\hat{y}_{t-1}\hat{p}_{t-2} - 0.344\hat{y}_{t-1}\hat{r}_{t-1} - 0.054\hat{y}_{t-1}\hat{y}_{t-2} \\
-0.099\hat{y}_{t-1}\hat{p}_{t-2} + 0.562\hat{y}_{t-1}\hat{r}_{t-2} - 0.011\hat{y}_{t-1}\hat{y}_{t-3} + 0.056\hat{y}_{t-1}\hat{p}_{t-3} - 0.207\hat{y}_{t-1}\hat{r}_{t-3} + 0.773\hat{p}_{t-1}^2 \\
-0.866\hat{p}_{t-1}\hat{r}_{t-1} - 0.041\hat{p}_{t-1}\hat{y}_{t-2} - 0.799\hat{p}_{t-1}\hat{p}_{t-2} + 1.074\hat{p}_{t-1}\hat{r}_{t-2} - 0.019\hat{p}_{t-1}\hat{y}_{t-3} - 0.377\hat{p}_{t-1}\hat{p}_{t-3} \\
-0.303\hat{p}_{t-1}\hat{r}_{t-3} + 4.885\hat{r}_{t-1}^2 - 0.160\hat{y}_{t-1}\hat{y}_{t-2} - 0.179\hat{y}_{t-1}\hat{p}_{t-2} - 12.499\hat{y}_{t-1}\hat{r}_{t-2} - 0.074\hat{y}_{t-1}\hat{y}_{t-3} \\
+0.155\hat{y}_{t-1}\hat{p}_{t-3} - 3.428\hat{y}_{t-1}\hat{r}_{t-3} + 0.111\hat{y}_{t-2}^2 + 0.176\hat{y}_{t-2}\hat{p}_{t-2} - 0.111\hat{y}_{t-2}\hat{r}_{t-2} - 0.047\hat{y}_{t-2}\hat{y}_{t-3} \\
-0.073\hat{y}_{t-2}\hat{p}_{t-3} - 0.278\hat{y}_{t-2}\hat{r}_{t-3} + 0.934\hat{p}_{t-2}^2 - 0.929\hat{p}_{t-2}\hat{r}_{t-2} - 0.033\hat{p}_{t-2}\hat{y}_{t-3} - 0.794\hat{p}_{t-2}\hat{p}_{t-3} \\
+0.665\hat{p}_{t-2}\hat{r}_{t-3} + 12.288\hat{r}_{t-2}^2 - 0.359\hat{r}_{t-2}\hat{y}_{t-3} - 0.064\hat{r}_{t-2}\hat{p}_{t-3} - 12.454\hat{r}_{t-2}\hat{r}_{t-3} + 0.100\hat{y}_{t-3}^2 \\
0.130\hat{y}_{t-3}\hat{p}_{t-3} + 0.264\hat{y}_{t-3}\hat{r}_{t-3} + 0.783\hat{p}_{t-3}^2 - 0.254\hat{p}_{t-3}\hat{r}_{t-3} + 4.827\hat{y}_{t-3}^2 \\
\end{align*}
\]

\[ (4.5) \]

4.4.5 Autoskedastic Function: Interpretation

Unlike the interpretation of the autoregressive function, which is linear, the interpretation of the autoskedastic functions, which is quadratic, needs more careful consideration. \( Var(y_t | \mathbf{Z}_{t-1}) \) is taken as an example to illustrate the interpretation of the autoskedastic functions. First, we need to be able to interpret the quadratic terms on the right hand side of equation (4.4) or (4.5), which are written in terms of deviation from the quadratic trending mean. Basically, the right hand side of equation (4.5) has following two types of terms
(a) The squared deviation from the trending mean (e.g. $\tilde{y}_{t-1} = (y_{t-1} - \mu_y(t-1))^2$): This term represents deviation (volatility), if any, of the growth rate of per-capita real GDP in the last quarter ($y_{t-1}$) from its mean. Whenever $y_{t-1}$ goes above or below the quadratically trending mean $\mu_y(t-1)$, $\tilde{y}_{t-1}^2$ becomes a positive quantity and affects $Var(y_t|Z_{t-1})$ positively (negatively) if the coefficient of $\tilde{y}_{t-1}^2$ is positive (negative) (positively in the case of equation (4.5) as the sign of $\tilde{y}_{t-1}$ is positive). The same kind of interpretation can be given to other quadratic terms.

(b) The cross product of deviation from the trending mean (e.g. $\tilde{y}_{t-1}\tilde{p}_{t-2} = (y_{t-1} - \mu_y(t-1))(p_{t-2} - \mu_p(t-2))$): This term is positive if both the growth rate ($y_{t-1}$) at time $t-1$ and the inflation rate ($p_{t-2}$) at time $t-2$ are either below or above their respective quadratic trends $\mu_y(t-1)$ and $\mu_p(t-2)$. On the other hand, if the growth rate ($y_{t-1}$) is above (below) its trend and inflation rate ($p_{t-2}$) is below (above) its trend, then this cross product term $\tilde{y}_{t-1}\tilde{p}_{t-2}$ becomes negative. Consequently, its effect on $Var(y_t|Z_{t-1})$ depends on whether its coefficient is negative or positive.

Now the following inferences are drawn from the estimated autokedastic functions. First, the estimation results of the autokedastic functions (conditional variance-covariance) show that all the variances $Var(y_t|Z_{t-1})$, $Var(p_t|Z_{t-1})$ and $Var(r_t|Z_{t-1})$ and the two covariances $Cov(y_t,r_t|Z_{t-1})$ and $Cov(p_t,r_t|Z_{t-1})$ increase with the increase in volatility in each of the previous three quarters – clearly indicated by highly significant and positive coefficients of the squared mean deviations such as $\tilde{y}_{t-1}^2$, $\tilde{p}_{t-1}^2$, $\tilde{r}_{t-1}^2$, $\tilde{y}_{t-2}^2$ and $\tilde{p}_{t-3}^2$. This suggests the positive dependence of volatility for all three variables: growth rate ($y_t$), inflation rate ($p_t$) and interest rate ($r_t$). Note that the signs of the coefficients of the autokedastic functions are crucially determined by the sign of the contemporaneous covariance matrix $\Omega$

Second, the conditional covariance $Cov(y_t,p_t|Z_{t-1})$ is negatively dependent on past volatility implied by highly significant and negative coefficients of the squared deviations such as
$\tilde{y}_{t-1}^2, \tilde{p}_{t-2}^2$ and $\tilde{r}_{t-3}^2$.

Third, all the six intercept terms are highly significant suggesting the existence of volatility in both the variances and the covariances even in the absence of any past departures of the variables from their quadratic trends. The intercept of $Cov(y_t, p_t | Z_{t-1}^0)$ is the only negative intercept showing negative covariance between growth rate and inflation rate.

Finally, just like the fitted values for the autoregressive function, one can see the $t$-plots of these variances and covariances (Figures 4.7-4.12). Taking Figure 4.7 as an example, it can be seen that the volatility is a lot higher before mid-1980s. The oil crisis shows up in spikes of volatility during the 1970s up to the mid 1980s. The plot also shows the signs of increased volatility ignited by the financial crisis of 2008. All the variances and covariances, except $Cov(y_t, p_t | Z_t^0)$ are positive. We can notice one important feature of the conditional variance-covariance in these figures: if a variance or covariance is positive (negative) it remains positive (negative) forever.
Figure 4.7: Fitted Conditional Variance for Growth Rate

Figure 4.8: Fitted Conditional Variance for Inflation Rate
Figure 4.9: Fitted Conditional Variance for Interest Rate

Figure 4.10: Fitted Conditional Covariance for Growth Rate and Inflation Rate
Figure 4.11: Fitted Conditional Covariance for Growth Rate and Interest Rate

Figure 4.12: Fitted Conditional Covariance for Inflation Rate and Interest Rate
4.5 M-S Testing for Heterogeneous St-VAR Model

When the M-S tests rejected the stationary Normal VAR(2) model in Chapter 2, they pointed towards a more dynamic, heterogeneous and heteroskedastic model. As a respecification strategy, the heterogeneous St-VAR(3; $\nu=3$) model is proposed and estimated in this chapter. But its statistical adequacy has not been established as yet. Before we can use it for inference purpose reliably, its own statistical adequacy has to be secured first. To do this, a new set of auxiliary equations for the new M-S tests, taking the heterogeneous St-VAR(3; $\nu=3$) model as the null model, has to be developed. In the next subsection, the standardized residuals of the St-VAR model are defined before developing the new auxiliary regressions based on these standardized residuals.

4.5.1 Standardized Residuals

The heteroskedastic and heterogeneous conditional variance-covariance suggests the standardized residuals

$$
\hat{u}_t = \begin{bmatrix}
\hat{u}_{yt} \\
\hat{u}_{pt} \\
\hat{u}_{rt}
\end{bmatrix} = L_t^{-1}(Z_t - \hat{Z}_t)
$$

where $L_tL_t^\top = \hat{\text{Cov}}(Z_t | \sigma(Z_{t-1}^0))$. Here, $L_t$ is changing with $t$ and $Z_{t-1}^0$ as opposed to the constant conditional variance-covariance in the case of the stationary Normal VAR model. This shows the importance of having the properly standardized residuals for M-S testing purposes.

In the St-VAR model, standardization of the residual is done using different weights for different $t$, the weight being the reciprocal of the conditional variance-covariance. This is
Figure 4.13: Unstandardised Growth Rate Residual

Figure 4.14: Standardised Growth Rate Residual
Figure 4.15: Unstandardised Inflation Rate Residual

Figure 4.16: Standardised Inflation Rate Residual
Figure 4.17: Unstandardised Interest Rate Residual

Figure 4.18: Standardised Interest Rate Residual
completely different from the standardization of the residuals from the Normal VAR model, where the weights are constant throughout. The difference between the standardized and the non-standardized residuals is most visible in Figures 4.13-4.18. The non-standardized residuals of the St-VAR(3;\(\nu=3\)) model (Figures 4.13, 4.15 and 4.17) have exactly the same probabilistic structure as the residuals of the stationary Normal VAR(2) model. It can be clearly seen that the homoskedastic and homogeneous variance is unsuitable for the estimated residuals. With the standardized residuals of the St-VAR(3,\(\nu=3\)) model in Figures 4.14, 4.16 and 4.18, it can be seen that the changing conditional variance has appropriately standardized the residuals. For example, let us take the residuals of the interest rate equation (Figures 4.14 and 4.15). When the volatility in the non-standardized residuals is high, the conditional variance also becomes high and lowers the standardized residuals. Similarly, when the volatility is low, the standardized residuals increase. Overall, the standardized residuals are adjusted differently at different places depending on the volatility of the data. This is one of the important aspects of a heteroskedastic and heterogeneous model.

Standardized residuals reject frequently claimed great moderation in US economy after 1980s (Blanchard and Simon, 2001; Summers, 2005). It clearly shows why a statistically inadequate model cannot be used to learn from data. Although the difference between the autoregressive functions of VAR and St-VAR models is not significant, the inference drawn from them can be miles apart.

The non-standardized residuals from the St-VAR model is very close to those of the Normal VAR model, partially due to the fact that they have the same form of the autoregressive function. The standardized residual from the Normal VAR model is mere a rescaled version of the non-standardized residual because the conditional variance is both homogeneous and homoskedastic. Hence the standardization does not have any real effect on the M-S tests results in the Normal VAR model. But the standardization has a significant effect on the M-
S tests for the St-VAR model because the residuals are rescaled by taking heteroskedasticity into account.

### 4.5.2 M-S Testing Auxiliary Equations for St-VAR Model

The auxiliary regressions for the M-S tests are built in an exactly the same way as it was done for the stationary Normal VAR(2) model in Chapter 2. The auxiliary regressions will contain all the terms from the right hand side of the conditional mean (or the conditional variance). The left hand side of the auxiliary regressions will contain different functions of the standardized residuals. The extra terms representing the potential departures from the null model (i.e. St-VAR(3;ν=3) are added to the right hand side of the auxiliary regressions and checked if they are statistically significant. The significant coefficients of the extra terms point to the potential departures and the insignificant coefficients indicate the lack of departures from the null model. As an example, a set of auxiliary regressions for the growth rate residuals is presented below in equations (4.7) and (4.8).

\[
\begin{align*}
\hat{u}_{yt} &= a_0 + a_1 \hat{u}_{yt-1} + a_2 \hat{u}_{yt-2} + b_1 \hat{y}_t + b_2 \hat{y}_t^2 + b_3 t^3 + b_4 t^4 + v_t \\
\hat{\sigma}^2_{yt} &= c_0 + c_1 \hat{\sigma}_{yt} + c_2 \hat{\sigma}_t^2 + c_3 \hat{\sigma}_{yt-1} + c_5 \hat{\sigma}_{yt-2} + d_1 t^5 + d_2 t^6 + v_t
\end{align*}
\]

(4.7)  
(4.8)

where

\[
\begin{bmatrix}
\hat{y}_t \\
\hat{p}_t \\
\hat{r}_t
\end{bmatrix}
= \hat{\delta}_0 + \hat{\delta}_1 t + \hat{\delta}_2 t^2 + \hat{A}_1^T Z_{t-1} + \hat{A}_2^T Z_{t-2} + \hat{A}_3^T Z_{t-3}
\]

\[
\begin{bmatrix}
\hat{y}_t \\
\hat{p}_t \\
\hat{r}_t
\end{bmatrix}
= \hat{\delta}_0 + \hat{A}_1^T Z_{t-1} + \hat{A}_2^T Z_{t-2} + \hat{A}_3^T Z_{t-3}
\]

\[
\hat{\sigma}^2_{yt} = \text{var}(y_t|\sigma(Z^0_{t-1}))
\]
\[ \widehat{\sigma}_{yt}^2 = \left( \frac{\nu}{\nu+3k-2} \right) \tilde{\omega}_y[1 + \frac{1}{\nu}(Z_{t-1}^0 - \hat{\mu}_0)\hat{Q}^{-1}(Z_{t-1}^0 - \hat{\mu}_0)] \]

In the auxiliary regression (4.7), \( \widehat{y}_t \) (fitted value) represents the linear combination of the terms in the conditional mean from the null model. \( \widehat{y}_t \) represents the fitted values minus the trend terms so that \( \widehat{y}_t^2 \) represents the pure departure from the linearity assumption. Similarly, in equation (4.8), \( \widehat{\sigma}_t \) represents the linear combination of the quadratic terms on the right hand side of the conditional variance \( \widehat{\text{var}}(y_t|\sigma(Z_{t-1}^0)) \). \( \widehat{\sigma}_{yt} \) represents the estimated \( \widehat{\text{var}}(y_t|\sigma(Z_{t-1}^0)) \) minus the trend components. In other words, \( \widehat{\sigma}_{yt} \) represents the pure heteroskedastic (i.e. the terms depending only on \( Z_{t-1}^0 \)) term of the conditional variance so that \( \widehat{\sigma}_{yt}^2 \) represents pure departure from the assumption of quadratic heteroskedasticity. This strategy on both equations (4.7) and (4.8) allows us to test the \( t \)-invariance assumption separately from the assumption of heteroskedasticity and the assumption second order dependence. Two different distributional tests are applied with an aim to detect the departures from the Student’s \( t \). The test results are shown in Table 4.7. Details of the tests applied are shown in Appendix F.

### 4.5.3 M-S Testing Results for St-VAR Model

Table 4.5 shows the null hypotheses and the M-S tests to be conducted. All the tests in Table 4.5 are \( F \)-tests on the auxiliary regressions such as (4.7) and (4.8). The numbers inside the parenthesis of \( F \) are the degrees of freedom for numerator and denominator respectively. The results of the M-S tests done are shown in Tables 4.6 and 4.7.

As indicated by the high p-values in the parentheses of Tables 4.6 and 4.7, the M-S tests indicate no serious departures from the probabilistic assumptions comprising the St-VAR model. Moreover, the heterogeneous St-VAR(3; \( \nu=3 \)) model is able to capture the second order temporal dependence in the data. Note that the respecification has not altered the
Table 4.5: M-S Tests for Heterogeneous St-VAR(3; ν = 3)

<table>
<thead>
<tr>
<th>Tests</th>
<th>Hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linearity F(237,1)</td>
<td>$H_0 : b_2 = 0$</td>
</tr>
<tr>
<td>t-invariance F(237,2)</td>
<td>$H_0 : b_3 = b_4 = 0$</td>
</tr>
<tr>
<td>Independence F(237,2)</td>
<td>$H_0 : a_1 = a_2 = 0$</td>
</tr>
<tr>
<td>Heteroskedasticity F(237,1)</td>
<td>$H_0 : c_2 = 0$</td>
</tr>
<tr>
<td>2nd Order Independence F(237,2)</td>
<td>$H_0 : c_4 = c_5 = 0$</td>
</tr>
<tr>
<td>2nd Order t-invariance F(237,2)</td>
<td>$H_0 : d_1 = d_2 = 0$</td>
</tr>
</tbody>
</table>

Structural equations (2.50) of the original DSGE model. It only extended the original statistical model (Table 2.3) but retained the parametric nesting of the structural model (2.50), giving rise to some additional overidentifying restrictions that can be tested. The likelihood ratio test can be performed to test the validity of the original structural model through the overidentifying restrictions.

Table 4.6: M-S Tests for St-VAR Model

<table>
<thead>
<tr>
<th></th>
<th>Growth Rate ($y_t$)</th>
<th>Inflation Rate ($p_t$)</th>
<th>Interest Rate ($r_t$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linearity</td>
<td>2.369 [0.125]</td>
<td>0.002 [0.963]</td>
<td>1.380 [0.241]</td>
</tr>
<tr>
<td>Homoskedasticity</td>
<td>1.765 [0.185]</td>
<td>0.201 [0.655]</td>
<td>0.106 [0.045]</td>
</tr>
<tr>
<td>1st Independence</td>
<td>0.461 [0.633]</td>
<td>0.429 [0.652]</td>
<td>2.649 [0.073]</td>
</tr>
<tr>
<td>2nd Independence</td>
<td>0.010 [0.905]</td>
<td>1.755 [0.175]</td>
<td>0.936 [0.181]</td>
</tr>
<tr>
<td>1st t-invariance</td>
<td>0.690 [0.502]</td>
<td>2.190 [0.114]</td>
<td>0.616 [0.541]</td>
</tr>
<tr>
<td>2nd t-invariance</td>
<td>0.824 [0.440]</td>
<td>1.766 [0.466]</td>
<td>1.720 [0.394]</td>
</tr>
</tbody>
</table>

Table 4.7: Skewness-Kurtosis and Kolmogorov-Smirnov Tests

<table>
<thead>
<tr>
<th></th>
<th>$y_t$</th>
<th>$p_t$</th>
<th>$r_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skewness-Kurtosis SK(2)</td>
<td>2.059[0.357]</td>
<td>3.200[0.202]</td>
<td>1.349[.510]</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov (KS)</td>
<td>0.042[0.426]</td>
<td>0.045[0.374]</td>
<td>0.046[0.358]</td>
</tr>
</tbody>
</table>

Having established the statistical adequacy of the heterogeneous St-VAR(3; ν=3) model, one can proceed to reliable inferences relating to various questions of interest including the
4.6 Impulse Response Function (IRF) Analysis

The estimated autoregressive function is a third order linear difference equation and the estimated autoskedastic function (conditional variance) is a quadratic difference equation. One important policy experiment to do is to solve these difference equations to derive the time path of $\hat{Z}_t$ and $\hat{Var}(Z_t|\sigma(Z_{t-1}))$. In the stationary Normal VAR model, however, the time path of the autoskedastic function is unnecessary because it is homoskedastic and homogeneous (or constant).

The IRF is a way to visualize these time paths. It shows by how much does the left hand side variable in the difference equation changes for some time periods if we change one of the right-hand side variables by one unit keeping every other variables constant. The IRF tracks the impact of any variable on the rest of the variables in the system. The effectiveness of a policy change such as a change in interest rate $r_t$ can be evaluated using the IRF.

Since the error terms in the error vector $u_t$ are correlated with each other, the effect of one shock, say $u_{gt}$, on rest of the variables, $y_t, p_t$ and $r_t$, cannot be computed in isolation. To get rid of the problem, Sims (1980) suggests to estimate the recursive VAR (structural VAR) model using Cholesky decomposition of the conditional variance-covariance matrix of the unrestricted Normal VAR model. In the St-VAR model, such a decomposition is to be done for each $t$ because of the heteroskedastic and heterogeneous nature of the conditional variance-covariance. In this research, the issue of Cholesky decomposition in the St-VAR model environment is ignored and the IRF is computed without any structural restrictions. Since the main objective, here, is to show the difference between the IRF of the stationary impulse response function (IRF) analysis, goodness of fit, Neymann-Pearson tests, prediction and other policy experiments that can potentially shed light on the phenomenon of interest.
Normal VAR(2) (statistically inadequate) model and the heterogeneous St-VAR(3; ν=3) (statistically adequate) model, avoiding the recursive estimation does not pose major obstacle. Since the conditional variance is heteroskedastic, IRF for both the autoregressive function and the autoskedastic function contain meaningful information. Let us call them mean IRF and variance IRF respectively.

(a) Mean IRF: The mean IRF is in the form of an ordered pair \((s, \hat{Z}_s)\). If we are to see the response of a rise in interest rate \(r_t\) on the rest of the variables over the time, we derive the mean IRF in following steps.

(i) Set \(Z_{t-2} = Z_{t-3} = 0, Z_{t-1} = [y_{t-1}, p_{t-1}, r_{t-1}] = [0, 0, \epsilon_0]\) where \(\epsilon_0\) is a small number.

(ii) Compute \(\hat{Z}_s\) for \(s = t + i, i = 0, 1, 2, ...\) to get the mean IRF \((s, \hat{Z}_s)\)

(b) Variance IRF: Similarly, the ordered pair \((s, \sqrt{Var}(Z_s|Z_{s-1}^0))\) is defined as a variance IRF.

4.6.1 Mean Impulse Response Function

For comparative purpose, the mean IRF of the stationary Normal VAR(2) model and the heterogeneous St-VAR(3; ν=3) model are plotted in the same graph (Figures 4.19 - 4.27). There is a big difference between the mean IRF from the VAR model and the St-VAR model implying different policy recommendations. Each mean IRF is produced by increasing one of the variables at time period \(t-1\) by 1%, keeping the other variables of time \(t-1, t-2\) and \(t-3\) constant at 0. Note that the trend of the heterogeneous St-VAR model is varying as opposed to the constant trend of the stationary Normal VAR model. Comparative interpretation of the mean IRF from the two models is presented as follows.

(a) Growth rate shock (Rise in \(y_{t-1}\) by 1%):

The growth rate declines at similar rates in the two models up to the third quarter.
But in the fourth quarter, the heterogeneous St-VAR($3;\nu=3$) model suggests a sharper decline in the growth rate before it starts rising in the fifth quarter and stabilizing around the trend from the $9^{th}$ quarter. But in the stationary Normal VAR model, the growth rate asymptotically reaches the constant trend at around the $7^{th}$ quarter. The effect of one percent rise in the growth rate on the inflation rate and the interest rate is much smaller in the heterogeneous St-VAR model than in the stationary Normal VAR model. The inflation rate and the interest rate converge much faster to its steady state (trend) in the heterogeneous St-VAR model.

(b) Inflation rate shock (Rise in $p_{t-1}$ by 1%):

As a result of a rise in the inflation rate by 1%, both the heterogeneous St-VAR model and the stationary Normal VAR model produce similar effects on the growth rate, the inflation rate and the interest rate for the first few quarters. The heterogeneous St-VAR model suggests a sharper decline after the third quarter and a sharper recovery after the fifth quarter in the growth rate. Almost similar effects on the future inflation rates are produced by both the models for up to the 25 quarters, though the effects produced by the heterogeneous St-VAR model is more volatile between the $2^{nd}$ and the $5^{th}$ quarter. The response by the interest rate is more volatile in the stationary Normal VAR model.

(c) Interest rate shock (Rise in $r_{t-1}$ by 1%):

The heterogeneous St-VAR model produces a sharper decline and a sharper recovery in the growth rate of per-capita real GDP. This indicates stronger evidence for the effectiveness of the monetary policy. After some quarters of sharp decline, the growth rate for some time rises above the trend before falling below the trend again. But the effects produced by the stationary Normal VAR model is completely different. The growth rate smoothly falls and sluggishly recovers. The effects on the inflation rate are also significantly different in the two models (see Figure 4.26). Both the models produce
similar effects on the interest rate.
Figure 4.19: Effect of 1% Rise in Growth Rate on Growth Rate

Figure 4.20: Effect of 1% Rise in Growth Rate on Inflation Rate
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8
0 5 10 15 20 25
0.00 0.02 0.04 0.06
Quarters
Interest Rate
Trend(VAR)
IRF(VAR)
Trend(St−VAR)
IRF(St−VAR)

Figure 4.21: Effect of 1% Rise in Growth Rate on Interest Rate

0 5 10 15 20 25
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8
Quarters
Growth Rate
Trend(VAR)
IRF(VAR)
Trend(St−VAR)
IRF(St−VAR)

Figure 4.22: Effect of 1% Rise in Inflation Rate on Growth Rate
Figure 4.23: Effect of 1% Rise in Inflation Rate on Inflation Rate

Figure 4.24: Effect of 1% Rise in Inflation Rate on Interest Rate
Figure 4.25: Effect of 1% Rise in Interest Rate on Growth Rate

Figure 4.26: Effect of 1% Rise in Interest Rate on Inflation Rate
4.6.2 Variance Impulse Response Function

Figures 4.28-4.30 show the effect of 1% rise in the growth rate, the inflation rate and the interest rate on the conditional variance of the growth rate, the inflation rate and the interest rate respectively. The effects on other variance-covariances will not be different except in scaling. The plots of the variance IRF show that the rise in the growth rate by 1% has permanent and highest effect on the volatility or the conditional variance of the model as the conditional variance never returns to pre shock-level. Moreover, the conditional variance seems to increase for up to the 25 quarters. Severe rises and falls in the volatility are seen up to the fifth quarters. For the growth rate and the inflation rate shock, the volatility remains at higher level for few quarters after the shock. The interest rate shock creates a sharp peak of the conditional variance. For the inflation rate and the interest rate shock, the volatility returns back to almost pre-shock level after 10 quarters.
Figure 4.28: Effect of 1% Rise in Growth Rate on Conditional Variance

Figure 4.29: Effect of 1% Rise in Inflation Rate on Conditional Variance
4.7 Testing the Theory

Using the estimation results in Table 2.2 of the structural model (2.41)-(2.42) and the model’s reduced form (2.50), the coefficients $\hat{\Psi}_0$, $\hat{\Psi}_1$ and $\hat{\Psi}_2$ are obtained in equation (2.51). The structural model in equation (2.43) can be viewed as a result of the structural restrictions on the autoregressive function of the St-VAR model. The same heterogeneous St-VAR(3; $\nu$=3) model can be re-estimated by imposing these restrictions on its autoregressive function. To show that the restrictions are not valid for the data through the likelihood ratio test, the restrictions imposed by the structural model on the autoskedastic function can be ignored because if the test rejects the overidentifying restrictions by ignoring the restrictions on the autoskedastic function, it is even more likely that the structural model will be rejected more strongly when the restrictions on the autoskedastic function are taken into account. The 27 restrictions on the autoregressive function to be tested are as follows
\[ A_1^\top = \hat{\Psi}_1, \quad A_2^\top = \hat{\Psi}_2, \quad A_3^\top = 0. \]

While doing the likelihood ratio test, the data set used covers the same time period 1983:1 through 2009:4 with 108 observations. The likelihood ratio test is defined as

\[ LR = 2[l_T(\hat{\Theta}_{H_1}) - l_T(\hat{\Theta}_{H_0})], \quad (4.9) \]

where \( l_T(\hat{\Theta}) \) is the maximized log-likelihood value under a given scenario (\( H_0 \) or \( H_1 \)). The \( LR \) is distributed as \( \chi^2 \) with 27 degrees of freedom. The result of the likelihood ratio test is presented in Table 4.8. The tiny p-value from the likelihood ratio test suggests that the Ireland’s DSGE model is strongly rejected by the data. Using this kind of model for any inference purpose will be highly misleading.

<table>
<thead>
<tr>
<th>Log-likelihood under</th>
<th>( H_1 )</th>
<th>( H_0 )</th>
<th>( LR )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5.093</td>
<td>-560.900</td>
<td>1111.200</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

The likelihood ratio test for Ireland (2004) model is presented in Appendix C. Sample size for the test in this model is 220 and there are 27 restrictions to be tested. The results in Table C.1 shows the rejection of the structural model by the data.

## 4.8 Identification Through Simulation

One of the important issues raised in the DSGE literature is the lack of proper identification when the structural DSGE model is not estimated directly. There is no easy way to go from the statistical parameters, parameters of the heterogeneous St-VAR model \( \theta_2 \), to the structural parameters, parameters of the DSGE model \( \psi_1 \), because the implicit function
\( G(\theta_2, \psi_1) = 0 \) is not only highly non-linear but also involves many algorithms such as the Schur decomposition of the structural matrices involved.

To get around this problem, identification strategy based on the simulation of the observations from the statistically adequate model is adopted. We already have the heterogeneous St-VAR(3;\( \nu = 3 \)) model as a statistically adequate model capturing the dynamics of the given data on growth rate, inflation rate and interest rate. Once the data is simulated from the statistically adequate heterogeneous St-VAR(3;\( \nu = 3 \)) model, the data is used to re-estimate the parameters of the Ireland’s (2011) structural DSGE model. If the distribution of the estimated parameters have a narrow interval of support, then these parameters can be considered identified. If the estimated parameters are scattered over a large range of values, then they are under identified. The simulation technique used is discussed in the following subsections.

### 4.8.1 Simulation

Observations are simulated for the variables: growth rate \((y_t)\), inflation rate \((p_t)\) and interest rate \((r_t)\). The simulated vectors of these variables, denoted by \(Z_t^* = [y_t^* \ p_t^* \ r_t^*]^\top\), are generated using the estimated parameters of the heterogeneous St-VAR (3;\( \nu = 3 \)) model as follows

\[
Z_t^* = \delta_0 + \delta_1 t + \delta_2 t^2 + \hat{A}_1^\top Z_{t-1} + \hat{A}_2^\top Z_{t-2} + \hat{A}_3^\top Z_{t-3} + u_t^*, \quad t \in \mathbb{N},
\]

\[
u + 3k - 2Var(u_t^*|\sigma(Z_t^0))= \left( \frac{\nu}{\nu+3k-2} \right) \Omega \left\{ \left[ 1+\frac{1}{\nu}(Z_{t-1}^0 - \hat{\mu}_{tk}(t))\hat{Q}^{-1}(Z_{t-1}^0 - \hat{\mu}_{tk}(t)) \right] \right\}.
\]

The estimates used for the simulation are estimated using the data from the whole sample period (1948:1–2010:3). The simulated sample size is 121. Changing the sample size does
not change the results much unless the sample size provides enough degrees of freedom to estimate all the parameters of the heterogeneous St-VAR(3;ν=3) model.

### 4.8.2 Simulation Results

The data is simulated 1218 times from (4.10). The simulated data is then used to re-estimate the 10 structural parameters of the Ireland’s (2011) model and 12 structural parameters of Ireland (2004) model separately. The histograms of the estimates are in Appendix D. The histograms show that the parameters are identified within a narrow interval of the support for Ireland (2011) model (Figures D.1-D2). Although we do not have explicit identifying restrictions going from the St-VAR(3;ν=3) model to the Ireland’s structural DSGE model, the simulation helped to know that the parameters are exactly identified. But some of the parameters of Ireland (2004) model are not exactly identified, which can be seen in histograms in Appendix D (Figures D.3-D.4). This indicates that Ireland’s (2011) strategy to identify the parameters has worked. However, this strategy does not guarantee the statistical adequacy.

When comparing the DSGE models in Ireland (2004) and Ireland (2011), it has it be noted that the one does not nest the other completely. For example, the Taylor rule in Ireland (2011) paper is a special case of the Taylor rule in Ireland (2004) paper. On the other hand, the price adjustment equation in Ireland (2004) is a special case of the price adjustment equation in Ireland (2011). Moving from Ireland (2004) model to Ireland (2011) model one has to impose four parameter restrictions and add two more parameters (number of parameters reduces from 14 to 12). The difference between Ireland’s two models (2004 and 2011) is discussed in detail in Appendix B.
4.9 Substantive Assumptions Revisited

Now, we have a statistically adequate model in the form of a St-VAR(3;ν=3) model, which rejected the null model (Ireland’s (2011) structural DSGE model). This puts one in a position to raise the validity of the substantive assumptions of the DSGE model without falling into Duhem’s ambiguity; see Spanos (2010). Since we have already made sure that statistical premises are valid for the data, the only thing that can go wrong is either the theory is wrong or the data we have does not represent the variables of the theoretical model. Following are some of the major substantive assumptions that can easily be unrealistic.

(a) Functional form of the utility function and the welfare function is ad hoc.

(b) Identical households assumption is very unrealistic.

(c) The finished goods producing firms are not allowed to use labor directly.

(d) The price adjustment equation (2.14) (Rotemberg, 1982) is only one of the many in the literature.

(e) The linerization using the first order Taylor’s expansion can be very crude.

(f) The monetary policy rule may require some substantive changes. Taylor’s rule is derived under the assumption of constant conditional variance of $p_t$.

(g) The shock processes are assumed to be either AR(1) with restrictions or unit root. There is neither substantive nor statistical reason behind such assumptions.

So the model can go wrong because of the invalidity of any one of these substantive assumptions. When we derive the more realistic macroeconomic model, all of these assumptions have to be addressed properly.
4.10 New Way of Macroeconomic Modeling

The DSGE models aim to explain the key macroeconomic phenomena like business cycles in terms of individual inter-temporal decision making facing stochastic shocks. The theory-oriented transformations of the data and the restrictions on them are totally misplaced and statistically uninformed as we saw in Chapter 2 and in the likelihood ratio test.

Modeling with the least possible constraints on the vector stochastic process \( \{Z_t, t=1, 2, 3...\} \) can be done only by taking the data in its most original form. Once we transform the data through processes such as differencing, log and ratio, we already impose the restrictions on the statistical model, which has little chance of being tested against the data. In our case the variables can be expressed in terms of its original components as follows

(a) The per-capita real GDP can be further separated into real GDP and population size.

(b) The inflation rate can be further expressed into GDP deflator or even in terms of nominal GDP and real GDP.

(c) The interest rate is the only variable in its original form.

The whole Ireland’s DSGE model can be expressed in terms of the original variables. Furthermore, we can keep variables like consumption \((C_t)\), which is dropped using an identity of the model, in the statistical model. This allows us to test the validity of the identity (restrictions) imposed by the structural model.

In most of the DSGE models, there are some unobserved variables which forces one to use filtering such as the Kalman filter and the particle filter. The use of these filters cannot be justified when the unobserved variables either functions of some observed variables, some stochastic processes like random walk or both. In the Ireland’s model, \( \hat{g}_t, \hat{y}_{ht} \) and \( \hat{c}_t \) are
unobserved variables, which are defined as follows:

\[
\hat{g}_t = \ln(Y_t) - \ln(Q_t) - \ln(\bar{g})
\]

\[
\hat{y}_{ht} = \ln(y_{ht} / y_h) = \ln(Y_t) - \ln(H_t) - \ln(\bar{y}_h)
\]

\[
\hat{c}_t = \ln(c_t / \bar{c}) = \ln(C_t) - \ln(H_t) - \ln(\bar{c})
\]

Here, \(C_t\) and \(Y_t\) are the observable variables, because \(C_t\) is the consumption spending and one of the observable is defined as \(\hat{y}_t = \ln(Y_t) - \ln(Y_{t-1}) - \ln(y)\). It has to be noted that the only unobserved term in above three unobservable variables is \(Z_t\) whose logarithm is a random walk process with a drift. Solving the random walk process backward, it can be shown that it is a process with a trending mean and a trending variance as follows

\[
\ln(H_t) = \ln(H_0) + t \ln(h) + \Sigma_{i=1}^{t} \varepsilon_i
\]

(4.13)

Here, \(H_t \sim N(\ln(H_0) + t \ln(h), t \sigma_h^2)\). Note that the respecified Student’s \(t\) VAR model already has conditional distribution of the variables with trending mean and trending variance built into it. Ideally, the solution of the DSGE model can be expressed in terms of the observables \(Z_t = (Y_t, r_t, P_t, C_t)\) only.

### 4.11 St-VAR Model for Other Data Sets

During the research the St-VAR model was estimated for two more cases

(a) St-VAR\((3;\nu=3)\) for the vector \(Z_t = [Y_t \ C_t \ p_t \ r_t]^\top\) for the US data.

(b) St-VAR\((1;\nu=3)\) for the vector \(Z_t = [Y_t \ p_t \ r_t]^\top\) for the UK data.
The St-VAR turned out to be statistically adequate in both of these cases. This provides a very strong support for the use of St-VAR model for macroeconomic time series data.
Appendix B

Ireland’s 2004 and 2011 Models

Table B.1: DSGE Model: 2011

<table>
<thead>
<tr>
<th>Behavioral equations:</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) [ \ln \left( \frac{Y_t}{Q_t} \right) = \ln \left( \frac{Y_{t-1}}{Q_{t-1}} \right) - \ln \left( \frac{H_t}{H_{t-1}} \right) ]</td>
</tr>
<tr>
<td>(ii) [ \ln \left( \frac{P_t/P_{t-1}}{P_t} \right) = \frac{\alpha}{1+\beta} \ln \left( \frac{P_{t-1}/P_{t-2}}{P_t} \right) + \frac{\beta}{1+\beta} \ln \left( \frac{H_t}{H_{t-1}} \right) + \frac{\psi}{1+\beta} \ln \left( \frac{\Lambda_t}{\Lambda_{t-1}} \right) + \frac{\gamma}{1+\beta} \ln \left( \frac{\lambda_t}{\lambda_{t-1}} \right) - \frac{1}{1+\beta} \tilde{a}_t ]</td>
</tr>
<tr>
<td>(iii) [ \ln \left( \frac{r_t}{r_{t-1}} \right) = \ln \left( \frac{r_{t-1}}{r_{t-2}} \right) + \frac{\rho_p}{\rho_y} \ln \left( \frac{P_t}{P_{t-1}} \right) + \frac{\rho_y}{\rho_y} \ln \left( \frac{Y_t}{Y_{t-1}} \right) + \delta \ln \left( \frac{\theta_t}{\theta_{t-1}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{\Lambda_t}{\Lambda_{t-1}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{\lambda_t}{\lambda_{t-1}} \right) - \frac{1}{\gamma+\beta} \tilde{a}_t ]</td>
</tr>
<tr>
<td>(iv) [ \ln \left( \frac{Y_t}{Y_{t-1}} \right) = \ln \left( \frac{Y_{t-1}}{Y_{t-2}} \right) - \ln \left( \frac{H_t}{H_{t-1}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{\Lambda_t}{\Lambda_{t-1}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{\lambda_t}{\lambda_{t-1}} \right) - \frac{1}{\gamma+\beta} \tilde{a}_t ]</td>
</tr>
<tr>
<td>(v) [ \ln \left( \frac{Q_t}{Q_{t-1}} \right) = \ln \left( \frac{Q_{t-1}}{Q_{t-2}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{Q_{t-1}}{Q_{t-2}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{H_t}{H_{t-1}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{\Lambda_t}{\Lambda_{t-1}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{\lambda_t}{\lambda_{t-1}} \right) - \frac{1}{\gamma+\beta} \tilde{a}_t ]</td>
</tr>
<tr>
<td>(vi) [ \ln \left( \frac{H_t}{H_{t-1}} \right) = \frac{\gamma}{\gamma+\beta} \ln \left( \frac{H_{t-1}}{H_{t-2}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{\Lambda_t}{\Lambda_{t-1}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{\lambda_t}{\lambda_{t-1}} \right) - \frac{1}{\gamma+\beta} \tilde{a}_t ]</td>
</tr>
<tr>
<td>(vii) [ \ln \left( \frac{R_t}{R_{t-1}} \right) = \ln \left( \frac{R_{t-1}}{R_{t-2}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{R_{t-1}}{R_{t-2}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{Y_t}{Y_{t-1}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{Q_t}{Q_{t-1}} \right) - \frac{1}{\gamma+\beta} \tilde{a}_t ]</td>
</tr>
<tr>
<td>(viii) [ \ln \left( \frac{a_t}{a_{t-1}} \right) = \ln \left( \frac{a_{t-1}}{a_{t-2}} \right) + \delta \ln \left( \frac{\theta_t}{\theta_{t-1}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{\Lambda_t}{\Lambda_{t-1}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{\lambda_t}{\lambda_{t-1}} \right) - \frac{1}{\gamma+\beta} \tilde{a}_t ]</td>
</tr>
<tr>
<td>(ix) [ \ln \left( \frac{\theta_t}{\theta_{t-1}} \right) = \ln \left( \frac{\theta_{t-1}}{\theta_{t-2}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{\Lambda_t}{\Lambda_{t-1}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{\lambda_t}{\lambda_{t-1}} \right) - \frac{1}{\gamma+\beta} \tilde{a}_t ]</td>
</tr>
<tr>
<td>(x) [ \ln \left( \frac{H_t}{H_{t-1}} \right) = \ln \left( \frac{H_{t-1}}{H_{t-2}} \right) + \delta \ln \left( \frac{\theta_t}{\theta_{t-1}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{\Lambda_t}{\Lambda_{t-1}} \right) + \frac{\gamma}{\gamma+\beta} \ln \left( \frac{\lambda_t}{\lambda_{t-1}} \right) - \frac{1}{\gamma+\beta} \tilde{a}_t ]</td>
</tr>
</tbody>
</table>

Parameters: \( \varphi := (\alpha, \beta, \gamma, \psi, \rho_p, \rho_a, \rho_y, \rho_y, \sigma^2_\alpha, \sigma^2_\beta, \sigma^2_\gamma, \sigma^2_\psi) \)
Table B.2: DSGE Model: 2004

Behavioral equations:

(i) \( \ln \left( \frac{Y_t}{Q_t} \right) = \alpha_x \ln \left( \frac{Y_{t-1}}{Q_{t-1}} \right) + (1 - \alpha_g) E_t \ln \left( \frac{Y_{t+1}}{Q_{t+1}} \right) - \left\{ \ln \left( \frac{Y_t}{Q_t} \right) - E_t \ln \left( \frac{P_t}{P_{t-1}} \right) \right\} + (1 - \omega)(1 - \rho_a) \ln (a_t) \)

(ii) \( \ln \left( \frac{P_t}{P_{t-1}} \right) = \beta \left( \alpha_p \ln \left( \frac{P_{t-1}}{P_{t-2}} \right) \right) + (1 - \alpha_p) E_t \ln \left( \frac{P_{t+1}}{P_t} \right) + \psi \ln \left( \frac{Y_t}{Q_t} \right) - \left( \frac{1}{\eta} \right) \ln \left( \frac{Y_t}{Q_t} \right) \)

(iii) \( \ln \left( \frac{r_t}{r_{t-1}} \right) - \ln \left( \frac{r_{t-1}}{r_t} \right) = \rho_p \ln \left( \frac{P_t}{P_{t-1}} \right) + \rho_y \ln \left( \frac{Y_t}{Y_{t-1}} \right) + \rho_g \ln \left( \frac{Y_t}{Q_t} \right) + \varepsilon_r \)

(iv) \( \ln \left( \frac{Y_t}{Q_t} \right) = \ln \left( \frac{H_t}{q_h} \right) - \omega \ln (a_t) \)

(v) \( \ln \left( \frac{Y_t}{Y_{t-1}} \right) = \ln \left( \frac{Y_{t+1}}{h_{t+1}} \right) - \ln \left( \frac{Y_{t-1}}{H_{t-1}} \right) + \ln \left( \frac{H_t}{H_{t-1}} \right) \)

Shocks:

(vi) \( \ln a_t = \rho_a \ln a_{t-1} + \varepsilon_a \), (vii) \( \ln \left( \frac{\theta_t}{\theta_{t-1}} \right) = \ln \left( \frac{\theta_{t-1}}{\theta_t} \right) + (\psi \rho_a) \ln \left( \frac{\theta_{t-1}}{\theta_t} \right) + \varepsilon_\theta \),

(viii) \( \ln \left( \frac{H_t}{H_{t-1}} \right) = \varepsilon_h \)

Parameters: \( \varphi := (\alpha_g, \rho_a, \beta, \psi, \alpha_p, \rho_p, \rho_y, \rho_g, \omega, \sigma_a^2, \sigma_\theta^2, \sigma_h^2, \sigma_r^2) \)

The four differences between the structural models of 2004 paper and 2011 paper are as follows:

(a) Functional form of the utility function. For 2004 paper and 2011 paper are respectively,

\[
2004 : \beta^t [a_t \ln (C_t + \ln (M_t/P_t) - (1/\eta)l_t^p)] \\
2011 : \beta^t [a_t \ln (C_t - \gamma C_{t-1}) + \ln (M_t/P_t) - l_t]
\]

Note that one utility function does not nest the other. But if we impose \( \gamma = 0 \) and \( \eta = 1 \) on both of them, we get the same utility function.

(b) The cost of nominal price adjustment:

\[
2004 : \frac{\phi}{2} \left[ \frac{P_t(i)}{\bar{P}P_{t-1}(i)} - 1 \right]^2 Y_t \\
2011 : \frac{\phi}{2} \left[ \frac{P_t(i)}{P_{t-1}^\alpha \bar{P}^{1-\alpha} P_{t-1}(i)} - 1 \right]^2 Y_t
\]
Note that if we impose $\alpha = 0$, the price adjustment equation (B.4) reduces to (B.3).

(c) The Taylor rule:

$2004 : \ln r_t - \ln r_{t-1} = \rho_p \ln(p_t/\bar{p}) + \rho_y \ln(y_t/\bar{y}) + \rho_g \ln(g_t/\bar{g}) + \varepsilon_{rt}$ \hspace{1cm} (B.5)

$2011 : \ln r_t - \ln r_{t-1} = \rho_p \ln(p_t/\bar{p}) + \rho_y \ln(y_t/\bar{y}) + \varepsilon_{rt}$ \hspace{1cm} (B.6)

Imposing $\rho_g = 0$ on (5) gives (6)

(d) There are additional parameters in the 2004 model, $\alpha_g$ and $\alpha_p$, which are absent in the 2011 model. These parameters constitute the coefficients of the additional lagged and expected values of the output gap ($g_t$) and inflation ($p_t$). These extra terms are added after the linearisation of the structural model in the 2004 model in a hope to make the linearized model more data friendly. But in the 2011 model, no extra terms are added in the linearized model.

We can notice that if we impose the 6 restrictions $\eta = 1, \rho_g = 0, \alpha_g = 0, \alpha_p = 0, \alpha = 0, \gamma = 0$ on both the models, we get a single model. But one model cannot nest the other model completely.

All other differences between the two models are the results of above mentioned four differences. The differences between the two models can be seen more clearly once the models are expressed in terms of levels of the variables (see Table B.1 and Table B.2 for detail).
Appendix C

Likelihood Ratio Test: Ireland’s 2004 Model

There are 27 restrictions on the conditional mean to be tested. The results of the likelihood ratio tests are presented in Table B.1.

<table>
<thead>
<tr>
<th>Log-likelihood under</th>
<th>$H_1$</th>
<th>$H_0$</th>
<th>$LR$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5.093</td>
<td>1111.200</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table C.1: Likelihood Ratio Test for Ireland (2004) Model
Appendix D

Identification through Simulation

In this Appendix, the histograms of the simulation done in Chapter 4 are presented. To create the histograms, there are two steps. First, statistically adequate model (St-VAR model in this case) is used to simulate observations. Second, the newly simulated data set is used to estimate the structural parameters by maximizing Ireland (2011) likelihood function. If Ireland (2011) model can be identified, then the estimates from each simulated data should closely fall around some constant value. However, if the model is not identified, the estimates will take a wide range of values. Lower the scattering of estimates, as shown by the histograms, more precisely the corresponding parameter is identified. For example: histogram for \( \hat{\alpha} \) and \( \hat{\gamma} \) in Figure D.1 indicate that the parameters are identified.
Figure D.1: Identification of Ireland (2011) Model ($N = 1200$)
Figure D.2: Identification of Ireland (2011) Model (N = 1200)
\[ \omega^* = 0.0617 \]

\[ \alpha^* x = 0.0836 \]

\[ \alpha^* p = 0.3597 \]

\[ \rho^* y = 0.2536 \]

\[ \rho^* x = 0.0347 \]

Figure D.3: Identification of Ireland (2004) Model \((N = 955)\)
\( \hat{\rho}_a = 0.947 \)
\( \hat{\sigma}_a = 0.0405 \)
\( \hat{\rho}_e = 0.9625 \)
\( \hat{\sigma}_e = 0.0012 \)
\( \hat{\sigma}_z = 0.0109 \)
\( \hat{\sigma}_r = 0.0031 \)

Figure D.4: Identification of Ireland (2004) Model \((N = 955)\)
Appendix E

Autoskedastic Function of St-VAR model

In Chapter 4, the conditional variance of growth rate $g_t$ is presented in equation form. There are five remaining variance-covariance estimated from the same St-VAR($3; \nu = 3$) model, which are presented in this Appendix.
Table E.1: Conditional Variance of $p_t$: $VAR(p_t|Z_{t-1}^0)$

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.018[0.000]</td>
<td>$\tilde{r}<em>{t-1}\tilde{p}</em>{t-3}$</td>
<td>0.021[0.376]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}_{t-1}^2$</td>
<td>0.014[0.000]</td>
<td>$\tilde{r}<em>{t-1}\tilde{r}</em>{t-3}$</td>
<td>0.458[0.000]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}<em>{t-1}\tilde{p}</em>{t-1}$</td>
<td>0.018[0.000]</td>
<td>$\tilde{y}_{t-2}^2$</td>
<td>0.015[0.000]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}<em>{t-1}\tilde{r}</em>{t-1}$</td>
<td>-0.046[0.000]</td>
<td>$\tilde{y}<em>{t-2}\tilde{p}</em>{t-2}$</td>
<td>0.024[0.000]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}<em>{t-1}\tilde{y}</em>{t-2}$</td>
<td>-0.007[0.000]</td>
<td>$\tilde{y}<em>{t-2}\tilde{r}</em>{t-2}$</td>
<td>-0.015[0.260]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}<em>{t-1}\tilde{p}</em>{t-2}$</td>
<td>-0.013[0.000]</td>
<td>$\tilde{y}<em>{t-2}\tilde{y}</em>{t-3}$</td>
<td>-0.006[0.000]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}<em>{t-1}\tilde{r}</em>{t-2}$</td>
<td>0.075[0.000]</td>
<td>$\tilde{y}<em>{t-2}\tilde{p}</em>{t-3}$</td>
<td>-0.010[0.004]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}<em>{t-1}\tilde{y}</em>{t-3}$</td>
<td>-0.001[0.271]</td>
<td>$\tilde{y}<em>{t-2}\tilde{r}</em>{t-3}$</td>
<td>0.037[0.000]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}<em>{t-1}\tilde{p}</em>{t-3}$</td>
<td>0.007[0.042]</td>
<td>$\tilde{p}_{t-2}$</td>
<td>0.125[0.000]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}<em>{t-1}\tilde{r}</em>{t-3}$</td>
<td>-0.028[0.001]</td>
<td>$\tilde{p}<em>{t-2}\tilde{r}</em>{t-2}$</td>
<td>-0.124[0.004]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{p}<em>{t-1}\tilde{r}</em>{t-1}$</td>
<td>0.103[0.000]</td>
<td>$\tilde{p}<em>{t-2}\tilde{y}</em>{t-3}$</td>
<td>-0.004[0.183]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{p}<em>{t-1}\tilde{y}</em>{t-2}$</td>
<td>-0.116[0.000]</td>
<td>$\tilde{p}<em>{t-2}\tilde{p}</em>{t-3}$</td>
<td>-0.106[0.000]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{p}<em>{t-1}\tilde{r}</em>{t-2}$</td>
<td>-0.005[0.082]</td>
<td>$\tilde{p}<em>{t-2}\tilde{r}</em>{t-3}$</td>
<td>0.089[0.000]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{p}<em>{t-1}\tilde{p}</em>{t-2}$</td>
<td>-0.107[0.000]</td>
<td>$\tilde{r}_{t-2}^2$</td>
<td>1.643[0.000]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{p}<em>{t-1}\tilde{r}</em>{t-2}$</td>
<td>0.144[0.000]</td>
<td>$\tilde{r}<em>{t-2}\tilde{y}</em>{t-3}$</td>
<td>-0.048[0.000]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{p}<em>{t-1}\tilde{y}</em>{t-3}$</td>
<td>-0.003[0.457]</td>
<td>$\tilde{r}<em>{t-2}\tilde{p}</em>{t-3}$</td>
<td>-0.009[0.820]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{p}<em>{t-1}\tilde{p}</em>{t-3}$</td>
<td>-0.050[0.000]</td>
<td>$\tilde{r}<em>{t-2}\tilde{r}</em>{t-3}$</td>
<td>-1.665[0.000]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{p}<em>{t-1}\tilde{r}</em>{t-3}$</td>
<td>-0.041[0.089]</td>
<td>$\tilde{r}_{t-3}^2$</td>
<td>0.013[0.000]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{r}_{t-1}^2$</td>
<td>0.653[0.000]</td>
<td>$\tilde{y}<em>{t-3}\tilde{p}</em>{t-3}$</td>
<td>0.017[0.000]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{r}<em>{t-1}\tilde{y}</em>{t-2}$</td>
<td>-0.021[0.008]</td>
<td>$\tilde{y}<em>{t-3}\tilde{r}</em>{t-3}$</td>
<td>0.035[0.000]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{r}<em>{t-1}\tilde{p}</em>{t-2}$</td>
<td>0.024[0.302]</td>
<td>$\tilde{p}_{t-3}^2$</td>
<td>0.105[0.000]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{r}<em>{t-1}\tilde{r}</em>{t-2}$</td>
<td>-1.671[0.000]</td>
<td>$\tilde{p}<em>{t-3}\tilde{r}</em>{t-3}$</td>
<td>-0.034[0.113]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{r}<em>{t-1}\tilde{y}</em>{t-3}$</td>
<td>0.010[0.223]</td>
<td>$\tilde{r}_{t-3}^2$</td>
<td>0.645[0.000]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table E.2: Conditional Variance of $r_t$: $VAR(r_t|\mathbf{Z}^0_{t-1})$

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.003[0.000] &amp; $\tilde{r}<em>{t-1}\tilde{p}</em>{t-3}$ &amp; 0.003[0.375]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}^2_{t-1}$</td>
<td>0.002[0.000] &amp; $\tilde{r}<em>{t-1}\tilde{r}</em>{t-3}$ &amp; 0.072[0.000]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}<em>{t-1}\tilde{p}</em>{t-1}$</td>
<td>0.003[0.000] &amp; $\tilde{y}^2_{t-2}$ &amp; 0.002[0.000]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}<em>{t-1}\tilde{r}</em>{t-1}$</td>
<td>-0.007[0.000] &amp; $\tilde{y}<em>{t-2}\tilde{p}</em>{t-2}$ &amp; 0.004[0.000]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}<em>{t-1}\tilde{y}</em>{t-2}$</td>
<td>-0.001[0.000] &amp; $\tilde{y}<em>{t-2}\tilde{r}</em>{t-2}$ &amp; -0.002[0.266]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}<em>{t-1}\tilde{p}</em>{t-2}$</td>
<td>-0.002[0.000] &amp; $\tilde{y}<em>{t-2}\tilde{y}</em>{t-3}$ &amp; -0.001[0.000]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}<em>{t-1}\tilde{r}</em>{t-2}$</td>
<td>0.012[0.000] &amp; $\tilde{y}<em>{t-2}\tilde{p}</em>{t-3}$ &amp; -0.002[0.004]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}<em>{t-1}\tilde{r}</em>{t-3}$</td>
<td>0.000[0.272] &amp; $\tilde{y}<em>{t-2}\tilde{r}</em>{t-3}$ &amp; 0.006[0.000]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}<em>{t-1}\tilde{p}</em>{t-3}$</td>
<td>0.001[0.041] &amp; $\tilde{p}_{t-2}$ &amp; 0.019[0.000]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{y}<em>{t-1}\tilde{r}</em>{t-3}$</td>
<td>-0.004[0.001] &amp; $\tilde{p}<em>{t-2}\tilde{r}</em>{t-2}$ &amp; -0.019[0.004]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{p}<em>{t-1}\tilde{r}</em>{t-1}$</td>
<td>0.016[0.000] &amp; $\tilde{p}<em>{t-2}\tilde{y}</em>{t-3}$ &amp; -0.001[0.184]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{p}<em>{t-1}\tilde{y}</em>{t-2}$</td>
<td>-0.018[0.000] &amp; $\tilde{p}<em>{t-2}\tilde{p}</em>{t-3}$ &amp; -0.017[0.000]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{p}<em>{t-1}\tilde{r}</em>{t-2}$</td>
<td>-0.001[0.084] &amp; $\tilde{p}<em>{t-2}\tilde{r}</em>{t-3}$ &amp; 0.014[0.000]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{p}<em>{t-1}\tilde{p}</em>{t-2}$</td>
<td>-0.017[0.000] &amp; $\tilde{r}^2_{t-2}$ &amp; 0.257[0.000]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{p}<em>{t-1}\tilde{y}</em>{t-3}$</td>
<td>0.022[0.000] &amp; $\tilde{r}<em>{t-2}\tilde{y}</em>{t-3}$ &amp; -0.007[0.000]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{p}<em>{t-1}\tilde{r}</em>{t-3}$</td>
<td>0.000[0.458] &amp; $\tilde{r}<em>{t-2}\tilde{p}</em>{t-3}$ &amp; -0.001[0.819]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{p}<em>{t-1}\tilde{p}</em>{t-3}$</td>
<td>-0.008[0.000] &amp; $\tilde{r}<em>{t-2}\tilde{r}</em>{t-3}$ &amp; -0.260[0.000]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{p}<em>{t-1}\tilde{r}</em>{t-3}$</td>
<td>-0.006[0.087] &amp; $\tilde{y}^2_{t-3}$ &amp; 0.002[0.000]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{r}^2_{t-1}$</td>
<td>0.102[0.000] &amp; $\tilde{y}<em>{t-3}\tilde{p}</em>{t-3}$ &amp; 0.003[0.000]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{r}<em>{t-1}\tilde{y}</em>{t-2}$</td>
<td>-0.003[0.006] &amp; $\tilde{y}<em>{t-3}\tilde{r}</em>{t-3}$ &amp; 0.006[0.000]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{r}<em>{t-1}\tilde{p}</em>{t-2}$</td>
<td>0.004[0.304] &amp; $\tilde{p}^2_{t-3}$ &amp; 0.016[0.000]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{r}<em>{t-1}\tilde{r}</em>{t-2}$</td>
<td>-0.261[0.000] &amp; $\tilde{p}<em>{t-3}\tilde{r}</em>{t-3}$ &amp; -0.005[0.119]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{r}<em>{t-1}\tilde{y}</em>{t-3}$</td>
<td>0.002[0.221] &amp; $\tilde{r}^2_{t-3}$ &amp; 0.101[0.000]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table E.3: Conditional Covariance of \( \{y_t, p_t\} \): \( \text{VAR}(y_t, p_t|Z_{t-1}^0) \)

|       | \( \hat{y}_t \) & \( \hat{\tilde{y}}_t \) & \( \hat{\rho}_t \) & \( \hat{\rho}_t^2 \) |
|-------|---------------|---------------|---------------|---------------|
| 1     | -0.008[0.000] | 0.007[0.000]  | -0.005[0.389] |               |
| \( \hat{y}_t \) & -0.008[0.009] | -0.007[0.000] | -0.006[0.006] |               |
| \( \hat{\tilde{y}}_t \) & 0.022[0.001]  | 0.003[0.001]  | 0.001[0.275]  |               |
| \( \hat{\rho}_t \) & 0.006[0.018]  | 0.001[0.001]  | 0.002[0.044]  |               |
| \( \hat{\rho}_t^2 \) & -0.035[0.001] | 0.004[0.040]  | 0.001[0.006]  |               |
| \( \hat{y}_t \) & 0.013[0.010]  | 0.054[0.001]  | 0.025[0.000]  | 0.001[0.234]  |
| \( \hat{\tilde{y}}_t \) & 0.048[0.000]  | 0.003[0.133]  | 0.021[0.006]  | -0.011[0.005] |
| \( \hat{\rho}_t \) & 0.050[0.000]  | 0.019[0.011]  | 0.025[0.000]  | 0.045[0.000]  |
| \( \hat{\rho}_t^2 \) & -0.067[0.006] | 0.001[0.474]  | 0.002[0.820]  | 0.039[0.000]  |
| \( \hat{y}_t \) & 0.001[0.133]  | 0.024[0.001]  | 0.039[0.000]  | 0.007[0.000]  |
| \( \hat{\rho}_t \) & 0.050[0.000]  | 0.019[0.011]  | 0.025[0.000]  | 0.045[0.000]  |
| \( \hat{\rho}_t^2 \) & -0.067[0.006] | 0.001[0.474]  | 0.002[0.820]  | 0.039[0.000]  |
| \( \hat{y}_t \) & 0.001[0.133]  | 0.024[0.001]  | 0.039[0.000]  | 0.007[0.000]  |
| \( \hat{\rho}_t \) & 0.050[0.000]  | 0.019[0.011]  | 0.025[0.000]  | 0.045[0.000]  |
| \( \hat{\rho}_t^2 \) & -0.067[0.006] | 0.001[0.474]  | 0.002[0.820]  | 0.039[0.000]  |
| \( \hat{y}_t \) & 0.001[0.133]  | 0.024[0.001]  | 0.039[0.000]  | 0.007[0.000]  |
| \( \hat{\rho}_t \) & 0.050[0.000]  | 0.019[0.011]  | 0.025[0.000]  | 0.045[0.000]  |
| \( \hat{\rho}_t^2 \) & -0.067[0.006] | 0.001[0.474]  | 0.002[0.820]  | 0.039[0.000]  |
Table E.4: Conditional Covariance of \(\{y_t, r_t\}: VAR(y_t, r_t|Z_{t-1}^0)\)

<table>
<thead>
<tr>
<th></th>
<th>(y_{t-1}^2)</th>
<th>(\tilde{y}_{t-1})</th>
<th>(\tilde{y}_{t-1})</th>
<th>(\tilde{y}_{t-1})</th>
<th>(\tilde{y}_{t-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y_{t-1}^2)</td>
<td>0.004 [0.000]</td>
<td>0.003 [0.000]</td>
<td>0.004 [0.001]</td>
<td>0.003 [0.000]</td>
<td>0.003 [0.000]</td>
</tr>
<tr>
<td>(\tilde{y}_{t-1})</td>
<td>-0.010 [0.009]</td>
<td>-0.011 [0.000]</td>
<td>-0.011 [0.004]</td>
<td>-0.003 [0.000]</td>
<td>-0.001 [0.000]</td>
</tr>
<tr>
<td>(\tilde{y}_{t-1})</td>
<td>0.016 [0.008]</td>
<td>0.002 [0.066]</td>
<td>0.002 [0.066]</td>
<td>0.026 [0.000]</td>
<td>0.026 [0.000]</td>
</tr>
<tr>
<td>(\tilde{y}_{t-1})</td>
<td>0.006 [0.025]</td>
<td>0.021 [0.000]</td>
<td>0.021 [0.000]</td>
<td>-0.026 [0.019]</td>
<td>-0.026 [0.019]</td>
</tr>
<tr>
<td>(\tilde{y}_{t-1})</td>
<td>-0.024 [0.001]</td>
<td>-0.001 [0.010]</td>
<td>-0.001 [0.010]</td>
<td>0.018 [0.006]</td>
<td>0.018 [0.006]</td>
</tr>
<tr>
<td>(\tilde{y}_{t-1})</td>
<td>0.030 [0.005]</td>
<td>0.022 [0.000]</td>
<td>0.022 [0.000]</td>
<td>0.341 [0.000]</td>
<td>0.341 [0.000]</td>
</tr>
<tr>
<td>(\tilde{y}_{t-1})</td>
<td>0.000 [0.046]</td>
<td>0.011 [0.000]</td>
<td>0.011 [0.000]</td>
<td>-0.010 [0.008]</td>
<td>-0.010 [0.008]</td>
</tr>
<tr>
<td>(\tilde{y}_{t-1})</td>
<td>0.010 [0.001]</td>
<td>-0.008 [0.011]</td>
<td>-0.008 [0.011]</td>
<td>0.345 [0.000]</td>
<td>0.345 [0.000]</td>
</tr>
<tr>
<td>(\tilde{y}_{t-1})</td>
<td>0.135 [0.000]</td>
<td>0.135 [0.000]</td>
<td>0.135 [0.000]</td>
<td>0.005 [0.000]</td>
<td>0.005 [0.000]</td>
</tr>
<tr>
<td>(\tilde{y}_{t-1})</td>
<td>-0.004 [0.010]</td>
<td>-0.004 [0.010]</td>
<td>-0.004 [0.010]</td>
<td>0.000 [0.000]</td>
<td>0.000 [0.000]</td>
</tr>
<tr>
<td>(\tilde{y}_{t-1})</td>
<td>0.005 [0.324]</td>
<td>0.005 [0.324]</td>
<td>0.005 [0.324]</td>
<td>0.022 [0.000]</td>
<td>0.022 [0.000]</td>
</tr>
<tr>
<td>(\tilde{y}_{t-1})</td>
<td>-0.347 [0.000]</td>
<td>-0.347 [0.000]</td>
<td>-0.347 [0.000]</td>
<td>0.000 [0.000]</td>
<td>0.000 [0.000]</td>
</tr>
<tr>
<td>(\tilde{y}_{t-1})</td>
<td>0.002 [0.259]</td>
<td>0.002 [0.259]</td>
<td>0.002 [0.259]</td>
<td>0.134 [0.000]</td>
<td>0.134 [0.000]</td>
</tr>
</tbody>
</table>
Table E.5: Conditional Covariance of \( \{p_t, r_t\} \): VAR(\(p_t, r_t|Z_{t-1}^0\))

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>( \tilde{p}_{t-1} )</th>
<th>( \tilde{r}_{t-1} )</th>
<th>( \tilde{r}_{t-1} )</th>
<th>( \tilde{r}_{t-1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.001</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.422]</td>
</tr>
<tr>
<td>( \tilde{y}_{t-1}^2 )</td>
<td>0.001</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{y}<em>{t-1} \tilde{p}</em>{t-1} )</td>
<td>0.001</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{y}<em>{t-1} \tilde{r}</em>{t-1} )</td>
<td>0.000</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{y}<em>{t-1} \tilde{y}</em>{t-2} )</td>
<td>0.000</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{y}<em>{t-1} \tilde{p}</em>{t-2} )</td>
<td>-0.001</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{y}<em>{t-1} \tilde{r}</em>{t-2} )</td>
<td>-0.001</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{y}<em>{t-1} \tilde{y}</em>{t-3} )</td>
<td>0.000</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{y}<em>{t-1} \tilde{p}</em>{t-3} )</td>
<td>0.001</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{y}<em>{t-1} \tilde{r}</em>{t-3} )</td>
<td>-0.002</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{p}<em>{t-1} \tilde{p}</em>{t-1} )</td>
<td>0.007</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{p}<em>{t-1} \tilde{r}</em>{t-1} )</td>
<td>-0.008</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{p}<em>{t-1} \tilde{y}</em>{t-2} )</td>
<td>0.000</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{p}<em>{t-1} \tilde{p}</em>{t-2} )</td>
<td>-0.007</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{p}<em>{t-1} \tilde{y}</em>{t-3} )</td>
<td>0.000</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{p}<em>{t-1} \tilde{p}</em>{t-3} )</td>
<td>-0.003</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{p}<em>{t-1} \tilde{r}</em>{t-3} )</td>
<td>-0.003</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{r}<em>{t-1} \tilde{p}</em>{t-1} )</td>
<td>0.044</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{r}<em>{t-1} \tilde{y}</em>{t-2} )</td>
<td>-0.001</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{r}<em>{t-1} \tilde{p}</em>{t-2} )</td>
<td>0.002</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{r}<em>{t-1} \tilde{r}</em>{t-2} )</td>
<td>-0.113</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{r}<em>{t-1} \tilde{y}</em>{t-3} )</td>
<td>0.001</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
<tr>
<td>( \tilde{r}<em>{t-1} \tilde{r}</em>{t-3} )</td>
<td>0.044</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
<td>0.001[0.000]</td>
</tr>
</tbody>
</table>
Chapter 5

Statistical Adequacy of CAPM

5.1 Introduction

The Student’s t VAR model, extensively discussed in Chapters 3-4, has the potential to
provide a statistically adequate model for many applications in econometric modeling, espe-
cially in relation to financial data. In this chapter we revisit the capital asset pricing model
(CAPM) in an attempt to bring out the weaknesses of current empirical research in this
area. It is shown that the implicit statistical model underlying the traditional CAPM, a
static panel data regression model with coefficients that vary with the cross section ordering,
is statistically misspecified. An attempt to respecify this model with a view to ensure sta-
tistical adequacy leads one to the Student’s t VAR and its various reparametrizations. The
respecified model is found to be capable of embedding the original CAPM allowing one to
test the overidentifying restrictions. In the next section, the structural CAPM is described
briefly.

5.2 CAPM: Structural Model

The CAPM, often credited to Sharpe (1964), builds on Markowitz (1952) portfolio theory
and focuses on risk premiums, the difference between the expected return from a portfolio
of risky assets and a risk-free rate of return, using the following substantive assumptions
(a) The market of each asset is in an equilibrium.

(b) All market participants have the same forecasts on expected returns and risks.

(c) All market participants use optimal portfolio diversification; every one holds a tangency portfolio of risky assets as well as a risk free asset.

(d) The market rewards optimal portfolio diversification but punishes non-optimal portfolio selection.

The key concept underlying the CAPM is the capital market line (CML)

$$\mu_R - \mu_f = \frac{\sigma_R}{\sigma_M} (\mu_M - \mu_f) \quad (5.1)$$

which relates the risk premium on an efficient portfolio $R$ to that of a return on a market portfolio $R_M$, where $\mu_R = E(R), \sigma_R = Var(R), \mu_M = E(R_M), \sigma_M = Var(R_M)$ and $\mu_f$ denotes the risk-free return rate. The CAPM relates the risk premium of asset $k$ to its $\beta_k = \frac{\sigma_R}{\sigma_M}$ to define a linear regression-type model

$$(r_k - \mu_f) = \beta_k (r_M - \mu_f) + \varepsilon_k, \quad k = 1, 2, ..., m, \quad (5.2)$$

where the error term satisfies the properties

$$E(\varepsilon_k) = 0, \quad Cov(\varepsilon_k, r_M) = 0, \quad Var(\varepsilon_k) = \sigma_k^2 - \beta_k^2 \sigma_M^2, \quad k = 1, 2, ..., m.$$ 

Hence, in practice $\beta_k$ is the key parameter of substantive interest because it can be used to measure the sensitivity of asset return $k$ to market movements; see Lai and Xing (2008). The structural model imposes the restriction of zero intercept in equation (5.2).
5.3 Statistical Model

In order to see how the CAPM can be embedded into a statistical linear regression model, let \( y_k = r_k, X = r_M \) and \( Z = \mu_f \). The statistical formulation of the panel linear regression model

\[
M_\theta(z) : \ y_{kt} = \beta_0k + \beta_1kX_t + \beta_2kZ_t + u_{kt}, \ k=1,..,m, \ t=1,2,..., \ (5.3)
\]

nests parametrically the CAPM (5.2) via the substantive restrictions

\[
\beta_0 = 0, \ \beta_1 + \beta_2 = 1, \ k = 1,2,...,m \quad (5.4)
\]

Now the complete specification of the panel data model (5.3) is shown in Table 5.1 with following notation.

\[
B_T^i = (\beta_1^T, \beta_2^T), \ \beta_i = (\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}), \ i = 1, 2.
\]

<table>
<thead>
<tr>
<th>Table 5.1: The Multivariate Normal Regression Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical GM: ( y_t = \beta_0 + B_T^iX_t + u_t, \ t \in T )</td>
</tr>
<tr>
<td>[1] Normality: ( (y_t</td>
</tr>
<tr>
<td>[2] Linearity: ( E(y_t</td>
</tr>
<tr>
<td>[3] Homoskedasticity: ( Var(y_t</td>
</tr>
<tr>
<td>[4] Independence: ( {(y_t</td>
</tr>
<tr>
<td>[5] ( t )-invariance: ( \theta := (\beta_0, B_T^i, V) ) are ( t )-invariant for all ( t \in T ).</td>
</tr>
</tbody>
</table>

Using time series data \( y_t = (y_{1t}, y_{2t}, \ldots, y_{mt}) \) and \( X_t = (X_t, Z_t) \) for \( t = 1,2,\ldots, T \), the model in Table (5.1) can be estimated using maximum likelihood. Now, consider the following weekly log returns of \( m = 3 \) stocks representing different sectors, during the period 2/16/1982 to 5/21/2007 (\( T = 1306 \))^1.

^1All data for this chapter has been taken from Yahoo! Finance and FRED.
(a) Citigroup Inc. (*CITI*) - banking.

(b) General Motors (*GM*) - manufacturing.

(c) Pfizer Inc. (*PFE*) - pharmaceuticals.

The stock returns are all based on Close Prices. The market portfolio is represented by the *S&P 500* index and the risk free asset by the 3-month Treasury bill (*3mtb*).

### 5.4 Exploratory Data Analysis (EDA)

The key role played by Exploratory Data Analysis in the Probabilistic Reduction (PR) approach to modeling has already been discussed in Chapter 2. The *t*-plots of the five variables (Figures 5.1-5.5) exhibit second order temporal dependence through changing volatility. This can be a symptom of heteroskedastic and/or heterogeneous autoskedastic functions. Each of the *t*-plot indicates some outlier like observations, which is a typical feature of random numbers generated from Student’s *t* distribution. Since these are time series data, there is high possibility of strong autocorrelation in the data. In Chapter 4, it has been shown that differencing alone is very unlikely to turn the non-stationary data into stationary.
Figure 5.1: Three Month Treasury Bill Log Returns

Figure 5.2: CITI Log Returns
Figure 5.3: PFE Log Returns

Figure 5.4: GM Log Returns
5.5 Unrestricted Static CAPM

The panel data regression model (5.3), when estimated without the restrictions, is reported in Table 5.2 with the p-values inside square brackets.

Table 5.2: Unrestricted Static CAPM Model

<table>
<thead>
<tr>
<th></th>
<th>CITI</th>
<th>PFE</th>
<th>GM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const.</td>
<td>0.0003 [0.4730]</td>
<td>0.0004 [0.2810]</td>
<td>-0.0002 [0.5690]</td>
</tr>
<tr>
<td>3mtb</td>
<td>1.3475 [0.0000]</td>
<td>0.9435 [0.0000]</td>
<td>1.0238 [0.0000]</td>
</tr>
<tr>
<td>SP500</td>
<td>0.0248 [0.1290]</td>
<td>-0.0108 [0.4900]</td>
<td>-0.0101 [0.5560]</td>
</tr>
</tbody>
</table>

If one were to take these estimates at face value it is clear that the restrictions on the constant (intercept) terms

\[ \beta_{0k} = 0, k = 1, 2, 3, \]
would be accepted by data, but the coefficient (beta) restrictions

\[ \beta_{1k} + \beta_{2k} = 1, k = 1, 2, 3, \]

are likely to be rejected. It is important to note that the overwhelming majority of applied papers testing the CAPM often ignore the second set of restrictions; see Lai and Xing (2008). Hence it is usually inferred that the CAPM is supported by the data. However, such inferences are unwarranted, because, as shown below, the estimated model in Table 5.2 is statistically misspecified.

5.6 M-S Testing of Structural CAPM

To test the statistical adequacy of the estimated model in Table 5.2, one needs to apply thorough Misspecification (M-S) testing to assess the validity of assumptions [1]-[5] in Table 5.1. Assumptions [2]-[5] are tested via auxiliary regressions developed in following subsection.

5.6.1 M-S Testing Auxiliary Regressions for CAPM

The M-S tests for assessing the validity of the assumptions in Table 5.1 are based on following auxiliary regressions

\[ \hat{u}_t = \delta_0 + \delta_1 t + \delta_2 t^2 + A_1 Z_{t-1} + A_2 Z_{t-2} + b_1 x_t + b_2 z_t + b_3 x_t^2 + b_4 z_t^2 + \nu_{1t} \]  \hspace{1cm} (5.5)

\[ \hat{u}_t^2 = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + D_1 Z_{t-1}^2 + D_2 Z_{t-2}^2 + d_1 x_t + d_2 z_t + d_3 x_t^2 + d_4 z_t^2 + \nu_{2t} \]  \hspace{1cm} (5.6)
where $\mathbf{Z}_t = (y_t, X_t, Z_t)$, $\mathbf{y}_t = (y_{1t}, y_{2t}, y_{3t})$ and $\widehat{u}_t$ is the standardized residuals of the unrestricted static CAPM model. The standardized residuals are defined in the Chapter 4. The M-S tests done are summarized in Table 5.3 along side the corresponding null hypotheses. The summary of the distributional tests done are presented in Appendix F. Only the Kolmogorov’s test results are presented because other tests (e.g. Skewness-Kurtosis test) also reject the null with zero $p$-values.

### Table 5.3: M-S Tests for Unrestricted CAPM

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>$H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linearity $F(1304,2)$</td>
<td>$b_3 = b_4 = 0$</td>
</tr>
<tr>
<td>$t$-invariance $F(1304,2)$</td>
<td>$\delta_1 = \delta_2 = 0$</td>
</tr>
<tr>
<td>Independence $F(1303,6)$</td>
<td>$A_1 = A_2 = 0$</td>
</tr>
<tr>
<td>Heteroskedasticity $F(1303,4)$</td>
<td>$d_1 = d_2 = d_3 = d_4 = 0$</td>
</tr>
<tr>
<td>2nd order Independence $F(1303,6)$</td>
<td>$D_1 = D_2 = 0$</td>
</tr>
<tr>
<td>2nd order $t$-invariance $F(1303,2)$</td>
<td>$\gamma_1 = \gamma_2 = 0$</td>
</tr>
</tbody>
</table>

5.6.2 M-S Testing Result for CAPM

The M-S test results reported in Table 5.4, indicate that all the probabilistic assumptions underlying the Linear Regression model in (5.2), with the exception of linearity, are invalid. The Normality assumption, when tested, is invalid for all the CAPM equations. Hence, any tests of the substantive restrictions (5.4) will be untrustworthy.

5.7 Respecification of CAPM

The primary aim of respecifying the original statistical model (Table 5.1) is to find a statistical model which is statistically adequate vis-a-vis data. The M-S testing results in previous section indicate massive departures from the Normality, homoskedasticity, Independence and
Table 5.4: M-S Tests on Unrestricted Static CAPM

<table>
<thead>
<tr>
<th></th>
<th>CITI</th>
<th>PFE</th>
<th>GM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linearity</td>
<td>2.960 [0.052]</td>
<td>0.211 [0.810]</td>
<td>0.190 [0.827]</td>
</tr>
<tr>
<td>Homoskedasticity</td>
<td>7.754 [0.000]</td>
<td>2.037 [0.087]</td>
<td>2.2818 [0.024]</td>
</tr>
<tr>
<td>1st Independence</td>
<td>5.983 [0.000]</td>
<td>0.3488 [0.911]</td>
<td>0.567 [0.7569]</td>
</tr>
<tr>
<td>2nd Independence</td>
<td>5.482 [0.000]</td>
<td>13.761 [0.000]</td>
<td>3.970 [0.001]</td>
</tr>
<tr>
<td>1st t-invariance</td>
<td>1.230 [0.293]</td>
<td>1.146 [0.318]</td>
<td>0.235 [0.791]</td>
</tr>
<tr>
<td>2nd t-invariance</td>
<td>6.825 [0.001]</td>
<td>3.149 [0.043]</td>
<td>20.209 [0.000]</td>
</tr>
<tr>
<td>Normality (KS)</td>
<td>36.29 [0.000]</td>
<td>36.29 [0.000]</td>
<td>36.29 [0.000]</td>
</tr>
</tbody>
</table>

$t$-invariance assumptions. The first attempt to respecify led to the heterogeneous Student’s $t$ VAR(1; $\nu$) model (see Table 3.4), but it turned out that the $t$-invariance assumption continued to be rejected by the data. Closer examination of the data revealed three distinct periods, period 1: 2/16/1982-10/9/1989, period 2: 10/16/1989-3/17/2003 and period 3: 3/24/2003-5/21/2007, during which there have been serious structural changes in the parameters. When estimated separately for the three periods the Student’s $t$ VAR(1; $\nu = 3$) turned out to be approximately statistically adequate.

### 5.8 M-S Testing of St-VAR Model

Table 5.5 and Table 5.6 report the M-S testing results for the three periods for the heterogeneous St-VAR(1; $\nu = 3$) model. The auxiliary regressions are same as those used in Chapter 4. The Kolmogorov’s (KS) test is used as a test for distributional misspecification; see Appendix F for detail.

The test results indicate that the heterogeneous St-VAR(1; $\nu = 3$) model is statistically adequate for the given data. The only misspecification detected in Table 5.5 is of second order dependence in 3mtb equation for the period 3. Given that the problem is not detected for period 1 and 2, the problem is not likely to be severe. In Table 5.6, some departures
from Student’s $t$ distribution is still detected for PFE and 3mtb for period 3 and period 2 respectively. But as compared with the departures from the Normality assumption, these departures are minor. Hence the model can be considered statistically adequate.

<table>
<thead>
<tr>
<th>Table 5.5: M-S Tests for St-VAR(1; $\nu = 3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period 1: 2/16/1982-10/9/1989</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Linearity</td>
</tr>
<tr>
<td>Heterosk.</td>
</tr>
<tr>
<td>1st Indepd.</td>
</tr>
<tr>
<td>2nd Indepd.</td>
</tr>
<tr>
<td>1st t-Invar.</td>
</tr>
<tr>
<td>2nd t-Invar.</td>
</tr>
</tbody>
</table>

|                                    | CITI   | PFE    | GM     | SP500   | 3mtb   |
| Linearity       | 0.035  [0.852] | 1.168  [0.280] | 0.091  [0.762] | 3.408  [0.065] | 0.411  [0.522] |
| Heterosk.       | 0.533  [0.466] | 3.249  [0.072] | 2.605  [0.107] | 0.815  [0.367] | 0.773  [0.380] |
| 1st Indepd.     | 0.964  [0.382] | 0.606  [0.546] | 1.144  [0.319] | 0.339  [0.712] | 2.963  [0.052] |
| 2nd Indepd.     | 0.490  [0.613] | 0.741  [0.477] | 0.489  [0.614] | 0.326  [0.722] | 0.193  [0.825] |
| 1st t-Invar.    | 0.011  [0.989] | 0.588  [0.556] | 0.315  [0.730] | 1.387  [0.250] | 1.577  [0.207] |
| 2nd t-Invar.    | 1.598  [0.203] | 1.795  [0.167] | 0.170  [0.844] | 1.658  [0.191] | 0.440  [0.644] |

|                                    | CITI   | PFE    | GM     | SP500   | 3mtb   |
| Linearity       | 0.231  [0.631] | 0.053  [0.818] | 0.707  [0.401] | 0.119  [0.731] | 0.108  [0.743] |
| Heterosk.       | 0.031  [0.861] | 0.321  [0.572] | 0.182  [0.670] | 0.226  [0.635] | 0.054  [0.817] |
| 1st Indepd.     | 0.190  [0.827] | 0.360  [0.698] | 0.137  [0.872] | 0.011  [0.989] | 2.679  [0.071] |
| 2nd Indepd.     | 0.607  [0.546] | 0.276  [0.759] | 1.017  [0.363] | 3.050  [0.049] | 4.037  [0.019] |
| 1st t-Invar.    | 0.593  [0.554] | 0.048  [0.954] | 1.090  [0.338] | 0.662  [0.517] | 1.977  [0.141] |
| 2nd t-Invar.    | 0.058  [0.943] | 0.041  [0.960] | 0.127  [0.881] | 0.034  [0.966] | 0.418  [0.659] |
### Table 5.6: Kolmogorov-Smirnov Test

<table>
<thead>
<tr>
<th></th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CITI</td>
<td>0.068 [0.052]</td>
<td>0.054 [0.032]</td>
<td>0.078 [0.147]</td>
</tr>
<tr>
<td>GM</td>
<td>0.049 [0.301]</td>
<td>0.045 [0.120]</td>
<td>0.076 [0.168]</td>
</tr>
<tr>
<td>PFE</td>
<td>0.043 [0.461]</td>
<td>0.053 [0.040]</td>
<td>0.124 [0.003]</td>
</tr>
<tr>
<td>SP500</td>
<td>0.054 [0.192]</td>
<td>0.056 [0.024]</td>
<td>0.070 [0.238]</td>
</tr>
<tr>
<td>3mtb</td>
<td>0.067 [0.053]</td>
<td>0.071 [0.002]</td>
<td>0.105 [0.016]</td>
</tr>
</tbody>
</table>

### 5.9 Dynamic Multivariate Student’s $t$ CAPM

To embed the static CAPM in the statistically adequate model, the heterogeneous St-VAR model is reparameterized into dynamic multivariate Student’s $t$ CAPM model (see section 3.10 in Chapter 3 for detail specification of multivariate Student’s $t$ dynamic linear model). It should be noted that any reparameterized model of a statistically adequate model is also a statistically adequate model in itself; so the statistical adequacy of the dynamic multivariate Student’s $t$ CAPM model need not be checked again. The estimation results of the dynamic multivariate Student $t$ CAPM models are presented in Tables 5.7, 5.8 and 5.9.

### Table 5.7: Dynamic CAPM for Period 1

<table>
<thead>
<tr>
<th></th>
<th>CITI</th>
<th>PFE</th>
<th>GM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0064 [0.7809]</td>
<td>0.0056 [0.7921]</td>
<td>0.0419 [0.0266]</td>
</tr>
<tr>
<td>$t$</td>
<td>-0.0137 [0.6995]</td>
<td>-0.0306 [0.4125]</td>
<td>-0.0496 [0.0930]</td>
</tr>
<tr>
<td>$t^2$</td>
<td>0.0047 [0.8276]</td>
<td>0.0053 [0.7919]</td>
<td>0.0381 [0.0320]</td>
</tr>
<tr>
<td>SP500$_t$</td>
<td>0.9161 [0.0000]</td>
<td>1.0921 [0.0000]</td>
<td>0.8944 [0.0000]</td>
</tr>
<tr>
<td>3mtb$_t$</td>
<td>0.0033 [0.8904]</td>
<td>-0.0067 [0.7505]</td>
<td>-0.0126 [0.5105]</td>
</tr>
<tr>
<td>CITI$_{t-1}$</td>
<td>-0.0463 [0.3970]</td>
<td>-0.0260 [0.5732]</td>
<td>-0.0393 [0.3408]</td>
</tr>
<tr>
<td>PFE$_{t-1}$</td>
<td>0.0352 [0.5697]</td>
<td>-0.1690 [0.0009]</td>
<td>-0.0286 [0.5487]</td>
</tr>
<tr>
<td>GM$_{t-1}$</td>
<td>-0.0091 [0.8921]</td>
<td>-0.1349 [0.0124]</td>
<td>-0.0781 [0.1459]</td>
</tr>
<tr>
<td>SP500$_{t-1}$</td>
<td>0.1100 [0.4200]</td>
<td>0.3318 [0.0043]</td>
<td>0.1230 [0.2682]</td>
</tr>
<tr>
<td>3mtb$_{t-1}$</td>
<td>0.0188 [0.5625]</td>
<td>-0.0432 [0.1280]</td>
<td>0.0081 [0.7211]</td>
</tr>
</tbody>
</table>
Table 5.8: Dynamic CAPM for Period 2

<table>
<thead>
<tr>
<th></th>
<th>CITI</th>
<th>PFE</th>
<th>GM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0038 [0.5474]</td>
<td>0.0036 [0.5724]</td>
<td>-0.0050 [0.4510]</td>
</tr>
<tr>
<td>t</td>
<td>-0.0398 [0.0025]</td>
<td>-0.0232 [0.0220]</td>
<td>-0.0203 [0.0894]</td>
</tr>
<tr>
<td>t²</td>
<td>0.0029 [0.5846]</td>
<td>0.0010 [0.8482]</td>
<td>-0.0028 [0.6221]</td>
</tr>
<tr>
<td>SP500t</td>
<td>1.6125 [0.0000]</td>
<td>0.9477 [0.0000]</td>
<td>1.0081 [0.0000]</td>
</tr>
<tr>
<td>3mtbₜ</td>
<td>-0.0082 [0.7140]</td>
<td>0.0197 [0.4558]</td>
<td>0.0027 [0.9174]</td>
</tr>
<tr>
<td>CITIₜ₋₁</td>
<td>-0.1329 [0.0005]</td>
<td>0.0598 [0.0764]</td>
<td>-0.0060 [0.8583]</td>
</tr>
<tr>
<td>PFEₜ₋₁</td>
<td>0.0247 [0.4507]</td>
<td>-0.0555 [0.1361]</td>
<td>-0.0212 [0.6110]</td>
</tr>
<tr>
<td>GMₜ₋₁</td>
<td>-0.0171 [0.6524]</td>
<td>-0.0568 [0.1112]</td>
<td>-0.0469 [0.2090]</td>
</tr>
<tr>
<td>SP500ₜ₋₁</td>
<td>0.1574 [0.1119]</td>
<td>-0.0683 [0.4839]</td>
<td>0.2497 [0.0257]</td>
</tr>
<tr>
<td>3mtbₜ₋₁</td>
<td>0.0078 [0.8082]</td>
<td>-0.0346 [0.3200]</td>
<td>0.0083 [0.8211]</td>
</tr>
</tbody>
</table>

Table 5.9: Dynamic CAPM for Period 3

<table>
<thead>
<tr>
<th></th>
<th>CITI</th>
<th>PFE</th>
<th>GM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0043 [0.9408]</td>
<td>0.0036 [0.9666]</td>
<td>0.0421 [0.7688]</td>
</tr>
<tr>
<td>t</td>
<td>-0.0068 [0.9458]</td>
<td>-0.0191 [0.8594]</td>
<td>-0.0477 [0.7005]</td>
</tr>
<tr>
<td>t²</td>
<td>0.0046 [0.9363]</td>
<td>0.0053 [0.9495]</td>
<td>0.0412 [0.7695]</td>
</tr>
<tr>
<td>SP500ₜ</td>
<td>0.9858 [0.0000]</td>
<td>0.9766 [0.0000]</td>
<td>1.4781 [0.0000]</td>
</tr>
<tr>
<td>3mtbₜ</td>
<td>0.0120 [0.5803]</td>
<td>-0.0488 [0.1520]</td>
<td>0.0141 [0.8001]</td>
</tr>
<tr>
<td>CITIₜ₋₁</td>
<td>-0.0999 [0.1717]</td>
<td>-0.1012 [0.4285]</td>
<td>-0.0538 [0.7579]</td>
</tr>
<tr>
<td>PFEₜ₋₁</td>
<td>0.0122 [0.7998]</td>
<td>-0.0424 [0.5803]</td>
<td>0.0214 [0.8676]</td>
</tr>
<tr>
<td>GMₜ₋₁</td>
<td>0.0233 [0.4256]</td>
<td>-0.0075 [0.8814]</td>
<td>-0.0330 [0.6658]</td>
</tr>
<tr>
<td>SP500ₜ₋₁</td>
<td>0.2141 [0.1007]</td>
<td>0.2301 [0.2787]</td>
<td>-0.1016 [0.7687]</td>
</tr>
<tr>
<td>3mtbₜ₋₁</td>
<td>0.0305 [0.2808]</td>
<td>-0.0093 [0.8158]</td>
<td>-0.0113 [0.8714]</td>
</tr>
</tbody>
</table>
5.10 Fitted Values

The difference between the fitted value $\hat{Z}_t$ and the adjusted fitted value $\hat{\hat{Z}}_t$, defined in Chapter 4, can be easily seen in Figures 5.6-5.8. The fitted values are plotted only for period 1 as example. $\hat{Z}_t$ captures the dynamic, heteroskedastic and heterogeneous nature of the autoskedastic functions while as $\hat{\hat{Z}}_t$ captures only the autoregressive function.

Figure 5.6: Fitted Values for CITI Log Returns
Figure 5.7: Fitted Values for PFE Log Returns

Figure 5.8: Fitted Values for GM Log Returns
5.11 Autoskedastic Function

The estimated autoskedastic function for CITI log returns is plotted in Figure 5.9 as an example. The plot clearly indicates the changing conditional variance across time. After 3/24/2003, the sudden decline in volatility is estimated by the model. This sudden change in volatility can be one reason why a single model for the whole time period violated the $t$-invariance problem at first respecification attempt.

Figure 5.9: Fitted Conditional Variance for CITI Log Returns
5.12 Likelihood Ratio Tests

To test the statistical validity of the static CAPM model we need to test the following overidentifying restrictions (also given in equation 5.4)

\[ \beta_{0k} = 0, \ \beta_{1k} + \beta_{2k} = 1, \ k = 1, 2, \ldots, m \]

on the estimated dynamic Student’s $t$ CAPM in Tables 5.7-5.9. The likelihood ratio test results are presented in Table 5.10. The likelihood ratio tests clearly reject the static CAPM model for the data chosen making the use of CAPM model for decision making purpose highly unreliable and misleading. This is also an evidence against efficient market hypothesis.

<table>
<thead>
<tr>
<th>Table 5.10: Likelihood Ratio Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis</td>
</tr>
<tr>
<td>$H_0$</td>
</tr>
<tr>
<td>$H_1$</td>
</tr>
<tr>
<td>Sample Size</td>
</tr>
<tr>
<td>LR</td>
</tr>
<tr>
<td>Bartlett LR</td>
</tr>
</tbody>
</table>
Chapter 6

Summary and Conclusion

6.1 Introduction

The last sixty years of modeling in macroeconomics and finance has focused primarily on the substantive adequacy of the model, irrespective of its empirical adequacy. In this research, two frequently used structural models (DSGE and CAPM) are evaluated for their reliability using probabilistic reduction (PR) approach. In this approach, the statistical model implicit in the structural model is unveiled in terms of a complete and internally consistent list of probabilistic assumptions pertaining to the observables involved. If the assumptions of the statistical model are not validated using thorough misspecification (M-S) testing vis-a-vis the data chosen, the statistical reliability of the estimated model is at best unknown, calling into question any inference based on it, including forecasting and impulse response analysis. The DSGE model based on Ireland (2011) is used as a typical example of macroeconomic model while the CAPM model is taken as a typical example of a widely used financial model. To estimate the DSGE model, three major quarterly macroeconomic data, the growth rate, inflation rate and interest rate are used. The weekly asset price of PFE, GM and CITI, SP500 and three months treasury bill rate are used to estimate the CAPM model. Both the structural models (DSGE and CAPM) are found to be statistically misspecified. The respecification, guided by informal exploratory data analysis (EDA) and formal M-S testing led towards the Student’s t VAR (St-VAR) model. The M-S tests on the St-VAR
model suggests that it is statistically adequate for the both macroeconomic and financial data sets used to estimate the DSGE and the CAPM models respectively. Moreover, the St-VAR model can easily embed both the structural models allowing one to tests the overidentifying restrictions using a likelihood ratio test. To achieve statistical adequacy ones needs to allow the autoregressive and autoskedastic functions to be heterogeneous.

6.2 Contributions

There are basically following contributions of this dissertation.

(a) It is shown that the PR approach to econometric modeling, that separates, ab initio, the statistical from the substantive premises, provides a flexible framework for evaluating the reliability of structural models such as DSGE and CAPM. In this approach, all the information regarding probabilistic structures of the data is taken into consideration while specifying the statistical model. The use of graphical techniques is very important in unveiling the statistical information like trends, dependence and leptokurticity in the data.

(b) The St-VAR model is proposed as an alternative to statistical model underlying the DSGE model, which is shown to be a restricted Normal VAR model. In comparison to models such as DSGE and CAPM, which are based on Normality assumption, the St-VAR offers broader framework for modeling richer probabilistic structures present in both macroeconomic and financial data because it allows for fat tails as well as second order dependence. That is, the St-VAR model can account for the leptokurticity, the heteroskedasticity and the second order dependence present in the data. Unlike in GARCH type models, where the autoregressive and the autoskedastic functions are treated as
separate specifications, in the St-VAR model both of them are derived from the same conditional distribution without any ad-hoc parameter restrictions. The autoregressive function of the St-VAR model is linear and dynamic, which makes it very easy to embed most of the macroeconomic and financial models into it. This allows a researcher to apply a likelihood-ratio test for the validity of the overidentifying restrictions. The restrictions imposed by the DSGE model and the CAPM are found to be empirically unacceptable. In this research, it has also been shown that most of the coefficients of the autokskedastic functions are statistically very significant. This shows that the assumption of homoskedasticity assumption in the VAR model is very unrealistic for this type of data.

(c) This research also demonstrates how various M-S testing auxiliary regressions can be developed and used to secure the thoroughness of the tests.

(d) It is also shown that the impulse response function (IRF) produced by the Normal VAR model and the St-VAR model can be significantly different despite the fact that their autoregressive functions coincide. In Chapter 4, it was shown that the forecasting performance of the St-VAR model is far better than that of the structural VAR implied by the DSGE model. A key reason for this is that the statistical adequacy of the St-VAR model has been secured through thorough M-S testing. Second, forecasting based on the St-VAR model has to be done by taking distributional assumption and autokskedastic function into account. But under Normal VAR model, forecasting is done using only based on the autoregressive function.

(e) In this research, it has been shown that when there is a lack of explicit relationship between the structural parameters and the statistical parameters, the identification can be assessed through simulation. Once a statistically adequate model is secured, it can be used to simulate reliable replicas of the original data, which is then be used to estimate
the structural parameters. Repeating the process gives an empirical distribution for the estimators. A parameter is just identified if the empirical distribution of its estimator has to have narrow support. It has been shown in the Appendix D that most of the parameters in Ireland’s (2011) model are just identified while as some of the parameters in Ireland’s (2004) model are not identified.

(f) In this research, a software written in R code has been developed to estimate the parameters of the heterogeneous St-VAR model; see Appendix G. The software can be used to run the St-VAR for at least two variables, at least one lag and one degrees of freedom. The software generates results such as fitted values, residuals, conditional variance, M-S testing results and log-likelihood. This software is an important contribution of this research. But the code can be further improved to gain speed and efficiency.

6.3 Future Research

This research is a small but important step in macroeconomic modeling stemming from the probabilistic reduction approach. The possibility of multivariate macroeconomic models based on non-Normal distributions has been initiated. In future, this research can be extended in several directions.

(a) New statistical models can be estimated by expressing the variables of the DSGE models in their original form. For example, instead of modeling growth rate of real GDP per capita, one can model real GDP and population as separate variables. This approach reduces the unwarranted restrictions imposed by the equilibrium steady state transformation of data. Modeling GDP deflator instead of inflation proved to be difficult because of computational issues. Once the computational problems are cleared, the
St-VAR model can be estimated for all the variables in levels, without filtering the data.

(b) The St-VAR model can be estimated using the macroeconomic data without seasonal adjustment. This can reveal the potential influence of seasonal adjustment of raw data and the filters used on inference.

(c) The large number of parameters in the St-VAR model and the small sample size in macroeconomics restrict the number of variables to be considered for estimating the model. So estimating a restricted St-VAR, which allows more variables to be included, will be an interesting challenge to conquer. One important fact about the St-VAR model is that the weak exogeneity assumption does not hold and thus, the likelihood needs to treated differently.

(d) The VAR models based on other distributions can be further explored. The first step toward this can be taken by deriving a general VAR model for elliptically symmetric family of distributions. The distributions with non-linear autoregressive function and other forms of autoskedastic function can be interesting for modeling non-linearity and other types of heteroskedasticity; see Bhattacharyya (1943) for one such example. Within the family of Student’s $t$ distribution, the distributions having different degrees of freedom can be explored so that Normality and Student’s $t$ distribution can be incorporated in a single modeling framework. Moving into this direction can result into development of many useful models capable of being statistically adequate for data sets with variables of different probabilistic structures.

(e) Research for development of M-S testing auxiliary regressions for other non-Normal models can be one area of interest for future research.

(f) In this research, the asymptotic distribution of the estimators is used for inference purpose. But the approximation of probability can be very crude for a finite sample. Future
research on simulation technique to derive the finite sample distribution of the estimators of the St-VAR model can contribute further in the direction of making this model more practical.
References


Appendix F

M-S Tests for Distributional Assumptions

The distributional assumption is assessed via the standardized residuals of the model as defined in Chapters 3 and 4.

F.1 Skewness Kurtosis (SK): Normality

The Skewness-Kurtosis test for Normality is based on the work of Jarque and Bera (1987). The test is:

\[ SK(2) = \left( \frac{T}{6} \right) \hat{\alpha}_3^2 + \left( \frac{T}{24} \right) (\hat{\alpha}_4 - 3)^2, \]

where \( \hat{\alpha}_3 \) and \( \hat{\alpha}_4 \) are the estimated skewness and kurtosis of the standardized residuals, is asymptotically distributed as \( \chi^2(2) \) under the null. This test is biased in favour of rejection of the null when there are extreme observations. To deal with it, minimum and maximum of the residuals are removed before doing the test.

F.2 Skewness Kurtosis (SK): Student’s t

This Skewness-Kurtosis test based on the work of Praag and Wesselman (1989). The test is
\[ SK(2) = (T/6[1 + \kappa])\hat{\alpha}_3^2 + (T/24[1 + \kappa])(\hat{\alpha}_4 - 3 - \kappa)^2 \]

where \( \hat{\alpha}_3 \) and \( \hat{\alpha}_4 \) are the estimated skewness and kurtosis of the standardized residuals and 
\( \kappa = \frac{6}{\nu + \ell k - 1} \). \( SK(2) \) is asymptotically distributed as \( \chi^2(2) \) under the null. This test is biased in favour of rejection of the null when there are extreme observations. To deal with it, minimum and maximum of the residuals are removed before doing the test.

F.3 Kolmogorov-Smirnov (KS)

The Kolmogorov-Smirnov (Kolmogorov, 1950) statistic for a given cumulative distribution function \( F(x) \) is

\[ KS_T = \max_x |F_T(x) - F(x)| \]

where \( F_T(x) = \frac{1}{T} \sum_{i=1}^{T} I_{X_i \leq x} \) and \( I_{X_i \leq x} \) is the indicator function equal to 1 if \( X_i \leq x \) and equal to 0 otherwise. By the Glivenko-Cantelli theorem (Cantelli, 1933; Glivenko, 1933), if the sample comes from distribution \( F(x) \), then \( KS_T \) converges to 0 almost surely. Kolmogorov strengthened this result, by effectively providing the rate of this convergence.

F.4 Anderson-Darling (AD)

The Anderson-Darling (Anderson and Darling, 1954) statistic for a given cumulative distributions \( F(x) \) is

\[ AD_T = -T - \sum_{i=1}^{T} \frac{2i-1}{T} [\ln(F(X_i)) + \ln(1 - F(X_{T+1-i}))] \]

where the data is arranged as \( \{x_1 < x_2 < \ldots < x_T\} \). This \( AD \) test statistic can be compared against the critical values of the theoretical distribution.
Appendix G

R Code for St-VAR Estimation and M-S Testing

Run the command

St-VAR(Data,lag,degrees of freedom,maximum iteration,optimization method,
hessian, initial values, trend type)

with appropriate arguments. The St-VAR function is defined as follows.

```
library(MCMCpack) ; library(numDeriv) ; library(fMultivar)
library(moments) ; library(car) ; library(matlab)
source("C:/...../BlockTop.txt") ##Change directory accordingly
source("C:/...../trend.txt") ##Change directory accordingly
source("C:/...../Jacob.txt") ##Change directory accordingly
source("C:/...../MStVAR.txt") ##Change directory accordingly
source("C:/...../Par.txt") ##Change directory accordingly
StVAR<-function(Data,lag,v,maxiter,meth,hes,init,tre)
{
  if (lag-round(lag)!=0) return("lag must be an integer")
  if (lag<0) return("lag number must be positive")
  Z<-t(embed(Data,lag+1)) ; T<-ncol(Z)+lag ; l<-ncol(Data) ## No. of variables in VAR
  if (ncol(cbind(Data))<2) return("no. of variables must be greater than 1")
  if (v<0) return("degrees of freedom must be greater or equal to 0")
  if (maxiter<10) return("Iteration must be at least 10")
  if (2*T<3*((lag+3)*(l+1)*l/2)+3*l) return(list(((lag+3)*(l+1)*l/2)+3*l,"Too many parameters for given sample size. Reduce the number of lags."))
  L<-function(a)
```

183
```r
S <- BlockTop(a[((l*(l+1)/2)+(1-1)+1):((lag+3)*(l+1)*l/2)],l,lag)$S ## Var-Cov Matrix
F <- solve(S)/v ## Var-Cov and its Inverse
M0 <- rep(c(a[((lag+3)*(l+1)*l/2)+(1-1)+1):((lag+3)*(l+1)*l/2)+(lag+1)*l),l) ## Constant term
M1 <- c(a[((lag+3)*(l+1)*l/2)+(lag+1)*l+1:(((lag+3)*(l+1)*l/2)+(lag+1)*l+2*l])) ## Linear term
M2 <- c(a[((lag+3)*(l+1)*l/2)+(lag+1)*l+2*l+1:(((lag+3)*(l+1)*l/2)+(lag+1)*l+3*l))) ## Quadratic term

### Linear and Quadratic trends
trenz <- trend(M1, M2, T, lag, l)
m1 <- trenz$m1 ; m2 <- trenz$m2
if(tre == "Q") M <- (M0 + c(m1[, ,]) + c(m2[, ,]))
if(tre == "L") M <- (M0 + c(m1[, ,]))
if(tre == "C") M <- (M0)
D <- 1 + diag(t(Z - M) %*% F %*% (Z - M)) ## Quadratic form

## Likelihood function
LLn <- (T - lag) * const - 0.5 * (T - lag) * log(det(S)) - 0.5 * (v + (lag + 1) * l) * sum(log(D))
neg.LLn <- -LLn

if(init[1] == "na" & tre == "Q") int <- c(runif(((lag + 3) * (l + 1) * l/2) + 3 * l, 0, 0.1)) ## Initialization
if(init[1] == "na" & tre == "L") int <- c(runif(((lag + 3) * (l + 1) * l/2) + 2 * l, 0, 0.1)) ## Initialization
if(init[1] == "na" & tre == "C") int <- c(runif(((lag + 3) * (l + 1) * l/2) + 1), 0, 0.1)) ## Initialization
if(init[1] != "na") int <- init

const <- log(gamma((v + (lag + 1) * l)/2)) - log(gamma(v/2)) - 0.5 * l * (lag + 1) * log(pi * v)

op <- optim(int, L, hessian = hes, control = list(trace = 1, maxit = maxiter, reltol = 1e-12), method = meth)
a <- op$par ; Like <- op$value

### Parameters
PP <- Par(a, l, v, T, tre)
M0 <- PP$M0 ; M1 <- PP$M1 ; M2 <- PP$M2
Delta0 <- PP$Delta0 ; Delta1 <- PP$Delta1 ; Delta2 <- PP$Delta2
B1 <- PP$B1 ; s2 <- PP$s2 ; Q <- PP$Q
m1 <- PP$m1 ; m2 <- PP$m2

### Fitted values/Residuals/Con. Covariance
q <- v/(v + 1 * lag - 2)
Ct <- Ctt <- vector(length = (T - lag))
U <- muy < muyy = muyx %*% matrix(nrow = 1, ncol = (T - lag))
for(i in 1:(T - lag))
```

\(Z_{11}<-\text{matrix}(\text{nrow}=1, \text{ncol}=\text{lag})\)
for\(s\) in \(1:\text{lag}\)
{
\(Z_{11}[s]<-c(M_0[1:l]+m_1[,s+1,i]+m_2[,s+1,i])\)
}
\(Z_1<-c(Z_{11})\)
if\(\text{tre}=="Q"\) \(M<-M_0+c(m_1[,,i])+c(m_2[,,i])\)
if\(\text{tre}=="L"\) \(M<-M_0+c(m_1[,,i])\)
if\(\text{tre}=="C"\) \(M<-M_0\)
\(C_t[i] \leftarrow 1 + t(Z[[1+1]:((\text{lag}+1)*1),i]-M[[1+1]:((\text{lag}+1)*1),i])%*%Q%*%(Z[[1+1]:((\text{lag}+1)*1),i]-M[[1+1]:((\text{lag}+1)*1),i]))\)
if\(\text{tre}=="L"\) \(C_t[i] \leftarrow 1 + t(Z[[1+1]:((\text{lag}+1)*1),i])%*%Q%*%(Z[[1+1]:((\text{lag}+1)*1),i])#-M_0[[1+1]:((\text{lag}+1)*1)]\)
if\(\text{tre}=="C"\) \(C_t[i] \leftarrow C_t[i]\)
if\(\text{tre}=="Q"\) \(B_0<-M_0[1:l]+m_1[,1,i]+m_2[,1,i]-t(B_1)^%*%Z_1\)
if\(\text{tre}=="L"\) \(B_0<-M_0[1:l]+m_1[,1,i]-t(B_1)^%*%Z_1\)
if\(\text{tre}=="C"\) \(B_0<-M_0[1:l]-t(B_1)^%*%Z_1\)
\(\text{muy}[i]<-B_0+t(B_1)^%*%Z_1[i](1:(\text{lag}+1)),i]\)
if\(\text{tre}=="L"\) \(\text{muyy}[i]<-t(B_1)^%*%Z_1[i](1:(\text{lag}+1)),i]\)
if\(\text{tre}=="C"\) \(\text{muyy}[i]<-\text{muy}[i]\)
\(U[,i]<-t(Z[1:1,i])-\text{muy}[i]\)
}
}
if\(\text{hes}=="\text{TRUE}\)\) \(\text{VARth}<\text{solve}(op^hessian[]\)
if\(\text{hes}=="\text{TRUE} & tre=="Q"\)\)
{
\(Jc<-\text{jacobian}(a,\text{lag},1,v)\) \#jacobian\(J,a\)
\(SE<-\text{sqrt(diag(Jc^%*%VARth^%*%t(Jc))))\)
\(p\_value<-2*(1-\text{pt}(\text{abs}(c(Delta_0,Delta_1,Delta_2,B_1,vech(s2))))/\text{SE},(\text{T}-\text{lag})))\)
\(\text{COEF}<\text{round}(\text{cbind}(c(Delta_0,Delta_1,Delta_2,B_1,vech(s2)),c(\text{SE}[1:\text{length}(\text{SE})]),c(p\_value[1:\text{length}(\text{SE})])),8)\)
}
if\(\text{hes}=="\text{TRUE} & tre=="Q"\) \(\text{COEF}<\text{round}(\text{cbind}(c(Delta_0,Delta_1,Delta_2,B_1,vech(s2)),8)\)
if\(\text{hes}=="\text{TRUE} & tre=="L"\)\)
{
\(Jc<-\text{jacobian}(a,\text{lag},1,v)\) \#jacobian\(J,a\)
\(SE<-\text{sqrt(diag(Jc^%*%VARth^%*%t(Jc))))\)
p_value<-2*(1-pt(abs(c(Delta0,Delta1,B1,vech(s2)))/SE,(T-lag)))
COEF<-round(cbind(c(Delta0,Delta1,B1,vech(s2)),c(SE[1:length(SE)]),c(p_value[1:length(SE)])),8)
}
if(hes!='TRUE' & tre=='L') COEF<-round(cbind(c(Delta0,Delta1,B1,vech(s2))),8)
if(hes=='TRUE' & tre=='C')
{
Jc<-Jacob(a,lag,l,v)$J #jacobian(J,a)
SE<-sqrt(diag(Jc%*%VARth%*%t(Jc)))
p_value<-2*(1-pt(abs(c(Delta0,B1,vech(s2)))/SE,(T-lag)))
COEF<-round(cbind(c(Delta0,B1,vech(s2)),c(SE[1:length(SE)]),c(p_value[1:length(SE)])),8)
}
if(hes!='TRUE' & tre=='C') COEF<-round(cbind(c(Delta0,B1,vech(s2))),8)

################################
##MS-tests####################
################################
MS<-MS(U,lag,tre,s2,Ct,Ctt,muy,muyy,T,q,v)
Dist<-MS$Dist
MS<-MS$MS
if(hes=='TRUE' & tre=='Q') result<-list(beta=cbind(Delta0,Delta1,Delta2,t(B1)),coef=COEF,like=-
Like,sigma=s2,cvar=Ct,res=t(U),fitted=t(muy),ms=MS,dist=Dist,init=a,hes=op$hes,S=S)
if(hes!='TRUE' & tre=='Q') result<-list(beta=cbind(Delta0,Delta1,Delta2,t(B1)),coef=COEF,like=-
Like,sigma=s2,cvar=Ct,res=t(U),fitted=t(muy),ms=MS,dist=Dist,init=a,S=S)
if(hes=='TRUE' & tre=='L') result<-list(beta=cbind(Delta0,Delta1,t(B1)),coef=COEF,like=-
Like,sigma=s2,cvar=Ct,res=t(U),fitted=t(muy),ms=MS,dist=Dist,init=a,hes=op$hes,S=S)
if(hes!='TRUE' & tre=='L') result<-list(beta=cbind(Delta0,Delta1,t(B1)),coef=COEF,like=-
Like,sigma=s2,cvar=Ct,res=t(U),fitted=t(muy),ms=MS,dist=Dist,init=a,S=S)
if(hes=='TRUE' & tre=='C') result<-list(beta=cbind(Delta0,t(B1)),coef=COEF,like=-
Like,sigma=s2,cvar=Ct,res=t(U),fitted=t(muy),ms=MS,dist=Dist,init=a,hes=op$hes,S=S)
if(hes!='TRUE' & tre=='C') result<-list(beta=cbind(Delta0,t(B1)),coef=COEF,like=-
Like,sigma=s2,cvar=Ct,res=t(U),fitted=t(muy),ms=MS,dist=Dist,init=a,S=S)
return(result)
)

###Creating a Positive Definite Block Toeplitz Matrix
BlockTop<-function(a,l,lag)
{
C<-array(dim=c(1,l,lag+3))
for(x in 1:(lag+3))
{
C[,]<-c(xpnd(a[((l*(l+1)/2)*(x-1)+1):(x*(l+1)*l/2)]))}
A <- array(matrix(c(f,f[1:(l*l)]), nrow=l, ncol=(2*(lag+4)-1), dim=c(l,1,2*(lag+4)-1))
MM <- matrix(nrow=(lag+4)*(l), ncol=(lag+4)*(l))
for(i in 1:(lag+4))
{
  for(j in 1:(lag+4))
  {
    MM[(1*i-(l-1)):i*l,(1*j-(l-1)):j*l) <- A[,,j+(i-1)]
  }
}
S <- MM[1:((1+lag)*l),]%*%t(MM[1:((1+lag)*l),])
return(list(S=S))

## Trend coefficients

trend <- function(M1,M2,T,lag,l)
{
  m2 <- m1 <- array(dim=c(l,lag+1,T-lag))
t <- t(seq((1-lag)/1000,(T-lag)/1000,1/1000))
M11 <- M1%*%t ; M22 <- M2%*%((2*t^2-1)
for(i in (T-lag):(1))
{
  m1[,,(i)] <- M11[,(i+lag):(i)]
m2[,,(i)] <- M22[,(i+lag):(i)]
}

return(list(m1=m1,m2=m2))

## Creating the parameters of the conditional distribution
Par <- function(a,l,lag,v,T,tre)
{
  S <- BlockTop(a[((1*(1+l+1)/2)*((1+l)/2))^((lag+3)*1*(lag+1)*2)],1,lag)$S
  S11 <- S[1:1,1:1] ; S12 <- S[1:1,(1+l):((1+l)+1)] ; S22 <- S[(1+l):((1+l)+1),(1+l):((1+l)+1)]
  SS <- solve(S22) ; Q <- SS/v ; B1 <- SS%*%t(S12) ; s2 <- S11-S12%*%QSS%*%t(S12)
  M0 <- rep(c(a[((lag+3)*1*(lag+1)/2)+1]:((lag+3)*1*(lag+1)/2)+1)),(lag+1))
  M1 <- c(a[((lag+3)*1*(lag+1)/2)+1]:((lag+3)*1*(lag+1)/2)+1))
  M2 <- c(a[((lag+3)*1*(lag+1)/2)+2+1]:((lag+3)*1*(lag+1)/2)+3+1))
  if(tre == "C") M1 <- M2 <- rep(0,1) # Quadratic
  if(tre == "L") M2 <- rep(0,1)

  # Linear and Quadratic trend
trenz<-trend(M1,M2,T,lag,1)
m1<-trenz$m1; m2<-trenz$m2
if(tre="Q") M<-(M0+c(m1[,,])+c(m2[,,]))
if(tre="L") M<-(M0+c(m1[,,]))
if(tre="C") M<-M0
q<-v/(v+1*lag-2)
m12<-m22<-matrix(nrow=1,ncol=lag)
for(s in 1:lag)
{
 m22[,s]<-2*s*M2/1000
 m1[,s]<-s*M1/1000+((s/1000)^2)*M2
}
m222<-c(m22); m122<-c(m12)
Delta0<-M0[1:l]-t(B1)%*%c(rep(M0[1:l],lag)-m122)
Delta1<-M1-t(B1)%*%c(rep(M1,lag)-m222)
Delta2<-M2-t(B1)%*%rep(M2,lag)
return(list(m1=m1,m2=m2,M0=M0,M1=M1,M2=M2,Delta0=Delta0,Delta1=Delta1,Delta2=Delta2,Q=Q,s2=s2,S=S,B1=B1))

##Creating Jacobian Matrix for standard errors
Jacob<-function(a,lag,1,v)
{
 J<-function(a)
{
 S<-BlockTop(a[((1*(1+1)/2)*(1-1)+1):((lag+3)*(l+1)*l/2)],1,lag)
 S11<-S[1:l,1:l]; S12<-S[1:l,(l+1):(l*(lag+1))]; S22<-S[(l+1):(l+1):(l*(lag+1)),(l+1):(l*(lag+1))]
 SS<-solve(S22); Q<-SS/v; B1<-SS%*%t(S12); s2<-S11-S12%*%SS%*%t(S12)
 M0<-rep(c(a[((lag+3)*1/2):1]]); ((lag+3)*1/2)+1)]; (lag+3)*1/2)+2*1])
 M2<-c(a[((lag+3)*1/2)+1]+1)]; ((lag+3)*1/2)+2*1])
 if(tre="C") M1<-M2<-c(rep(0,1)
 if(tre="L") M2<-c(rep(0,1)
 m12<-m22<-matrix(nrow=1,ncol=lag)
 for(s in 1:lag)
{
 m22[,s]<-2*s*M2/1000
 m1[,s]<-s*M1/1000+((s/1000)^2)*M2
}
m222<-c(m22); m122<-c(m12)
Delta0<-M0[1:l]-t(B1)%*%c(rep(M0[1:l],lag)-m122)
Delta1<-M1-t(B1)*%o%rep(M1,lag)-m222 ; Delta2<-M2-t(B1)*%o%rep(M2,lag)
if(tre="Q") Cc<-c(Delta0,Delta1,B1,vech(s2))
if(tre="L") Cc<-c(Delta0,Delta1,B1,vech(s2))
if(tre="C") Cc<-c(Delta0,B1,vech(s2))
Cc
}
return(list(J=jacobian(J,a)))

## M-S tests for St-VAR model
MS<-function(U,lag,tre,s2,Ct,Ctt,muy,muyy,T,q,v)
{
  l<-nrow(U)
  st<-rmvst((T-lag),l,rep(0,l),Omega=diag(1,l),alpha=rep(0,l),df=v*lag+1)
  res<-matrix(nrow=(T-lag),ncol=l)
  for(i in 1:(T-lag))
  {
    spanos residual
    res[,i]<-U[,i]/sqrt((q*diag(s2)*Ct[i])) -sqrt((q*diag(s2)*Ct[i]))*st[,i]
  }
  t<-seq(1,T)
  ##MS tests for StVAR###
  hom<-tinv2<-Ind2<-Lin<-tinv<-Ind<-vector(length=l)
  for(i in 1:l)
  {
    polt<-poly(t,4,raw=FALSE)
    yhatt<-muyy[,i] ; y.hatt<-poly(yhatt,2,raw=FALSE)
    ms.g<-lm(res[(3:(T-(lag))),i]~polt[3:(T-(lag))]+y.hatt[3:(T-(lag))]+res[(2:(T-(lag+1))),i]+res[(1: (T-(lag+2))),i])
    cf<-names(coef(ms.g))
    Lin[i]<-linearHypothesis(ms.g,cf[7])[,2,6] ##Linearity
    if(tre="Q") tinv[i]<-linearHypothesis(ms.g,cf[4:5])[,2,6] ##t-invariance
    if(tre="L") tinv[i]<-linearHypothesis(ms.g,cf[3:5])[,2,6] ##t-invariance
    if(tre="C") tinv[i]<-linearHypothesis(ms.g,cf[2:5])[,2,6] ##t-invariance
    Ind[i]<-linearHypothesis(ms.g,cf[8:9])[,2,6] ## Independence
    var.h<-q*s2[1,i]*Ct[1:(T-lag)]
    var.hatt<-q*s2[1,i]*Ctt[1:(T-lag)] ; var.hb<-poly(var.hatt,2,raw=FALSE)[,1:2]
    if(tre="Q") tr<-6
    if(tre="L") tr<-4
    if(tre="C") tr<-2
    poltt<-poly(t,tr,raw=FALSE)[,1:tr]
ms.g2<-lm(I(res[3:(T-lag),i]^2)~var.h[3:(T-lag)]+var.hh[3:(T-lag),2]*var.h[2:(T-lag-1)]+var.h[1:(T-lag-2)]+poltt[3:(T-lag),(tr-1):tr])
cf<-names(coef(ms.g2))
if(tre=='Q') tinv2[i]<-linearHypothesis(ms.g2,cf[6:7])[2,6] ##t-inv
if(tre=='L') tinv2[i]<-linearHypothesis(ms.g2,cf[6:7])[2,6] ##t-inv
if(tre=='C') tinv2[i]<-linearHypothesis(ms.g2,cf[6:7])[2,6] ##t-inv
hom[i]<-linearHypothesis(ms.g2,cf[3])[2,6] ##homo
Ind2[i]<-linearHypothesis(ms.g2,cf[4:5])[2,6] ##Indep
}
##Skewness-Kurtosis test
res1<-matrix(nrow=T-lag-2,ncol=1)
for(i in 1:l)
{
 res1[,i]<-res[res[,i]!=max(res[,i]) & res[,i]!=min(res[,i]),i]
}
sk<-skewness(res1) ; kt<-kurtosis(res1)
N<-nrow(res1) ; kap<-6/(v+lag*l-4)
SK2<-(N/(6*(1+kap)))*sk^2+(N/(24*(1+kap)))*((kt-3-kap)^2)
p_chi<-pchisq(SK2,2,lower.tail=FALSE)
Students<-rbind(sk,kt,SK2,p_chi)
kolmo<-matrix(nrow=2,ncol=1)
for(i in 1:l)
{
 ks<-ks.test(res[,i],"pt",df=v+lag*l,alternative="greater")
kolmo[,i]<-rbind(ks$statistic,ks$p.value)
}
Dist<-round(matrix(rbind(kolmo[1:2],Students[3:4]),4,1,dnames=list(c("kolmo","p_kolmo","SK","p_SK"),c(rownames(U)))),4)
MS<-round(matrix(rbind(Ind,Inv,hom,Ind2,tinv2),6,1,dnames=list(c("Independence","Linearity","t-Invariance","Homosk","2nd Independ","2nd t-Inv"),c(rownames(U)))),4)
return(list(MS=MS,Dist=Dist))