© University of Cape Coast https://erl.ucc.edu.gh/jspui

UNIVERSITY OF CAPE COAST

A MONTE CARLO COMPARISON OF MINIMUM DESCRIPTION LENGTH MODEL SELECTION CRITERIA FOR ASYMMETRIC PRICE TRANSMISSION MODELS

IRENE KAFUI VORSAH AMPONSAH

2019

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

© Irene Kafui Vorsah Amponsah

University of Cape Coast

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

UNIVERSITY OF CAPE COAST

A MONTE CARLO COMPARISON OF MINIMUM DESCRIPTION LENGTH MODEL SELECTION CRITERIA FOR ASYMMETRIC PRICE TRANSMISSION MODELS

BY

IRENE KAFUI VORSAH AMPONSAH

Thesis submitted to the Department of Statistics of the School of Physical Sciences, College of Agriculture and Natural Sciences, University of Cape Coast, in partial fulfilment of the requirements for the award of Doctor of Philosophy degree in Statistics

JULY 2019

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

DECLARATION

Candidate's Declaration

I hereby declare that this thesis is the result of my own original research and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature...Date..10../03/2020.

IRENE KAFUI VORSAH AMPONSAH (MRS)

Supervisors' Declaration

We hereby declare that the preparation and presentation of the thesis were supervised in accordance with the guidelines on supervision of thesis laid down by the University.

Principal Supervisor's Signature..... .... Date..10../03/2020.

PROF. HENRY DE-GRAFT ACQUAH

Co-Supervisor's Signature.. Date...10./03/2020.

DR. NATHANIEL HOWARD

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

ABSTRACT

The Minimum Description Length (MDL) provides an attractive basis for statistical inference and model selection. However, little is known about the relative performance of its different formulations in Asymmetric Price Transmission (APT) modelling framework. To explore these issues, the study investigates different formulations of the MDL against commonly used alternatives (AIC and BIC) in terms of their ability to recover the true asymmetric data generating process (DGP) under various models, error sizes, asymmetric adjustment parameters and sample size conditions. Monte Carlo simulations results indicate that the performance of model selection method depend on sample size, level of asymmetry, noise levels and model complexity. The results further indicate that the different formulation of MDL, AIC and BIC all points to the true data generating process and clearly identifies the true model. In larger samples, rMDL is comparable to BIC and outperforms gMDL, nMDL, eMDL and AIC. At higher noise levels, AIC is comparable to eMDL and outperforms gMDL, nMDL, rMDL and BIC. AIC is comparable to nMDL and outperforms rMDL, gMDL, eMDL and BIC at strong levels of asymmetry. Empirically, application of a more complex model or increase in the number of asymmetric adjustment parameters improves the recovery of the true data generating process by the model selection methods. These results suggest that MDLs are very reliable and useful criteria in Asymmetric Price Transmission modelling. To achieve optimal APT linear models, one should always aim at stronger levels of asymmetry, lower noise and moderate to large samples.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

KEY WORDS

Asymmetric Price Transmission

Minimum Description Length

Model selection

Monte Carlo Simulation

Recovery Rate

True Data Generating Process

Digitized by Sam Jonah Library

© University of Cape Coast   https://erl.ucc.edu.gh/jspui

ACKNOWLEDGEMENTS

I am most grateful to God for all His providence throughout this journey of testimonies. Thank you Jesus. I am very much indebted to my principal supervisor, Prof. Henry De-Graft Acquah for his guidance, encouragement, and immerse understanding of a mother, lecturer and student through the good and bad times on this academic journey. The constructive criticisms were mixed with laughter and goodwill. God bless you abundantly.

I continue to thank Dr Nathaniel Howard the co-supervisor of this work. God bless you for all your guidance, constructive criticism and encouragement through this journey. I am grateful to the University of Cape Coast for financially supporting my PHD studies. I also appreciate the encouragement from all staff of both the Department of Statistics and Department of Mathematics of the University.

I especially thank my children for understanding a student mother who sometimes needs space to work and also my dear husband for his emotional support. I thank all who in diverse ways helped to bring this study to fruition.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

DEDICATION

To

Gideon Mensah

Alfred and Margaret Reynolds

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

TABLE OF CONTENTS

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Page

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

LIST OF TABLES

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

LIST OF FIGURES

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

LIST OF ACRONYMS

MDL	Minimum Description Length
NML	Normalized Maximum Likelihood
RNML	Renormalized Maximum Likelihood
APT	Asymmetric Price Transmission
ECM	Error Correction Model

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

LIST OF SYMBOLS

$\boldsymbol{\beta}$	The coefficient vector
$\hat{\boldsymbol{\beta}}$	$\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T y$, estimate of $\boldsymbol{\beta}$
D	The data set $\{x_1, x_2, \ldots, x_n\}$
h	The number of symbols to encode
$\mathbf{H}$	The hat matrix, defined as $\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$
$H(X)$	The entropy of a given data sequence
Hy	The predicted or fitted values
$\mathbf{I}$	Information matrix
k	The number of parameters in a given model
λ	A model or code
$L(x_i)$	The length of the code word assigned to symbol x_i
m	The number of trials in a binomial experiment
n	The number of observations
Ω	Is the population of possible values that can be observed
$\hat{P}_{ML}(D)$	The Maximum Likelihood estimate of the probability of the data
$P(x_i)$	The probability of symbol x_i
Q	Quartile
$q(D)$	The in hind-sight, optimal way of encoding D
U	The set of possible numbers that can be used to form a code word
$\lvert U \rvert$	The number of elements in the set U
W	Weight function

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

$\mathbf{X}$	Matrix containing the observed values of the variables in the model
z	A set of observed data in Ω
i.i.d	independent and identically distributed

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

CHAPTER ONE

INTRODUCTION

Model selection uses information criteria to achieve its goal and some new or different formulations of the old and popular methods are making great strides as researchers continue to find ways to choose the best model. This chapter seeks to provide guidance through the subject matter, emphasize the need for this study and states the objectives that will be worked on so as to make a meaningful contribution to knowledge.

The Background of the Study

Breiman (2001) explained that the extraction of information about a principal structure producing the data and the prediction of responses for future regressor variables are the two goals of analysing data. Modelling statistical data is a step in achieving these goals and so that the pattern in the observed data are revealed. Sampling is done from statistical populations from which data are drawn across a wide variety of disciplines.

Every population is distinctly associated with a probability distribution and an underlying method is estimated from data based on the specified family of distribution models. The performances of the competing models are then assessed and the best one selected. "The best model would have high prediction performance, and could illustrate which predictor variables are important and how these predictors affect the response of the underlying system" (Breiman, 2001). Stability, efficiency and consistency are used to evaluate the performance of the selection methods. To select the models, an algorithm is developed to create candidate models and also find a criterion for

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

classifying the contending models. Thus, the answer lies in information theory.

Information theory is a basic data communication tool which relates to the methodological methods that encode a signal for transmission, and the code gives a statistical description of the message produced. In reference to the work of the American electrical engineer Claude Shannon in the mid-twentieth century, information theory is also defined as a mathematical representation of the parameters and conditions affecting the processing and transmission of information. Information theory makes statistical inference and model selection very attractive (Markon & Krueger, 2004). However, not much is known in relation to the comparative performances of different Information-Theoretic Fit Criteria (ITFC) in certain areas of research such as linear models of econometric price transmission and asymmetric price adjustments.

Asymmetric Price Transmission (APT), informally called "rockets and feathers" occurs in price analysis when a decrease in price responds differently to an increase in price change, and this behaviour is dependent on the characteristics of changes in those prices (Meyer & Von Cramon-Taubadel, 2004). A typical example is when transportation fares increase quickly whenever prices of fuel increase, but the fares take time to decrease after fuel price decreases.

APT is said to occur when the regulation of prices is not consistent with respect to external or internal characteristics to the system (Tappata, 2008). Other examples explaining this asymmetric phenomenon include:

- Increases in prices of raw materials or farm produce lead to immediate increases in processed good or by-product prices, but decreases in

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

prices of raw materials slowly or do not equally translate into decrease in prices of processed goods. This asymmetry is referred to by Wlazlowski (2003) as time asymmetry, or

- The combination of the time asymmetry and the size asymmetry such that, in a situation when increases in prices of raw materials lead to bigger changes (in absolute values) in prices of finished goods than decreases. Notably, size asymmetry cannot occur on its own, otherwise price increases and price decreases would drift apart (GAO, 1993).

The consequences of the aforementioned (Peltzman, 2000) are that, the issue of APT has serious considerations in economic works. This is because it violates the theory of perfect competition and monopoly, and also market dynamics such as size, worldwide reliance on some products (e.g. oil) and the average household expenditure on some products. APT is therefore an important welfare problem and this has serious political and social implications.

Evidently, asymmetry is a subject of concern for linear econometric models and the way forward is to solve the problem with information criteria. Some information-theoretic fit criteria have been used extensively (the popular and widely used ones and their various extensions), yet other numerous information criteria exist and are yet to be explored. Thus, the relative neglect of some (effective but not popular) ITFC in research is especially relevant to asymmetric price transmission (APT) modelling, where ITFC have conventionally played an essential role in selecting models. In price analysis, different methods identify asymmetry at alternate rates which results in unrelated inferences and conclusions (Meyer & Von Cramon-Taubadel, 2004;

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Capps & Sherwell, 2007). Nonetheless, the aim of APT modelling is to choose a particular model which best identifies the underlying asymmetric DGP among contending models (Acquah & Von Cramon-Taubadel, 2009). This stimulates the interest in model selection methods. A commonly used criterion today is the Akaike's Information Criterion (AIC), which in some sense has become the "standard" (Akaike, 1973). However, another criterion is Bayesian Information Criterion (BIC) (Schwarz, 1978, Acquah & Von Cramon-Taubadel, 2009), which also proved to be a good model selection tool. These two aforementioned criteria are built on the premises of crediting models that explains the data well, while penalizing complex models.

The field of information theory has recently developed a new model selection technique, called the Minimum Description Length (MDL) principle. MDL is a principle of data compression. The basic idea is that understanding data is the same as finding a structure in it and thus being able to compress it. The reasoning behind this is that there is a one-to-one correspondence between code lengths in MDL and probabilities. In other words, there is no difference in maximizing a probability and minimizing a code length (Grünwald, 2007). The main idea behind this criterion, though much like the two others, is to find a simple, but best model that describes the data well. Consequently, there is a relationship between this goal and that of Asymmetric Price Transmission modelling mentioned earlier (i.e. selection of a particular model that best captures the true DGP among a group of contending models which leads to proper strategic deductions).

This idea of parsimony is not new; it can at least be traced back to the late thirteenth to early fourteenth century to scholar William of Ockham and the

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

more popular Ockham's razor principle that has been named after him. He has been said to have stated that "it is vain to do with more what can be done with fewer." (Li & Vitányi, 1997, p. 317) Ockham's razor is a philosophical principle of parsimony. It, roughly speaking, states that when deciding between two equally plausible theories of how something works, one should settle for the one that is simpler (Britannica, 2016). Implicitly, the principle of parsimony inspires the method of data analysis and statistical modelling, and it is also the soul of model selection. In the past, the need for selecting a model arises when researchers have to choose among the model classes based on data. MDL stems from an algorithmic or descriptive complexity theory of Kolmogorov, Chaitin and Solomonoff (Li & Vitanyi, 1997). The new classification of probability based on the length of the shortest binary computer program that describes an object (the descriptive complexity of the object) came about due to the desire to study the association between "mathematical formulations of randomness and their application to real world phenomena" (Li & Vitanyi, 1997).

MDL is based on the same kind of philosophy, where a short description of a phenomenon is preferred to a long description. In the words of Rissanen (2007) "the principle is very general, somewhat in the spirit of Ockham's razor, although the latter selects the simplest explanation among equivalent ones, while the MDL principle applies to all explanations of data whether equivalent or not".

There are several ways of modelling the data at hand, but in MDL, one is interested in the model generating the shortest code length (Grünwald, 2007). The model that provides the shortest code length is considered to be the "best"

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

model. Whether or not this is done using the "true" distribution is irrelevant (Grünwald, 2007). It is important to remember that, as Grünwald eloquently expresses it: "MDL is a *methodology for inferring models from data, not a statement about how the world works!*"

One major challenge is the inadequate knowledge of ITFC in APT modelling and this is coupled with poor understanding of how they work within econometric models. Emphasis is on criteria that use software or programming language to develop the algorithms. The underlying principle of ITFC is to choose the simplest statistical model that gives the true description of the data and model. Precisely, ITFC accentuate the amount of information required to be the most minimal. Thus, the most efficient and simplest model is selected.

Statement of the Problem

Price transmission and asymmetric adjustment analysis has come a long way in areas of model testing, specification and estimation (Acquah & Von Cramon-Taubadel, 2009) and econometric models of APT are no exceptions. Among these models are Wolfram's specification econometric model (1971) which was later improved by Houck (1977), the error correction model (Von Cramon-Taubadel, 1998) and models with a threshold (Cook, 2003; Goodwin & Serra, 2003; Cook & Holly, 2002; Abudulai, 2002; Goodwin & Piggott, 2001; Goodwin & Harper, 2000; Goodwin & Holt, 1999; Enders & Granger, 1998; Balke & Fomby, 1997; Tsay, 1989;). Although, these models have been applied extensively in APT modelling framework, there is still the discrepancy in determining which is most suitable and under which conditions.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

The variant of the Houck's model, restricted to the pre-cointegration setting, contains description in first and recursive sum of first differences. The post-cointegration approaches identify modifications of the Error Correction Models (ECM) which comprises the Standard ECM (Von Cramon-Taubadel, 1996, 1998; Granger & Lee, 1989) and a Complex ECM (Von Cramon-Taubadel & Loy, 1996). Furthermore, various authors (Cook, 2003; Meyer, 2003; Hansen & Seo, 2002; Abdulai, 2002; Goodwin & Piggott, 2001) also stipulated variants of the threshold model.

Despite the aforementioned, different methods or models are constantly used to analyse price dynamics and test for asymmetric adjustments in deriving policy conclusions; they remain discordant with each other which might lead to differences in interpretation and conclusions. Von Cramon-Taubadel and Loy (1999) thus extended the application of the asymmetric ECM for cointegrated data and established that the ECM was more suitable than using the conventional Houck approach. Subsequently, Capps and Sherwell (2007) in a related work found that the inference and conclusions derived from the Von Cramon-Taubadel and Loy ECM approach were different from the conventional Houck's approach and their empirical application. This supports the assertion that variant selection methods do point to different deductions.

Furthermore, another challenge is that researchers use uncommon objectives in choosing the different approaches under different conditions. Consequently, the absence of arduous and mutual objectives of criteria used in selecting the models is insufficient in deriving solutions of problems of APT. In a related study, Meyer and Von Cramon-Taubadel (2004) also

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

demonstrated that different methods detect asymmetry at different rates. Accordingly, this study uses artificial data to demonstrate that different methods detect asymmetry at different rates to support Meyer and Von Cramon-Taubadel (2004), Acquah and Von Cramon-Taubadel (2009) among others.

Another issue raised by Manera and Frey (2007) was that no attempt has been made (specifically using MDL principle and perhaps some less known information criteria) to solve the problem of which of the various APT models are mostly consistent or better fits a given dataset, notwithstanding numerous empirical studies carried out. This implies, very little basis exists for selecting the alternate approaches. Also, different predispositions to data irregularities (that is the impact of structural break on the measures of asymmetry) meticulously explained in the work of Von Cramon-Taubadel and Meyer (2000) exist.

Amid several contending existing models and the ability to develop new ones and not giving account of choice of the statistical model, it is obvious that the issue of choosing a credible process from different asymmetry tests using model selection criteria is an important methodological issue within the econometric models of APT.

This study fills the gap by addressing the methodological issues of model comparison and selection of the different econometric models of APT. The contribution of the study to knowledge lies in theoretically introducing and developing the MDL principle in the analysis of the APT models and establishing their use across different APT methods or model specifications during model selection. This entails the establishment or use of MDL criterion

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

as efficient model selection criterion to decide between the different methods of testing for asymmetry. Verifiably, this study estimates and compares the performance of model selection algorithms in an APT modelling framework.

Although, AIC and BIC have been widely used and are popular among Asymmetric Price Transmission modelling framework, other ITFC (e.g. KIC, RIC, etc.) clearly outperform them. Variants of the MDL principle (nMDL, rMDL, eMDL, gMDL) were extensively reviewed and they are performing better than AIC, BIC and their extensions in other areas of research. Empirically, MDLs have been applied in areas of regression, non-parametric curve estimation, cluster analysis, time series analysis, structural equation modelling, and graph theory, among others.

Theoretically, the MDL principle measures codes instead of probabilities by generating the shortest code length which implies that the best model, irrespective of whether a true distribution exists (BIC) or the estimated model fits well (AIC), is irrelevant. Fowler et al. (2011) established that the MDL principle is robust and does not follow any distributional assumptions. Thus, the MDL principle provides a criterion for selection of models, regardless of their complexity, without the restrictive assumption that the data is a sample from a “true” distribution. Furthermore, the MDL principle aside removing noise is also a natural safeguard against overfitting. The aforementioned evidently emphasize the various contributions that MDLs can make in the advancement of model selection in any discipline.

Thus, the research contribution is not restricted to the introduction and application of the MDL principle, but extends the empirical comparison of

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

some model selection methods (AIC, BIC to MDL) in an asymmetric price transmission modelling framework for which no studies have been undertaken.

Research Objectives

The general aim of this study is to introduce and demonstrate the usefulness of MDL principle in model selection across asymmetric price transmission models. This involves evaluating the performance of the formulations of MDL in comparison to AIC and BIC in the APT modelling framework.

Specific Objectives

Specifically, this research seeks:

1. To develop and implement the MDL principle to asymmetric price transmission linear model. This involves the development of MDL (rMDL, eMDL, nMDL and gMDL) computational algorithms using R programming language. Thus, the R-functions were developed for the various formulations of the MDL in the context of APT models.
2. To submit selected econometric models of asymmetric price transmission to an MDL model selection technique by:

   (a) Evaluating the performance of Information Criteria (Various types of MDL, AIC, BIC) in identifying the true asymmetric data generating process.

   (b) Demonstrating the conditions which improve the ability of information criteria to recover the true asymmetric data generating process in a Monte Carlo Experimentation under various conditions.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Thus, (a) and (b) will be examined under conditions of sample size, difference in asymmetric adjustment parameters, model complexity and stochastic variance across all study samples using 1000 Monte Carlo Simulations.

Limitations and De-limitations

Essentially, simulation method is used in generating data which is based on parameters used from experimental designs of renowned researchers (Acquah & Von Cramon-Taubadel, 2009; Holly et al., 2003). The objective of this simulation is brought to bear in the concept of "repeated samples" and this is important because this study is generally interested in the inference (like most researchers all over the world). That is, one does not simply want to describe what is going on in the one sample data we have at hand, rather one wants to be able to generalize the patterns found in sample data to all of the observations that could have been in the sample.

The concept of "repeated samples" is just theoretical or hypothetical and with limited resources, we cannot administer the same survey many times to different sample. In fact, if sufficient funding to field surveys to multiple samples was available, most researchers would either just increase their sample size initially to avoid repetition and use different questions (Carsey & Harden, 2014). However, in other cases, collecting another sample may be impossible (for instance, we cannot create a new set of regions to collect economic data from 1957).

Monte Carlo simulation solves this problem in that, it allows the analyst to easily create many samples of data in a computing environment, and

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

then assess patterns that appear across those repeated samples. Hence, this research used simulated data and Monte Carlo simulation to help us examine the extent to which the correct model is selected since the true data generating process is known in the repeated samples.

Organisation of the Study

The research is contained in five chapters. Chapter One, the introduction of the study comprises the background, research problem and its significance, research objectives, limitation and de-limitation of the study. Scholarly literature was reviewed in the second chapter which looked at contributions so far made with information criteria (AIC, MDL, BIC and a few others) used as tools of model selection. Emphasis was laid on MDL though not popular, yet gaining grounds in other disciplines and performing better. The third chapter discussed the theoretical overview of concepts employed in this study. Firstly, an empirical comparison of the overview of information theoretic criteria and how they are related structurally was emphasised. Then, the theory behind the minimum description length was also discussed and linked to statistical model building and model selection in linear regression. The last section gave an overview of asymmetric price transmission (APT) linear models and examined the econometric models of APT and their different methods or formulations employed in practise. Subsequently, in the fourth chapter, R-functions were developed for the MDL algorithms (first time ever) and the remaining model selection criteria and all APT models that would be used in this research. Further, R-codes were written for all computation and simulations that were carried out to achieve the objectives of

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

this study. Then, finally the results of all analysis were presented and discussed. The last chapter, Chapter Five, presented the summary of all chapters in this work, conclusions drawn to that effect and some useful recommendations given.

Chapter Summary

An extensive background to the study was discussed across the dimensions of the goal of statistical modelling, information theory and asymmetric price transmission. Specifically, emphasis was laid on characteristics of price input and output amid welfare considerations in the light of price asymmetry.

AIC and BIC were identified as commonly used Information-Theoretic Fit Criteria in the field of econometric APT modelling framework. It is also emphasized although not popular with APT models, the Minimum Description Length (MDL) criterion, based on the principle of data compression and parsimony, were extensively used in other areas of research.

The study mentioned problems of discordancy in alternative methods constantly used to analyse and test APT models and hence leading to differences in interpretations and conclusions. Also researchers do not appear to use a common objective set of criteria to choose the different approaches employed in different settings. Therefore, this research introduces MDL and compares it with common alternatives to address the methodological issues of model comparison and selection in the alternative econometric models of APT.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

The problem of "repeated samples" will be solved by generating data through simulation. Thus, the researcher is able to create many samples of data in a computing environment and then assess patterns that appear across those repeated samples. The study is organized into five chapters.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

CHAPTER TWO

LITERATURE REVIEW

Introduction

This chapter discusses the contributions of other researchers regarding the introduction and performance of information theoretic fit criteria with emphasis on Minimum Description Length Principle (MDL) in areas of modelling.

Information Theoretic Fit in Model Selection

Research is the pillar of problem solving through the collection of information in the form of data which is analysed for further inference. The question is how does one decide among competing explanations of data given limited resources? This is the problem of *model selection*. It stands out as one of the most important problems of inductive and statistical inference. Mathematically, a statistical model is defined as a set of probability distributions on the sample space (Cox & Hinkley, 1974). The practice of statistical modelling undergoes continual transformation as a result of both methodological developments and progress in the computing environment. The high-performance computers facilitate widespread advances in the development of statistical modelling theory to capture the underlying nature of a phenomenon. It is evident that the amount of information has been increasing both in size and variety, thanks to recent advancement in science and technology. With the advancement of computers and the information age, the challenge of understanding vast amounts of complicated data has led to the development of various types of statistical models.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Myung (2000) in his research entitled, 'the importance of complexity in model selection', highlighted the fact that model selection should not only be based on goodness-of-fit, but must also consider model complexity. The author argued that in cognitive psychology, the goal of mathematical modelling is to select one model from a set of competing models that best captures the underlying mental process but choosing the model that best fits a particular set of data will not achieve this goal. He explained further that a highly complex model can provide a good fit without necessarily bearing any interpretable relationship with the underlying process. The study showed that model selection based solely on the fit to observed data will result in the choice of an unnecessarily complex model that overfits the data, and thus generalizes poorly. He also used artificial data to explain selection methods. The study concluded that, to avoid selecting a powerful model, with little scientific significance, model selection should not be based solely on a model's ability to fit a particular sample data but instead should be based on its ability to capture the characteristics of the population; that is, its ability to generalize, and this leads to the correctness of the model.

There has been widespread use of information theoretic fit criteria, also known as information criteria, over many decades which take care of model complexity. Most researches do prefer information criteria (like AIC, BIC, DIC, MDL, etc. and their various extensions) in model selection. Comparison tests (such as the likelihood ratio, Lagrange multiplier, or Wald test, etc.) are only appropriate for comparing nested models, whilst, information criteria are model selection tools that can be used to compare any model fit to the same data. Despite their theoretical appeal and widespread use, there has been few

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

empirical investigations of the performance of information-theoretic criteria (especially) in most areas of research (e.g. structural equation modelling and asymmetric price transmission analysis). Even fewer investigations have compared the performance of different information-theoretic criteria, or examined the performance of information-theoretic criteria with regards to price models in particular. Most comparisons of information-theoretic criteria have focused on their use with generalized linear models (Acquah, 2010), and have focused on comparisons between AIC, BIC, Deviance Information Criterion (DIC), CAIC, Kullback Information Criterion (KIC), Residual Information Criterion (RIC), and many others.

Contributions of Information Criteria in Research

Investigations regarding the two (AIC, BIC) most popularly used criteria have generally demonstrated that BIC is consistent, that is, it tends to choose the true model with probability equal to one in large samples but performs poorly in small samples (Bickel & Zhang, 1992; Hurvich & Tsai, 1990). AIC, in contrast, is not consistent, but performs relatively well in small samples.

In addition, theoretical work has illustrated that, with reference to general linear models, AIC should theoretically converge faster to a limiting distribution than BIC, but that the limiting distribution of AIC should result in less than perfect model selection (Zhang, 1993). Zhang further suggested that, at least with regard to general linear models, BIC is nearly optimal in its rate of convergence for a criterion that depends on sample size.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Subsequently, Markon and Krueger's (2004) empirical investigations of the performance of information-theoretic fit criteria with factor analytic models have generally been consistent with existing work on generalized linear models. Ichikawa (1988), for example, examined the performance of AIC in selecting the number of factors in exploratory factor analysis. Simulation results indicated that the ability of AIC to select a true model rapidly increased with sample size. However, at the largest sample sizes, AIC continued to exhibit a slight tendency toward selecting overfitted models.

Hui et al. (2011) similarly investigated the performance of BIC in the context of blind source separation (which can be treated as a form of factor analysis). Their results suggested that BIC performs poorly for small sample sizes, but improves with increasing sample size to eventually choose the correct model with perfect probability.

Various studies suggested that the relative performance of information-theoretic fit criteria may be strongly influenced by distributional misspecification. Ichikawa and Konishi (1999), for example, treated AIC as an estimate of expected log-likelihood in the presence of cross-validation. Their results indicated that, when distributions are correctly specified, AIC performs relatively well. When distributions are misspecified, however, the performance of AIC decreases substantially.

Simulations by Fishler et al. (2002) suggested that the performance of BIC is relatively robust to distributional misspecification. Overall, existing work on information-theoretic criteria suggests that AIC performs relatively well in small samples, but is inconsistent and does not improve performance in large samples. BIC, in contrast, appears to perform relatively poorly in small

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

samples, but is consistent and improves in performance with sample size. There is some evidence to suggest that the relative performance of AIC and BIC is affected by distributional misspecification, such that the performance of AIC decreases substantially with misspecification, but BIC is robust to such effects.

Few studies have examined the performance of different information-theoretic criteria in a structural equation modelling context. Most investigations into the performance of information theoretic criteria with latent variable models have been limited to studies of latent class modelling, where likelihood-ratio tests are not appropriate. These studies have generally suggested that BIC outperforms AIC, in that it often exhibits lower error rates and greater power.

Information theoretic fit criteria (AIC, BIC, etc.) have also been applied extensively to asymmetric price transmission modelling framework but no work has been done using MDL and its extensions. These articles helped to show how the methods worked and also compares two or more criteria in selecting the best model given some conditions.

A typical example was a research on the comparison of AIC and BIC criteria in selection of asymmetric price relationship (Acquah, 2010). The study evaluated the performance of the two commonly used model selection criteria, AIC and BIC in discriminating between asymmetric price transmission models under various conditions. Monte Carlo experimentation indicated that the performance of the different model selection criteria are affected by the size of the data, the level of asymmetry and the amount of noise in the model used in the application. They concluded that BIC is

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

consistent and outperforms AIC in selecting the suitable asymmetric price relationship in large samples.

In recent times, 'the effect of outliers on the performance of AIC and BIC criteria in the selection of an asymmetric price relationship' was explored by Acquah (2017a). He argued that the goal of APT modelling, which is to select one model that best captures the asymmetric data generating process from a set of competing models, was affected by the presence of outliers in the data and this had a disproportionate impact on model ranking. The effect of outliers on the commonly used AIC and BIC in the selection of asymmetric price relationship were evaluated under conditions of different sample sizes. Monte Carlo experimentation revealed that the ability of the model selection methods to identify the true asymmetric price relationship decreased with increase in outliers in moderate and large samples. However, the effect of outliers on the performance of AIC and BIC in the selection of correct asymmetric model remain unclear with small samples.

Similarly, the Kullback's direct divergence (AIC and AIC corrected, AICc) was used extensively in APT model selection when it comes to choosing the optimal model among candidate models in econometrics (Acquah, 2017b). His article, 'criteria for APT model selection based on Kullback's symmetric divergence', emphasized alternative criteria that targets Kullback's symmetric divergence (KIC and KIC corrected, KICc) remains unexplored. Therefore, the article through Monte Carlo study evaluated the relative performance of the recently developed selection criteria based on KIC and KICc against commonly used alternatives; AIC and AICc in terms of their ability to recover the true asymmetric data generating process. The study

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

revealed that the performance of the model selection methods were influenced by the sample size, level of asymmetry and the amount of noise in the model used in the application. KICc was comparable to KIC and they both outperformed AIC and AICc in both small and large samples. Regarding the noise levels and difference in asymmetric adjustment parameters, again, KICc was comparable to KIC and they both outperformed AIC and AICc. The study concluded that the criteria based on Kullback's symmetric divergence were very reliable and useful criteria in asymmetric price transmission model selection.

However, these studies have also generally concluded that neither criteria AIC and BIC with their various extensions were completely satisfactory, and that other fit criteria exhibited greater power to select true models (Celeux & Soromenho, 1996; Lin & Dayton, 1997; Yang & Barron, 1998). Interestingly, Monte Carlo evidence suggests that early MDL approximations (Rissanen, 1978), essentially corrected forms of BIC, performed well in selecting latent class models (Yang & Barron, 1998).

Contributions of MDL in Research

Research has been done to explain the usefulness of the Minimum Description Length (MDL) Principle and its various formulations in model selection in some disciplines. The method has been explained from data compression's (computational statistics) point of view and how this basic principle, coming from information theory can be linked directly with information theoretic fit methods and statistical deduction.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Theoretical advancements

The fundamental idea behind the MDL principle is that any regularity in a given dataset can be used to compress the data, which is to describe it using fewer symbols than needed to describe the data literally (Grünwald, 2007). Generally, in Rissanen's formulation of MDL, any probability distribution is measured from a descriptive point of view and that it is not essentially the underlying data-generating mechanism (although it does not exclude such a possibility). Thus, MDL extends the more traditional random sampling approach to modelling. Many probability distributions can be compared in terms of their descriptive power and if the data in fact follow one of the models, then Shannon's celebrated source coding theorem (Cover & Thomas, 1991) states that this 'true' distribution gives the minimum description length of the data (on the average and asymptotically). Theoretic research has illustrated the usefulness of MDL:

A study conducted by Sewell (2007) used the MDL principle to minimize the sum of the length, in bits, of an effective description of the model and the length, in bits, of an effective description of the data when encoded with the help of the model.

Subsequently, in a seminal paper: "Modelling by shortest data description' (Rissanen,1978), the number of digits it takes to write down an observed sequence $x_1, x_2, x_3, ..., x_n$ of a time series which depended on the model with its parameters that one assumes to have generated the observed data was investigated. Accordingly, by finding the model which minimizes the description length, one obtains estimates of both the integer-valued structure parameters and the real-valued system parameters". The author further

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

indicated that MDL (Rissanen, 1978) uses an information theoretic measure known as Kolmogorov complexity. *Kolmogorov complexity* of a dataset is defined as the shortest description of the data.

Alpaydin (2004) also reveals that “if data is simple, it has a short complexity; for example, if it is a sequence of “0”s, we can just write “0” and the length of the sequence. If the data is completely random, then we cannot have any description of the data shorter than the data itself. If a model is appropriate for the data, then it has a good fit to the data, and instead of the data, we can send/store the model description. Out of all the models that describe the data, we want to have the simplest model so that it lends itself to the shortest description. So, we again have a trade-off between how simple the model is and how well it explains the data."

Further, Christianini and Shawe-Taylor (2000), noted that "as an example, the MDL principle proposes to use the set of hypotheses for which the description of the chosen function together with the list of training errors is shortest." "It is sometimes claimed that the minimum description length principle provides justification for preferring one type of classifier over another; specifically, “simpler” classifiers over “complex” ones. Briefly stated, the approach purports to find some irreducible, smallest representation of all members of a category (much like a “signal”); all variation among the individual patterns is then “noise.” The argument is that by simply fitting recognizers appropriately, the signal can be retained while the noise is ignored."

Duda, Hart and Stork (2001) also explained that, “*MDL principle* states that we should minimize the sum of the model’s algorithmic complexity and

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

the description of the training data with respect to that model". "Intuitively, MDL principle can be thought of as "recommending the shortest method for re-encoding the training data, where we count both the size of the hypothesis and any additional cost of encoding the data given this hypothesis" (Mitchell, 1997).

Further, the minimum description length principle is a formalization of Ockham's Razor in which the best hypothesis for a given set of data is the one that leads to the largest compression of the data. MDL was introduced by Rissanen in 1978; it is important in information theory and learning theory. Subsequently, according to Mackay (2003), "The MDL principle (Wallace and Boulton, 1968) states that one should prefer models that can communicate the data in the smallest number of bits".

Grünwald compiled some works on The Minimum Description Length Principle and Reasoning under Uncertainty (Grünwald, 2007). Firstly, Poland and Hutter, (2005) researched into Asymptotic of Discrete MDL for Online Prediction. Minimum description length (MDL) is an important principle for induction and prediction, with strong relations to optimal Bayesian learning. This paper dealt with learning processes which were independent and identically distributed (i.i.d.) by means of two-part MDL, where the underlying model class is countable. They considered the online learning framework (observations come in one by one, and the predictor is allowed to update its state of mind after each time step). They also identified two ways of prediction by MDL for this setup, namely, a static and a dynamic one. Where they proved that under the only assumption that the data is generated by a distribution contained in the model class, the MDL predictions converge to the

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

true values almost surely. This was accomplished by proving finite bounds on the quadratic, the Hellinger, and the Kullback–Leibler loss of the MDL learner, which are, however, exponentially worse than for Bayesian prediction. They demonstrated that these bounds are sharp, even for model classes containing only Bernoulli distributions and showed how these bounds imply regret bounds for arbitrary loss functions. The results apply to a wide range of setups, namely, sequence prediction, pattern classification, regression, and universal induction in the sense of algorithmic information theory among others.

Secondly, Poland and Hutter (2005) did another work on Convergence of Discrete MDL for Sequential Prediction where the study concentrated on the properties of the Minimum Description Length principle for sequence prediction, considering a two-part MDL estimator which is chosen from a countable class of models. This applies in particular to the important case of universal sequence prediction, where the model class corresponds to all algorithms for some fixed universal Turing machine (this correspondence is by enumerable semi-measures; hence the resulting models are stochastic). They proved convergence theorems similar to Solomonoff's theorem of universal induction, which also holds for general Bayes mixtures. The bound characterizing the convergence speed for MDL predictions is exponentially larger as compared to Bayes mixtures. They observed that there are at least three different ways of using MDL for prediction. One of these had worse prediction properties, for which predictions only converge if the MDL estimator stabilizes and hence they established sufficient conditions for this to

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

occur. Finally, some immediate consequences for complexity relations and randomness criteria were proven.

Hansen and Yu (2001) reviewed the minimum description length (MDL) for problems of model selection. They viewed statistical modelling as a means of generating descriptions of observed data, and thus the MDL framework discriminates between competing models based on the complexity of each description. Previous studies have confirmed that the aforementioned approach began with Kolmogorov's theory of algorithmic complexity, matured in the literature on information theory and has recently received renewed attention with the statistics community. Their research reviewed both the practical and theoretical aspects of MDL as a tool for model selection, emphasizing the rich connections between information theory and statistics. They argued that the boundary between these two disciplines were many interesting interpretations of popular frequentist and Bayesian procedures. Hence, MDL provides an objective umbrella under which rather desperate approaches to statistical modelling can co-exist and be compared. They illustrated the MDL principle by considering problems in regression, non-parametric curve estimation, cluster analysis and time series analysis. They emphasized in their work that since model selection in linear regression was an extremely common problem that arises in many applications, they derived several MDL criteria in this context and discussed their properties through a number of examples using real datasets. Classical problems in model selection were examined and also tried to apply MDL to more exotic modelling situations in engineering. In general, they concluded that as a matter of

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

principle for statistical modelling, the strength of MDL was that it can be intuitively extended to provide useful tools for new problems.

Pitt et al. (2002) also asked the unavoidable question 'how one should decide among competing explanations of data' in their article entitled "Toward a method of selecting among computational models of cognition". They argued that computational model of cognition are increasingly being advanced as explanations of behaviour and that the success of this line of inquiry depended on the development of robust methods to guide the evaluation and selection of these models. Their article introduced 'a method of selecting among mathematical methods of cognition' known as MDL, which provided an intuitive and theoretical well-grounded understanding of why one model should be chosen. The application of the adequacy of MDL was demonstrated in cognitive modelling in the areas of psychophysics, information integration and categorization.

The method of MDL (specifically, predictive MDL and Bayesian method of selection) from a theoretical point of view is also used by some researchers as bases to develop new methods. A typical example is the research conducted by Wagenmakers et al. (2006), which is entitled "accumulative prediction error and the selection of time series models". They devised the accumulative one-step-ahead prediction error (APE) as a data-driven method for model selection for very complex models. 'The APT method automatically takes the functional form of parameters into account, and the "plug-in" version and does not require the specification of priors. It was applied to data with natural ordering (time series) where they explored the possibility of using it to discriminate the short-range ARIMA (1, 1) model

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

from the long-range ARIMA (0, *d*, 0) model. They also applied it to model selection, which allows one to choose between different model selection methods.

Empirical Comparison

Research has proved that AIC and BIC (and their various extensions) are widely used criteria in various disciplines and some disciplines have introduced MDL, though less known, which is making great strides in contributing to the field of model selection. "The MDL principle is thus a relatively recent method for inductive inference. There is evidence of work done to compare the performance of the various types of MDL with some common and widely used information theoretic fit criteria (AIC, BIC, AICc, BICc, DIC, etc.) in some disciplines.

Sund (2001) in his lectures on statistical modelling (presented to research and development centre for welfare and Health [STAKES]), explored the minimum description length based on model selection in linear regression. He emphasized the fact that the soul of model selection is the principle of parsimony and this principle is in line with the rationale behind the MDL principle. He also emphasized that MDL has its roots in information theory and in the invariance theorem of Kolmogorov complexity but added computability challenges so arbitrary class of models should be used and do the coding with the help of model class. To achieve this, one must employ the most straightforward description length which is based on two-stage coding scheme by calculating the code length required to discretize model's parameter space and communicate the estimated parameter. Then in the second stage, the actual data string is coded using the distribution indexed by

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

the communicated parameter. He mentioned the various formulations of MDL (gMDL, nMDL, eMDL, PMDL) and then compared them to some criteria based on information theory (AIC, BIC, and their various extensions) in the linear regression modelling framework. Illustrating with a practical example, two linear regression models (*y* dependent variables) were built of ordinary model and hermit polynomial transformations of *x* regressors. He found some kind of functional dependency between the variables but the residuals were not normally distributed. However, proceedings with the analyses using straight forward computations via statistical package SURVO (SAS programming language), the results showed that all the selection criteria suggested the models were of fourth degree polynomials. Moreover, the ordinary and hermit polynomials gave the same description length for all criteria and according to the geometric interpretation of the exact MDL criteria are essentially just different parameterisations of the same model. Specifically, all formulations of MDL outperformed AIC and BIC in selecting both the ordinary and the hermit models using the given data.

Related studies also emphasize the fact that information theory offers a coherent, intuitive view of model selection. This perspective arises from thinking of a statistical model as a code, an algorithm for compressing data into a sequence of bits. Stine (2003) in his paper entitled “information theory and the MDL principle” explained that the description length is the length of this code for the data plus the length of the model to the data, whereas the length of the code for the model measures its complexity. He further explained that the minimum description length (MDL) principle picks the model with smallest description length, balancing fit versus complexity. Thus, Stine thinks

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

the conversion of a model into a code should be flexible; in that, one can represent a regression model, for example, with codes that reproduce the AIC and BIC as well as motivate other model selection criteria. Additionally, information theory allows one to choose among various types of non-nested models, such as tree-based models and regression identified from different sets of predictors. The results of analysis revealed that using AIC and BIC, the lurking problem of the best criterion depends on how many predictors that are 'useful' or work best for us. Recent research attempted to remove this 'need to know' by offering adaptive criteria that work almost as well if not better. So he concluded that MDL (also based on information theory) allows one to construct a wide range of model selection criteria, rewards theory and allows searches of large collection of models, hence providing an important heuristic guide to the development of customized criteria suited to the problem at hand. More importantly, MDL does not ask of the existence of a true model.

Markon and Krueger (2004) also emphasized the importance of further research into theory and computation of information-theoretic fit criteria in their paper entitled 'an empirical comparison of information-theoretic selection criteria for multivariate behaviour genetic models'. Thus, a particular condition or method (any criteria whether well-known or less known) shows strengths and weaknesses which needs investigation. Results indicated that performance depends on sample size, model complexity and distributional specifications. The Bayesian information criterion (BIC), is more robust to distributional misspecification than Akaike's information criterion (AIC) under certain conditions, and outperforms AIC in larger samples when comparing more complex models. Also, the study investigated the domain of

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

an approximation to the minimum description length (MDL) criterion, involving the empirical Fisher information matrix exhibited variable patterns of performance due to the complexity of estimating Fishers information matrices. Results showed that, Draper's information criterion (DIC) which shares features of the Bayesian and MDL criteria, performs similarly to or better than BIC. He argued that, MDL which is less used or explored was also competing very well in all specifications. He concluded that in the context of estimating information matrices based on MDL, NML and MDL estimators that do not rely on sample based estimates of the Fisher information matrix, although much more challenging to compute, may be more robust and efficient than the MDL estimators examined in their study. Other studies too have exhibited greater selection power of MDLs than that of AIC and BIC (Hansen & Yu, 2001). Additionally, the relative superior performance of DIC in this current study also suggested that other MDL estimators hold promise as model selection criteria.

Again, Myung et al. (2005) continued to support the fact that the minimum description length (MDL) principle is an information theoretic fit approach to inductive inference that originated in algorithmic coding theory. In their paper (which provided a tutorial review on the latest developments with special focus on NML) entitled, 'model selection by normalized maximum likelihood', data was viewed as codes to be compressed by the model. In their approach, models are compared on their ability to compress a dataset by extracting useful information in the data apart from random noise. They also emphasized the fact that the goal of model selection was to identify the model from a set of candidate models, which permits the shortest

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

description length (code) of the data. Further, they pointed out aside the original formalized problem by Rissanen who used the 'crude two-part-code' method in 1970s, many significant formulations of MDL have been derived especially in the 1990s with the culmination of the development of the refined 'universal code' method called the Normalized Maximum Likelihood (NML). The paper also provided an application example of the NML in cognitive modelling.

The last decade has seen some continuous use of different MDLs in various areas of research. Li et al. (2012) in their thesis work, entitled 'model selection via Minimum Description Length' emphasized the purpose and use of MDL. The study accentuated that MDL originated from data compression literature and has been considered for derivation of statistical model selection procedures and in the context of linear regression, most existing methods utilize the MDL principle focusing on models consisting of independent data. They considered data in the form of repeated measures using the MDL to focus on classical linear mixed-effect models. Their research objective was in two fold namely; concerns with population parameters and the other concerns with cluster/subject parameters. Regarding population level, they proposed a class of MDL procedures which incorporated the dependence structure within individual or cluster with data-adaptive penalties and enjoyed the advantages of BIC. When the number of covariates is large, the penalty term is adjusted by data-adaptive structure to diminish the under selection issue in BIC and try to mimic the behaviour of AIC. Theoretical justifications were provided from both data compression and statistical perspective and extensive numerical

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

experiments conducted to demonstrate the usefulness of the proposed MDL procedure on both population level and cluster level.

Fade et al. (2011) also addressed an original statistical method for unsupervised identification and concentration estimation of spectrally interfering gas components of unknown nature and number. They showed that such spectral minimizing can be efficiently achieved using information criteria derived from the MDL principle which outperformed the standard information criteria such as AICc and BIC. Their study emphasised within the context of spectroscopic applications that the most efficient MDL technique implemented showed good robustness to experimental artefacts.

Another study of statistical performance analysis using MDL source enumeration in array processing was carried out by Haddadi et al. (2010). Their study revealed that unfortunately, available theoretical analysis exhibited deviation from their simulation results. They then used the MDL source enumeration technique to present an accurate and insightful performance analysis for the probability of missed detection. They also showed that the statistical performance of the MDL is approximately the same under both deterministic and stochastic signal models. Their simulation results showed the superiority of the proposed analysis over available results.

In a related study in array response modelling, Costa et al. (2012) proposed a method for order selection using calibrated data. This method allowed for one to find the optimal number of basic functions for describing array steering vectors, according to the manifold separation principle. They argued that the proposed solution does not require heuristic design parameters and achieves asymptotically optimal (in the MSE sense) modelling of array

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

non-idealities from calibration measurements. They extended the normalized minimum description length (nMDL) to complex-value data and employed to choosing the optimal number of modes in the orthogonal decomposition of the array steering vector. BIC and the more recent exponentially embedded family (EEF) rules were employed as well. Extended simulations were carried out using a real-world antenna array and the various order selection rules were compared. The results indicated that the nMDL was consistent estimator of the optimal number of modes and its performance was close to the minimum MSE.

Relating to stochastic complexity-based model selection, Fade (2015) worked on false alarm rate control in optimal spectroscopy. He explained that stochastic complexity-based penalization criteria can prove efficient and robust in spectroscopy applications for unsupervised identification and concentration estimation of spectrally interfering chemical components. His paper showed how nMDL can be tailored to provide control of the detection performances in terms of probability of false alarm. Numerical experiments conducted on realistic simulated optical spectroscopy signals proved that the nMDL approach outperformed the standard information criteria (AIC, BIC) in terms of model selection performances. Moreover, the ability to control false alarm rates with the proposed modified nMDL criterion was demonstrated through simulations.

Likewise, Han et al. (2014) demonstrated how MDL plays a role in selection of one dependency estimators in Bayesian Network and overfitting criterion. The study explained that the Averaged One Dependency Estimator (AODE) integrated all possible Super-Parent-One-Dependency Estimators

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

(SPODEs) and estimated class conditional probabilities by averaging them. They argued that in an AODE network some redundant SPODEs may result in some bias of classifiers, as a consequence, it could reduce the classification accuracy substantially. So in their research a kind of MDL metrics was used to select SPODEs in whole or part and hence three classifiers were presented. The study revealed that the performance comparisons between them and AODE have shown not only the theoretical analyses were reasonable, but also efficient and effective. Experimental results indicated that the classifier using MDL score metrics had better performance than the original AODE, and at the same time, less overfitting. Additionally, further discussions and verifications of some properties of overfitting were also presented in the paper.

Similarly, another variant of MDL (gMDL) was proposed by Jiao et al. (2011) to source number estimation. In this paper, a source number estimator using the peak-to-average power ratio modified by Gerschgorin radii was proposed. The eigenvectors of the sample covariance matrix were first employed to calculate the peak-to-average power ratio and then a Gerschgorin transform was taken to the sample covariance matrix. Next, the new peak-to-average power ratio values modified by Gerschgorin radii were introduced to minimum description length criteria (PGMDL). Simulation results indicated that the proposed method shows better performance than the conventional detection methods such as AIC and MDL especially at lower signal to noise ratio (SNR), and has no affectedness by the intensity difference between the multiple sources.

MDL has seen its usefulness across various disciplines with various research areas and prediction is no exception. Harremoes and Brock (2018) in

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

their paper entitled Horizont Independent MDL, used conditional normalized maximum likelihood predictor to predict the future given the past. The authors explained that this strategy was however computationally involving and in general, depended on how many future symbols one want to predict. But for special exponential family models, the conditional normalized maximum likelihood predictor does not depend on the number of symbols that one wants to predict. In this case, the prediction strategy equals a Bayesian strategy based on Jeffrey's' prior. These special exponential families (Gaussian location family, the Gamma family and the inverse Gaussian family) can be characterized as those for which the conjugated exponential family has a saddle-point approximation, that is, exact after renormalization. They concluded that this approach can be used to construct exponential families with Horizont Independent MDL in higher dimensions.

Last but not least, application of MDL has also been extended to graph theory. Graph analytics was explained by Velampalli and Jonnalagedda (2017) as a useful tool for finding hidden patterns, relationships, similarities and anomalies in graphs. These tasks are useful in many application areas like protein analysis, fraud detection, health care, computer security, financial data analysis and many more. Therefore, MDL's property of universal coding or universal modelling is brought to bear in SUBstructure discovery (SUBDUE) algorithms to discover substructures. This paper applied MDL encoding to various graph datasets and in particular graph matching was solved using MDL. Further, comparative analysis was done to show how MDL value changed with respect to varying graph properties. A tool called subgen was

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

used to generate the graph datasets and statistical tests were applied to know which MDL values changed significantly.

Chapter Summary

The preview show widespread use of Information-Theoretic Fit Criteria (ITFC) in research. These include the AIC, BIC, MDL, DIC, RIC, SBIC, etc. They have been examined in their ability to recover the true asymmetric data generating process. The interest appears to stem from the fact that a good fit could be attained, especially using highly complex models, when in fact it lacks interpretable relationship with underlying process which leads to poor generalizations.

The literature appears unanimous on the performances of AIC and BIC. Results show that AIC has the tendency to select the true model over smaller samples size but fails in larger samples whilst the BIC rather performs better in larger samples. They have been examined in their ability to recover the true asymmetric data generating process.

It is also established that AIC and BIC are affected by extent of distributional misspecifications. As BIC is robust, performance of AIC decreases substantially with misspecifications. Sample size and model complexity are other influential conditions of performance. It points out that although ITFC is extensively applied to APT modelling framework, no work has been done using MDL and its extensions in comparing performance of types of MDL with commonly used ITFC. It is observed that the various methods are affected by sample size, level of asymmetry and amount of noise in the models used in the applications. It is clearly known that various methods

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

outperform the commonly used AIC and BIC in their ability to select true models. These are the KIC and KICc.

Then MDL was extensively reviewed. Various descriptions of the techniques are provided which basically explains the technique as a way of obtaining the simplest model that meets the shortest description. Various ways of application of the MDL have been explored and provides bases for developing new methods. It is widely applied in regression, non-parametric curve estimation, cluster analysis, time series analysis and graph theory.

The aforementioned contributions evidently emphasize the various contributions that MDL criterion can make in the advancement of model selection in any discipline, especially (in this study) to asymmetric price transmission analysis framework and adds on to knowledge in information theoretic fit model selection.

Thus, this research fills the gap by addressing the methodological issues of model selection and comparison in the alternative econometric models of asymmetric price transmission. The contribution to knowledge lies in theoretically introducing and developing the nMDL, rMDL, eMDL and gMDL to analysis of the asymmetric price transmission models within the context of a model selection and demonstrating their application across alternative asymmetric price transmission methods or model specifications during model selection. This entails the provision or application of Minimum Description Length algorithms as efficient model selection strategies to decide between the alternative approaches of testing for asymmetry. Empirically, the research evaluates and presents comparisons of the relative performance of the

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

model selection algorithms in an asymmetry price transmission modelling framework.

This will be the first incident of application of MDL in price analysis (especially the asymmetric price transmission) and perhaps other disciplines which have not been explored yet in research.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

CHAPTER THREE

RESEARCH METHODS

Introduction

This chapter introduces and explains the underlying concepts that guide this study namely: Information Theoretic Fit Criteria, Linear Models and Asymmetric Price Transmission. Emphasis is placed on MDL which is less known but gaining grounds in the area of model selection. Specifically, its development and introduction to linear models of Asymmetric Price Transmission (APT) against the backdrop of commonly used information theoretic fit criteria (AIC, BIC, etc.) is of utmost concern in this research.

According to Sund (2001), the purpose of all scientific studies is to search answers for the problem which is induced by inquisitiveness, practical need or the purposes of theory development keeping in mind that knowledge about our world is partial. Thus, one can state that the aim of statistical model fitting is to "understand" the system behind the studied phenomena via the observed data. Thus, the question of how one should decide among competing explanations of data is at the heart of the scientific enterprise. Over the decades, scientists have used an assortment of statistical tools to select among alternative models of data. However, there has not been an underlying theoretical framework to guide the enterprise and evaluate new developments.

Implicitly the principle of parsimony (or Ockham's Razor) has been the soul of model selection. To implement the parsimony principle, one has to quantify "parsimony" of a model relative to the available data. Applying this measure to a number of candidate models, the goal is to find a model, which is

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

a compromise between desirable yet conflicting properties: goodness-of-fit, generalizability and concision.

Overview of Information-Theoretic Criteria-An Empirical Comparison

Markon and Krueger (2004) argued that the basic principle of information theoretic model selection is to select statistical models that simplify description of the data and model. Specifically, information-theoretic methods emphasize minimizing the amount of information required to express the data and the model. This results in the selection of models that are the most parsimonious or efficient representations of the observed data.

There has been suggestion of a diversity of information criteria. In general, nonetheless, most information criteria can be considered as special cases of what Barron and Cover (1991) has called minimum complexity density estimators. Minimum complexity density estimators have the general form

$$i\left((X\left|m(\theta)\right)+i\left(m(\theta)\right)\right). \quad (1)$$

The first term represents the amount of information i required to express the data X, given the model of interest m with parameter θ. The second term represents the amount of information required to express the model itself. Generally, the first term is equivalent to the negative log-likelihood of the data calculated at the maximum likelihood estimate of the parameter. The second term can be thought of as a penalty for model complexity; it differs between different information-theoretic fit criteria, and usually uniquely defines a given criterion.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

As the goal of information-theoretic fit model selection is to select parsimonious models, models with the least criteria are selected. Information-theoretic inference is associated with likelihood-based inference, but has desirable properties that likelihood-based inference does not. Equation (1) shows that, comparing two models of the same class, the second term is the same for both models, and can be eliminated. Minimizing information thus coincides with maximizing likelihood for comparisons of nested models. Nevertheless, unlike likelihood-based procedures, information-theoretic methods can be applied to non-nested models as well. Barron and Cover (1991) further demonstrated that, minimum complexity density estimators satisfying certain general conditions are generally asymptotically consistent, that is, they will recover the true model in large samples, if the true model is one of the models under consideration (literature shows that AIC is an important example of an estimator that is not generally consistent).

Akaike's Information Criterion (AIC)

Akaike information criterion (AIC) (Akaike, 1974) is a penalized technique grounded on in-sample fit to estimate the likelihood of a model to predict or estimate future values (Mohammed et al., 2015). A good model is the one that has minimum AIC among all the other models. The AIC can be used for example to select between Standard Error Correction Model (SECM) and Complex Error Correction Model (CECM). AIC is the most popular and widely used information-theoretic criteria. Hypothetically, AIC derives from consideration of the Kullback–Liebler distance between a given model and the true model. Explicitly, AIC is an estimate of the relative expected Kullback-Liebler distance (The Kullback–Liebler distance is a function of the ratio of

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

two distributions, and can be thought of as reflecting the efficiency with which one distribution is approximated by another (Cover & Thomas, 1991) of a given model from the true model. It is defined as

$$AIC = -\ln\left(L\left(\mathbf{X}\middle|\hat{\theta}\right)\right) + k, \tag{2}$$

where, the first term is the negative maximum log-likelihood of the data $\mathbf{X}$ given the model parameter estimates, and k is the number of parameters in the model. As an estimate of relative Kullback–Liebler distance, AIC can be thought of as measuring the relative inefficiency of approximating the true model by the model of interest. Models producing smaller values of AIC can thus be thought of as more efficiently approximating the true model, where the true model is unknown. Another variant of AIC (CAIC/AICc) have also been extensively engaged in research since AIC is said to be bias in small samples (Takeuchi, 1976; Hurvich & Tsai, 1989; Sugiura, 1978; Burhnam & Anderson, 1998, Fujikoshi & Satoh, 1997).

Bayesian Information Criterion (BIC)

Bayesian information criterion (BIC) (Stone, 1979) is another criterion for model selection that measures the trade-off between model fit and complexity of the model. A lower AIC or BIC value indicates a better fit. AIC and BIC are currently among the most widely used information-theoretic fit criteria. BIC is usually explained in terms of Bayesian theory, especially as an estimate of the Bayes factor, the ratio of the posterior to the prior odds for two models, in comparisons of a model to a saturated model (Raftery, 1993; Schwarz, 1978). BIC is defined as

$$BIC = -\ln\left(L\left(\mathbf{X}\middle|\hat{\theta}\right)\right) + \frac{k}{2}\ln(N), \tag{3}$$

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

where N is again the sample size and k is the number of parameters in the model. As an estimate of the Bayes factor for the comparison of a model to the saturated model, BIC favours the saturated model when positive, and the alternative model when negative. If maximum likelihood is used to estimate parameters and the models are non-nested, then the Akaike information criterion (AIC) or the Bayes information criterion (BIC) can be used to perform model comparisons. The two criteria are very similar in form but arise from very different assumptions. The AIC is derived from information theory and it is designed to pick the model that produces a probability distribution with the smallest discrepancy from the true distribution (as measured by the Kullback–Liebler discrepancy). The BIC is derived from a large sample asymptotic approximation to the full Bayesian model comparison (Jerome et al., 2014). Research (Pauler, 1998; Draper, 1995) continues to explore further, various modifications of BIC or Schwarz information criterion (also known as SIC, SBC, SBIC) and many of these modifications have been typically incorporated into Deviance information criterion (DIC).

Minimum Description Length (MDL)

Although BIC is generally formulated in terms of Bayesian theory, it has another interpretation that is currently less well known. In addition to its Bayesian interpretation, BIC can be interpreted as an asymptotic estimate of the normalized maximum likelihood (NML) or minimum description length (MDL) criterion and was originally described by Rissanen (1983, 1986, 1989, 1996, 2001). The goal of MDL model selection is to minimize the amount of information (e.g., in bits) required to describe the data and the model. In a series of papers, Rissanen (1996, 2001) has demonstrated that the code length

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

of a dataset – the amount of information required to describe the data using a given model – can be expressed as the logarithm of the normalized maximum likelihood.

The normalized maximum likelihood is given by

$$NML = \frac{L\left(\mathbf{X}\middle|\hat{\theta}(\mathbf{X})\right)}{\int L\left(\mathbf{X}\middle|\hat{\theta}(\mathbf{X})\right)dX}, \tag{4}$$

where the numerator is the maximum likelihood and the integral in the denominator is taken over the sample space. The normalized likelihood can be thought of as the likelihood divided or ‘‘normalized’’ by the sum of the possible likelihoods. As an index of complexity, the denominator is consistent with intuitive notions of model complexity, in that more complex models will fit a large number of datasets equally well, producing large possible likelihoods. When all possible datasets are equally likely under a given model, the denominator will be large, and the normalized maximum likelihood will be small. Relationships between BIC and NML are made apparent when approximations to the NML are considered. The negative logarithm of the NML can be approximated using sum of squares by

$$MDL = -\ln\left(L\left(\mathbf{X}\middle|\hat{\theta}\right)\right) + \frac{k}{2}\ln\left(\frac{N}{2\pi}\right) + \ln\int\sqrt{|I(\theta)|}\,d\theta, \tag{5}$$

where $I(\theta)$ is the expected Fisher information matrix and the integral in the third term is evaluated over the parameter space (Barron et al., 1998; Myung, et al., 2000; Rissanen, 1996, 2001). As mentioned above in the criteria, the first term is the negative log-likelihood of the data evaluated at the maximum likelihood estimates. The second and third terms are measures of model complexity and do not depend on observed data. The second term reflects

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

what might be thought of as ‘‘parametric complexity,’’ or model complexity associated with number of parameters. The third term reflects what might be thought of as ‘‘structural complexity,’’ or model complexity associated with features beyond the number of parameters. The second term of Equation (5) can be expanded to produce an equation comprising BIC plus two terms that does not depend on sample size and become negligible in large samples.

Comparing this form of Equation (5) to Equation (3), it is evident that BIC is therefore asymptotically equivalent to MDL. Thus, in addition to its Bayesian interpretation, BIC can be thought of as an asymptotic estimate of the shortest description of the data given a model of interest. Comparisons between DIC which can be thought of as a more accurate version of BIC—and the MDL estimate in Equation (5) further reinforce relationships between the Bayesian and NML paradigms, as DIC approximates MDL even more closely than BIC. The negative logarithm of the NML and its MDL approximation given in Equations (4) and (5) are computationally challenging, given that the integrals are computed over the sample and parameter spaces, respectively. A more computationally tractable approximation to MDL (Rissanen, 1989; see also Hansen & Yu, 2001) is given by

$$MDL = -\ln\left(L\left(\mathbf{X}\middle|\hat{\theta}\right)\right) + \frac{1}{2}\ln\left|I(\hat{\theta})\right|, \tag{6}$$

where the first term is the negative logarithm of the maximum likelihood, and the second term includes the observed Fisher information matrix evaluated at the maximum likelihood estimates. It has been demonstrated that this formulation of MDL, like other MDL approximations, converges to the same value as BIC as sample size increases (Hansen & Yu, 2001). The observed

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Fisher information matrix in the second term can be calculated from the covariance matrix of parameter estimates.

There are various extensions of the MDL principle such as to achieve simplicity or rather improve them to achieve the best model. The next section delves into the basic theory of the MDL principle and tries to explain the key concepts underlying them.

Overview of Minimum Description Length Principle

In order to understand where MDL and its various extensions come from and thus why it is a reasonable criterion for model selection, a brief theoretical background on MDL and where it comes from is needed.

The principle of MDL draws heavily from data compression in terms of amount of information that is required usually in bits to describe the data and the model. Data compression is reducing the number of bits needed to represent data. This is achieved through a process of modifying, encoding or converting the bits structure of data in such a way that it consumes less space on disk. It enables reduction of the storage size of one or more data instances or elements. Data compression is also known as source coding or bit-rate reduction. Compressing data can save storage capacity, speed up file transfer, and decrease costs for storage hardware and network bandwidth.

Data Compression

Compression is performed by a program that uses a formula or algorithm to determine how to shrink the size of the data. For instance, an algorithm may represent a string of bits or 0s and 1s with a smaller string of 0s and 1s by using a dictionary for the conversion between them, or the formula

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

may insert a reference or pointer to a string of 0s and 1s that the program has already seen. An algorithm is a procedure or formula for solving a problem, based on conducting a sequence of specified actions. A computer program can be viewed as an elaborate algorithm. In mathematics and computer science, an algorithm usually means a small procedure that solves a recurrent problem.

There is a one-to-one correspondence between maximizing probabilities and minimizing code lengths (Grünwald, 2007). Since in linear models we are interested in finding a model that, with high probability, can predict future observations of the data well, this is synonymous to finding a way to compress the data such that the total code length is short. To understand what a code length is, however, one first needs to understand what a code is. After all, it is difficult to measure the length of something unless one knows what to measure. Codes can be created in an infinite number of ways, and if we are to compare models based on code lengths instead of probabilities, it is important to ensure that the codes are all constructed in accordance with some kind of objective rule and not by subjective thought.

A code is a one-to-one matching of symbols to another set of symbols and where a symbol is a sign that holds information of some sort. The evolution of codes (uniform codes, see Appendix A) became necessary since researchers are looking for ways to make their algorithms more efficient. This leads to the more desirable prefix-free codes when given a sequence (or pattern in the data) and no code word is a prefix of another code word. This removes ambiguity in a given message or information in the dataset and provides us with the length of the code that is desirable (shortest). It then makes intuitive sense to select a short code word to symbols that have a high

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

probability of occurrence in the sequence. Kraft's inequality provides the means to find out if shorter codes can be achieved (Kraft, 1949; Rissanen, 2007) but it should be noted that it only measures the excessive length of the code word in the code. Kraft's inequality is silent on how well the code is expected to work. Thus, according to Fowler and Linblad (2011), Kraft's inequality provides us with a limit for the minimum amount of redundant bits in code words but Shannon's code (heavily draws from the principle of entropy) gives such a limit when it comes to expected code length (see Appendix C for more on Kraft's inequality and entropy). Although, the Shannon code fulfils the Kraft inequality (see Appendix D for proof), it can still in some cases lead to non-optimal codes (Shannon, 1948; Cover & Thomas, 2006). This implies there is more information in the data that could be used to compress the data further. Therefore, another more efficient and optimal way of choosing a prefix-free code given a relative frequency of symbol is by Huffman's algorithm (Roos, 2009a and Cover & Thomas, 2006 given in appendix B). Also, arithmetic coding is another common method used in both lossless and lossy data compression algorithms. It is an entropy encoding technique, in which the frequently seen symbols are encoded with fewer bits than rarely seen symbols. But Nelson (2014) states that, "Arithmetic encoders are better suited for adaptive models than Huffman coding, but they can be challenging to implement".

Linear models are interested in finding a model that with a high probability can predict future observations of the data well, and this is synonymous to finding a way to compress data such that the total code length

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

is short. This brings a relationship between MDL and the problem of model selection and data compression.

Building a Statistical Model

This section of the study, talks about how the MDL principle can be used in statistical modelling. It starts with a theoretical "ideal" solution and continues to a discussion about the two most common model selection criteria used today, namely: AIC and BIC, and finally ends up in the renormalized maximum likelihood criterion (NML- an extension of MDL).

Code processing

Since data compression, in MDL terms, is equivalent to probability maximization, compression is also highly relevant for statisticians. It would therefore be interesting, from a statistical point of view, to be able to find a way to compress a data sequence as much as possible.

The Kolmogorov complexity is defined as the length of the shortest program that, when run on a Turing machine (see Appendix E), takes the binary code, prints the uncompressed data and then stops (Grünwald, 2007). The idea behind this is that if the shortest program producing the data can be found, then the best way of compressing the data at hand has been found too. The problem with Kolmogorov Complexity is two-fold, namely,

1) Depends on the programming language chosen

2) Is non-computable.

The first of these two problems is usually trivial since the difference in length of code between programming languages is just a constant. However, for short sequences of data, this constant term might not be negligible. This is

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

due to the fact that each programming language that is Turing-complete can translate all other Turing-complete programming languages via a compiler. A language is said to be Turing-complete if it can simulate a Universal Turing machine (Burgin, 2007). In layman terms, if the programming language can do everything a Universal Turing machine can do, then it is a compiler (a complier is a program that can translate code from a programming language into another programming language; Bornat, 2008). Thus, the length of the shortest program for each language differs only at most by the length of the code needed to create the compiler (Solomonoff, 1964; Rissanen & Tabus, 2005).

The Kolmogorov complexity is non-computable hence it is a non-trivial issue; it can never be used directly to evaluate code lengths. Other measures, which can be thought of as approximations of the Kolmogorov complexity, are instead needed (Rissanen, 1983). One such solution is the so-called two-part code.

Rissanen (1978) published two-part code for model encryption which was also known as "Crude MDL" (Grünwald, 2007). The basic idea behind it is the following: given a dataset

$$\mathbf{D} = \{x_1, x_2, \ldots, x_n\},$$

it is desirable to compress the data as much as possible. Other authors, including Rissanen (2007) and Grünwald (2007), used the notation X^n instead of **D** for the data. Their approach avoids confusion with the so-called Kullback-Leibler divergence, that often is denoted D(P||Q), as well as makes it clear that the set is full of x's, but Roos (2009a, 2009b), use **D** instead, because we find it more intuitive as it avoids confusion with mathematical power

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

notation. The best way of compressing the data would be to use the Huffman code, described earlier, which is based on the relative frequencies in the data. The problem with this approach, however, is that the decoder needs to know beforehand which code has been used to compress the data. Without this code, it is impossible to decode the encoded message.

Rissanen's solution to this dilemma was to create a two-part code, where the first part of the code contains the code used and the second part is the message that has been compressed with the said code.

$$L(\mathbf{D},\lambda) := -\log_2 P(\mathbf{D}|\lambda) + L(\lambda)$$

where $L(\mathbf{D},\lambda)$ is the total length needed to describe the data, $P(\mathbf{D}|\lambda)$ is the likelihood of the data, given the model, and $L(\lambda)$ is the number of bits needed to define the code (Rissanen, 1983). The idea, in model selection, is to pick the model that minimizes $L(\mathbf{D},\lambda)$.

While this approach may be intuitively appealing, it can be shown that more advanced versions of MDL, such as normalized maximum likelihood, are strictly better; that is, they never achieve longer total code length than the two-part code and sometimes perform better (Rissanen & Tabus, 2005). In other words, better model selection criteria can be found.

Optimal Codes

By now it should be clear that not all ways of encoding symbols are equally optimal. It is not possible, however, in advance to predict exactly the future sequences of strings that are to be encoded and thus the code that have been created may not be, in hind-sight, the optimal one (Turing, 1937; Sipser, 2006; Grünwald, 2007). A process called Regret was seen to help achieve optimality.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

According to Grünwald (2007), given that $q(\mathbf{D})$ is in hind-sight, the optimal way of encoding a data sequence and that $\hat{P}_{ML}(\mathbf{D})$ is the method actually used to encode it. The Regret is defined to be

$$REG = -\log_2\left[q(\mathbf{D})\right] - \left(-\log_2\left[\hat{P}_{ML}(\mathbf{D})\right]\right)$$

In other words, the Regret can be thought of as the extra number of bits needed to encode a sequence if $\hat{P}_{ML}(\mathbf{D})$ is used instead of $q(\mathbf{D})$ (Grünwald, 2007). $\hat{P}_{ML}(\mathbf{D})$ is the maximum likelihood estimate of the probability of the data and $q(\mathbf{D})$ is, in hind-sight, the optimal way of encoding $\mathbf{D}$. Hence, it is a form of error measurement. The lower the Regret, the better the model is. Thus, a code with low Regret, relative to sample size, will yield an asymptotically optimal model. This brings us to the universal code.

A universal code (Grünwald, 2007), is a code for which the increase in Regret goes to infinity slower than the sample size does. Fowler and Lindblad (2011) explained that despite the name (universal) implying otherwise, the model is not universal in regards to all model classes, and it is only the one that it belongs to.

This implies that, in the long run, the model (a model and a code are in essence the same thing in MDL) that one is working with converges to the optimal model (Roos, 2009b). In other words, it is finding a model that fits the true model very well.

A code is said to be universal if

$$\lim_{n\to\infty}\frac{1}{n}\left(REG\right) = 0,$$

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Two-part codes are universal codes (Roos, 2009b; Grünwald, 2007), which might bring up the question why are criteria like Normalized Maximum Likelihood-NML (an extension is RNML, etc.) needed? Since in most practical applications one does not have an infinite amount of observations, so the rate at which the universal code approaches the limit is of high interest. The quicker it approaches the limit, the less dependent it is on having a large sample at hand.

The Normalized Maximum Likelihood Criterion

In order to select a model that can compress data as much as possible, we wish to find the model that minimizes

$$L(\mathbf{D}, \lambda) := -\log_2 P(\mathbf{D}|\lambda) + L(\lambda)$$

There are various ways to estimate $P(\mathbf{D}/\lambda)$, though it is, according to Grünwald (2007), preferable to estimate it by maximizing the following formula:

$$P_{NML} := \frac{\hat{f}_{ML}(\mathbf{D}|\lambda)}{\int_{\mathbf{D}\in\mathbf{\Omega}} \hat{f}_{ML}(\mathbf{D}|\lambda)d\mathbf{D}},$$

where $\mathbf{\Omega}$ is the population of possible values that can be observed. That is, the numerator of P_{NML} is the maximum likelihood estimate of the distribution, given the model. The denominator is the integral of the maximum likelihood estimates of the distribution over all possible datasets (Roos et al., 2005). The reason the maximum likelihood estimator is normalized is to make it a density function (Grünwald, 2007).

The NML criterion would thus favour the model minimizing

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

$$L(\mathbf{D}, \lambda) = -\log_2 \int_{\mathbf{D}\in\mathbf{\Omega}} \left[\hat{f}_{ML}(\mathbf{D}|\lambda) d\mathbf{D} \right] + \log_2 \left[\int_{\mathbf{D}\in\mathbf{\Omega}} \hat{f}_{ML}(\mathbf{D}|\lambda) d\mathbf{D} \right] + L(\lambda)$$

The first term of this criterion rewards models that describe the data well, while the second term penalizes complex models, similar to the criteria AIC and BIC. The last term is usually negligible (Grünwald, 2007). The problem with P_{NML} is that the integral over all possible datasets is almost always infinite, resulting in a measurement that cannot be calculated. To get around this problem, the integral needs to be restricted to an interval, that ensures that the integral is finite (Roos, 2004). This leads to the following approximate form of P_{NML} in linear regression modelling:

$$P_{NML} \approx \frac{\hat{f}_{ML}\left(\mathbf{D}; \gamma, \hat{\boldsymbol{\beta}}(\mathbf{D}), \hat{\sigma}^2(\mathbf{D})\right)}{\int_{\mathbf{\Omega}(\hat{\sigma}_o, \mathbf{R})} \hat{f}_{ML}\left(\mathbf{z}; \gamma, \hat{\boldsymbol{\beta}}(\mathbf{z}), \hat{\sigma}^2(\mathbf{z})\right) d\mathbf{z}},$$

where γ is a subset of all possible variables that could be included in the model (Rissanen & Tabus, 2005). $\mathbf{\Omega}\left(\sigma_o{}^2, R\right)$ means that we integrate over all possible data, $\mathbf{Z}$, such that the estimated residual variance ($\hat{R}$) is greater than or equal to some value, $\sigma_o{}^2$, so as to exclude saturated and overly overfitted models, and that

$$\hat{R} = \frac{1}{n} \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}}^T$$

is less than or equal to R . This is to make the number of possible data under consideration countable and thus limits the integral to an area that is not infinitely large.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

The integral can now be calculated, but a problem arises, namely: the resulting formula cannot be interpreted correctly from an MDL standpoint due to the fact that the last term depends on the data (Grünwald, 2007). The problem is solved by a second normalization. That is, after some rather tedious calculations was done to simplify the integration and the criterion was then called the Renormalized Maximum Likelihood criterion. The proof can be found in Roos et al. (2005).

$$RNML := -\ln \hat{f}\left(y;\gamma\right) =$$
$$\frac{n-k}{2}\ln\hat{\sigma}^2 + \frac{k}{2}\ln\hat{R} - \ln\Gamma\left(\frac{n-k}{2}\right) - \ln\Gamma\left(\frac{k}{2}\right) + \frac{n}{2}\ln\left(n\pi\right) + \ln\left[\ln\left(\frac{\sigma_2^2}{\sigma_1^2}\ln\frac{R_2^2}{R_1^2}\right)\right],$$

where k is the number of parameters in the model, n is the number of observations and is described as the estimated residual variance

$$\hat{R} = \frac{1}{n}\hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}}^T,$$

which in non-matrix notation can be written as

$$RNML = \hat{R} = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i^2,$$

the variation in $\hat{y}$ expressed by the model. Furthermore, since the variance cannot be estimated directly, due to yet another σ_1^2 and σ_2^2 integral over a range that is most often infinite, it is bounded within the interval, as is R between R_1^2 and R_2^2 (Roos, 2004). These values are arbitrarily chosen to be reasonable values that the data can obtain. That is, variance close to zero, or very high, is excluded. As can be shown, as long as the intervals are not too narrow, they do not greatly affect the criterion's approximation.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

By the use of the Stirling's approximation of the gamma functions (see Appendix F), as well as the fact that

$$\frac{n}{2}\ln(n\pi)+\ln\left[\ln\left(\frac{\sigma_2^2}{\sigma_1^2}\ln\frac{R_2^2}{R_1^2}\right)\right]$$

does not depend on *k* (and thus does not affect the value for which the function is minimized with regard to *k*), the formula can be simplified. The approximation has also been multiplied with two, since such a multiplication does not affect the value for which the function is minimized (Roos 2004). This gives the simplified form:

$$RNML \approx (n-k)\ln\hat{\sigma}^2+k\ln\hat{R}+(n-k-1)\ln\left(\frac{1}{n-k}\right)-(k-1)\ln k$$

from which the following alternative form follows by simple algebra (The derivation can be found in Appendix G):

$$RNML = (n-k)\ln\left(\frac{\hat{\sigma}^2}{n-k}\right)+k\ln\left(\frac{\hat{R}}{k}\right)+\ln\left[k(n-k)\right]$$

By minimizing the aforementioned function, the code length is minimized and thus the probability is maximized (though strictly speaking this is only an approximation), and thereby the best model is selected. It should be noted that from an MDL perspective, a good model does not necessarily need to be the same as the "true" model. Since this approximation of the *RNML* criterion is just an approximation, the correct notation for the criterion should thus be something in the line of $[\hat{R}NML]$. However, to simplify the notation the hat will be dropped throughout this thesis.

The RNML criterion is developed for linear regression analysis and is as such, not directly applicable to logistic cases. However, according to

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Arnoldsson (2011), the criterion could be adapted to the exponential family by adding the weight function, W, to R . Such an adaption might prove to perform well.

Sum of Squares Representation of Criteria

The basic problem of model selection (e.g. linear least squares regression) is how to choose between competing linear regression model. One battles with whether the model is too small ("underfit" the data; poor predictions; high bias and low variance) or the model being too big ("overfit" the data; poor predictions; low bias and high variance) or how one chooses a model to be just right (balance bias and variance to get good predictions). This study will apply the derivation of exact formulas for the different Minimum Description Length (MDL) criteria to Asymmetric Price Transmission (APT) linear models using the method of least squares.

The simplest linear regression model to be fit is of the form

$$\mathbf{y} = \mathbf{X}\mathbf{b}+\mathbf{e},$$

where **y** is a vector corresponding to the dependent variable, **X** is a matrix of regressors, **b** is the vector of regression coefficients and **e** is the vector of error terms.

The idea is to find a linear combination of regressor variables, which explains the systematic variation in the dependent variable. The ordinary least squares (OLS) technique is used to estimate the regression coefficients in that the error sum of squares is to be minimised. If normality with mean zero and constant variance is assumed for the error terms, the OLS estimates coincide with the maximum likelihood estimates. The main problem is to choose

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

appropriate regressors for the model. In practice, Sund (2001) argues that the domain knowledge is usually the best criteria to find interesting models, but sometimes a kind of "objective" assistance can be very helpful. As a matter of fact there are lots of different model selection criteria available.

This study gives a practical view of model selection based on MDL and the exact derivation for the special case of RNML, which forms the basis of all the MDLs presented, is shown in Appendix G. For generality and comparison purposes, some popular non-MDL-based model selection formulas are given in their least squares expressions. In the following formulas various authors have suggested other multipliers which keep the general form of these criteria. In other words, the numerical values obtained from the formulas are not directly comparable (the "scale" is not same for all formulas). However, in every case the smaller value means the better model.

In the following Equations (7, 8, 9, 10, 11, 12, 13), **y**, **X**, **b** and **e** are as in the regression model, *SSE* is the error sum of squares, $\hat{\sigma}^2$ is the variance estimated from fitting the full model, *k* is the number of parameters, *n* is the number of observations, $S = SSE / (n\text{-}k)$, $F = (\mathbf{y'y} - SSE) / kS$ and R^2 is the coefficient of determination.

Akaike's Information Criteria (AIC):

$$AIC = n\log\left(\frac{SSE}{n}\right) + 2k \qquad (7)$$

Bayesian Information Criteria (BIC):

$$BIC = n\log\left(\frac{SSE}{n}\right) + 2(k+2)\left(\frac{n}{\sigma^2}\right) - 2\left(\frac{n\sigma^2}{SSE}\right) \qquad (8)$$

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Rissanen's MDL (basic) = $rMDL = \frac{n}{2}\log_2\left(\frac{SSE}{n}\right)$ (9)

Schwarz's Bayesian Information Criteria SBIC (two-stage MDL devised by Rissanen):

$$SBIC = n\log\left(\frac{SSE}{n}\right) + k\log n \tag{10}$$

G-prior mixture MDL (gMDL):

$$gMDL = \begin{cases} \frac{n}{2}\log S + \frac{k}{2}\log F + \log n & , R^2 \geq k/n \\ \frac{n}{2}\log\left(\frac{\mathbf{y}'\mathbf{y}}{n}\right) + \log n & , \text{otherwise} \end{cases} \tag{11}$$

Normalised Maximum Likelihood MDL (nMDL):

$$nMDL = \frac{n}{2}\log S + \frac{k}{2}\log F + \frac{1}{2}\log(n-k) - \frac{3}{2}\log k \tag{12}$$

Exact Normalised Maximum Likelihood MDL (eMDL):

$$eMDL = (n-k)\log\frac{SSE}{n} + k\log\left[\mathbf{b}^1\mathbf{x}^1\mathbf{xb}\right] + (n-k-1)\log\left(\frac{n}{n-k}\right) - (k+1)\log k \tag{13}$$

In the following predictive Equations (14, 15), y_i and X_i correspond to the values of the response and regressor variables on the i^{th} observation, b_{i-1}

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

and $\hat{\sigma}^2{}_{i-1}$ are the estimates based on $(i-1)$ first observations, m is the first integer so that b_i is uniquely defined and n is the number of observations.

Predictive MDL (PMDL):

$$PMDL = \sum_{i=m+1}^{n} \left[\log \hat{\theta}^2_{i-1} + \frac{(y_i - x_i b_{i-1})^2}{\hat{\theta}^2_{i-1}} \right] \tag{14}$$

Predictive Least Squares (PLS):

$$PLS = \sum_{i=m+1}^{n} (y_i - X_i b_{i-1})^2 \tag{15}$$

In summary, research emphasizes the fact that, understanding the data means the ability to remove redundancies in the data. Hence to discover regular statistical features, the ultimate measure of the success of understanding must be the length with which the data can be described (principle of Parsimony). Indeed, if such a shortest description of the data, to be called stochastic complexity, is found in terms of the models of a selected class, there is nothing further anyone can teach us about the data; we know all there is to know (Sund, 2001). This is the rationale behind the MDL (minimum description length) principle.

The minimum description length principle epitomizes a significantly different basis for model selection and generally statistical inference. One of such distinctive features is that there is no need to assume anything about the data generation mechanism. Sund (2001) argues that in particular, unlike in traditional statistics, it is not needed that the data form a sample from a

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

population with some probability law. Hence, the objective is not to estimate an assumed but "unknown" distribution, be it inside or outside the proposed class of models, but to find good models for the data. Most essentially, the principle permits comparison of any two models, regardless of their type. It is also important to realise that the MDL principle has nothing to say about how to select the suggested family of model classes. In fact, this is a problem that cannot be adequately formalised. In practise the selection of models is based on human judgement and prior knowledge of the kinds of models that have been used in the past, perhaps by other researches. In addition, the application of the principle requires the calculation of the stochastic complexity, which can sometimes be a difficult task.

The MDL has its roots in information theory and in the invariance theorem of Kolmogorov Complexity. The Kolmogorov Complexity of a sequence is defined to be the length of the shortest computer program that prints the sequence and then halts (halting makes the code a prefix code, i.e. none of the code words is a prefix of another). Unfortunately, the Kolmogorov Complexity is not computable. The idea behind the MDL is to scale things down in a way that it becomes possible to compute the complexity: instead of using a code based on a universal computer language, we should use an arbitrary class of models and do the encoding with the help of this model class. However, there are different forms of description length based on a model, even in that sense that they achieve the universal coding lower bounds. The most straightforward description length is the one based on two-stage coding scheme. The idea in the first stage is to calculate the code length required to discretize model's parameter space and communicate the estimated

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

parameter. In the second stage the actual data string is coded using the distribution indexed by the communicated parameter. In more complicated situations more than two stages of coding might be required.

In the mixture form of description length we base our description of a data string on a distribution that is obtained by taking a mixture of the members in the family with respect to a probability density function on the parameters. The mixture description length results in integral formula, which has closed form expression only in special cases. An analytical approximation to the mixture can be in certain situations obtained by Laplace's expansion and essentially results in a two-stage description length which is called the stochastic information complexity.

Hansen and Yu (2001) explained that the recent form of description length bases on the normalised maximum likelihood (NML) coding scheme. In general, the NML description of a data string works by restricting the second stage of coding to a data region identified by the parameter estimate. This criterion is not only sensitive to functional form and the number of parameters but also invariant under re-parameterisation. From the differential geometric point of view, these criteria select the model that gives the highest value of the maximised likelihood per distinguishable distribution, which may be called the "normalised maximised likelihood". In other words, the model complexity is related to the number of (distinguishable) probability distributions that a model can generate, not to the functional form of a model or its number of parameters.

Moderately, a method is to consider the description length from a predictive coding point of view. Predictive coding according to Sund (2001)

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

means that modelling the conditional density for the possible values of the "next" observation using the part of the data which is already "seen". This means the joint distribution of a data string can be written as a product of conditional distributions. He further explains that if each of the conditionals shares the same parameter, then the joint distribution based on the particular model class is free of unknown parameters and the cost of encoding a data string can be directly seen from the joint distribution. This method of description length is called predictive description length and it is used especially in situations where the data is sensibly ordered.

Overview of Asymmetric Price Transmission Linear Models

This section delves into literature on asymmetric price transmission and theories. Empirical literature is discussed to demonstrate how the different econometric models which measure these asymmetrical behaviours detect asymmetries at different rates or culminates in different inferences and conclusions. Subsequently, the MDL principle and its various extensions discussed above which provides a better and simple (though from a different background than the popular AIC and BIC) framework for comparing competing models is proposed to guide the rigorous comparison of the alternative methods of testing for asymmetry.

Asymmetric price transmission (sometimes abbreviated as APT and informally called "rockets and feathers") refers to pricing phenomenon occurring when downstream prices react in a different manner to upstream price changes, depending on the characteristics of upstream prices or changes in those prices.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

The simplest example is when prices of ready products increase promptly whenever prices of inputs increase, but take time to decrease after input price decreases.

Over the years, the analysis of price transmission and asymmetric adjustment have matured with many developments in model specification, estimation and testing. Some developments are the construction and application of various econometric models of asymmetric price transmission.

Asymmetric Price Transmission

The relationship between two prices (P_A and P_B) at different levels of the marketing chain can be estimated as follows:

$$P_{A,t} = \beta_o + \beta_1 P_{B,t} + \varepsilon_t \tag{16}$$

where β_1, β_2 are regression coefficients; and ε the random error for a give price series with time t.

Recent empirical studies analysing whether prices rise faster than they fall, have categorised the price dynamics into symmetric and asymmetric processes and denoted as

$$P_{A,t} = \beta_o + \beta_1^+ P_{B,t}^+ + \beta_1^- P_{B,t}^- + \varepsilon_t \tag{17}$$

Thus, inference on price whether symmetric or asymmetric is achieved through hypothesis testing using an F-test as follows:

Test for Symmetric Adjustment

$$\begin{aligned} H_o &: \beta_1^+ = \beta_1^- \\ H_A &: \beta_1^+ \neq \beta_1^- \end{aligned} \tag{18}$$

According to Gauthier and Zapata (2001), asymmetry is defined as an un-reciprocal relationship between rises and falls in prices, and an example is

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

farm and retail prices. Researchers are keenly interested in those processes for which the transmission differs according to whether the prices are increasing or decreasing (i.e. asymmetric price transmission). Several empirical researches show that the price transmissions are asymmetric and studies of various products and services (including gasoline, agriculture products and bank deposit rates) reveal that prices are more likely to rise to input price increases than they are to decrease in the wake of cost reduction. Peltzman (2000) significantly broadens the evidence for this asymmetrical price behaviour. In a study of 77 consumer and 165 producer goods, Peltzman found that on the average, the immediate response to a cost increase is at least twice the response to a cost decrease. This phenomenon presents an interesting empirical constancy that needs to be explained.

The issue of APT continues to receive significant attention in the economic literature for two prominent reasons. First, its presence is not in line with predictions of the conventional economic theory (e.g. perfect competition and monopoly) which postulates that under some regularity assumption (such as non-kinked convex or concave demand functions) prices should respond symmetrically to cost increases and cost reductions. The forgoing discussion is consistent with Peltzman (2000), who finds asymmetric price transmission to be the rule, rather than the exception and argues that it poses a real challenge to standard economic theory, since it does not predict or explain the existence of asymmetries. Hence, APT reveals gaps in economic theory.

Secondly, APT also presents important welfare and policy implications (Von Cramon-Taubadel & Meyer, 2000). 'It implies a different distribution of

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

welfare than would be obtained under symmetry, since it alters the timing and size of welfare changes'.

Von Cramon-Taubadel and Meyer (2000) added that the presence of asymmetric price transmission was often considered to be evidence of market failure (for example exercise of market power), 'signalling in addition to redistribution, the associated net welfare losses'. This means the redistribution and net welfare loss provide 'a *prima facie'* case for policy intervention.

Von Cramon-Taubadel (1998) and Meyer (2003) provided a concise discussion of the definition of asymmetry in the context of price transmission under three main groupings namely: a) asymmetry with reference to the speed and magnitude, b) asymmetry affecting vertical or spatial price transmission and, c) positive or negative asymmetry.

Particularly, positive asymmetry defines a set of reactions in which any price movement that squeezes the margins is transmitted more rapidly than an equivalent that stretches the margin. On the other hand, asymmetric price transmission is negative, when any price movement that stretches the margin is transmitted more rapidly than those that squeeze the margin. Figure 1 shows asymmetry with respect to the speed and magnitude with price P on the vertical axis and time t on the horizontal axis.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

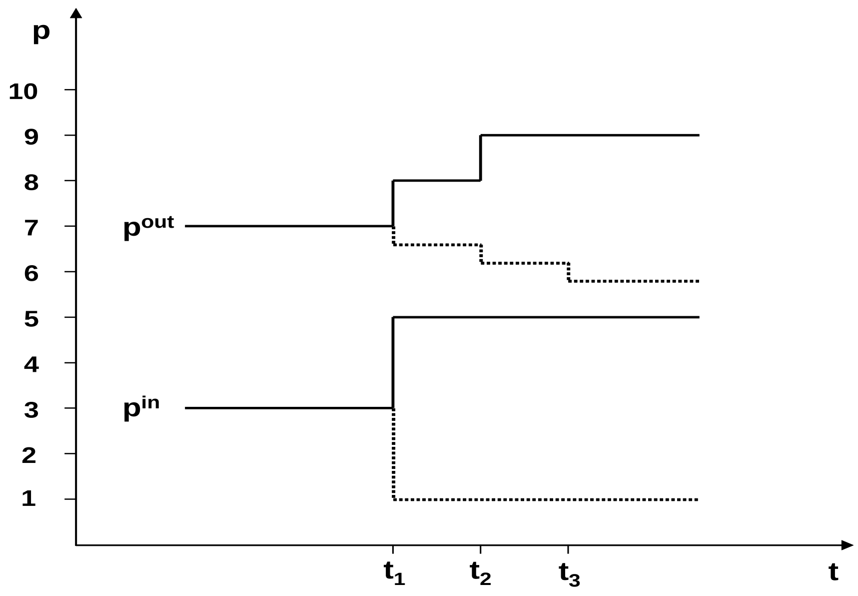


Figure 1: Example of type of Asymmetry (speed and magnitude)

It is evident from Figure 1 that a decrease in the input price $\left(P^{in}\right)$ takes three periods and is not fully transmitted to the output price $\left(P^{out}\right)$. While an increase in the input price $\left(P^{in}\right)$ takes only two periods and is fully transmitted to the output price $\left(P^{out}\right)$.

Problem of Asymmetric Price Transmission

Numerous factors which contribute to asymmetric price transmission have been projected in research. Firstly, a frequently cited source of APT is market power (Kinnucan & Forker 1987; Miller & Hayenga, 2001; McCorrisston, 2002; and Lloyd et al., 2003). “Oligopolistic processors, for example, might react collusively more quickly to shocks that squeeze their margin than to shocks that stretch it, resorting in asymmetric short run transmission in an attempt to hide the exercise of market power behind the ‘confusion’ created by major shocks, processors could also react less completely to the shocks that stretch their margins leading to asymmetric long run transmission”.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

In the same way, asymmetric price transmission could result if traders in the local market believing that competitors will follow an increase in the local market prices as price in the central market rise, but that they will not respond to falling prices in the central market by granting an equivalent reduction. It is however important to mention that concentration of price is probably a necessary but certainly not a sufficient condition for the exercise of market power, as the theoretical and empirical evidence on the relationship between these two phenomena is inconclusive (Goodwin and Serra, 2003). Arguing from the oligopoly perspective, both positive and negative asymmetries are conceivable depending on the market structure and conduct. Thus, several studies of market power and asymmetry that focus on specific markets deserve to be mentioned. In support of the aforementioned line of thought, Borenstein et al. (1997) analysed vertical price transmission for crude oil to gasoline prices, and concluded that downward stickiness of retail prices for gasoline in an oligopolistic environment will lead to positive asymmetry.

Then again, Ward (1982) points out that market power can lead to negative asymmetry if oligopolists are reluctant to risk losing market share by increasing output prices. A firm facing a kinked demand curve that is either convex or concave to the origin was also considered by Bailey and Brorsen (1989). They argued that if a firm believes that no competitor will match a price increase but rather match a price cut (concave), negative asymmetry will result. On the other hand, if the firm assumes that all firms will match an increase but none will match a price slash (convex), positive asymmetry will result. Hence, it is not clear a priori whether market power will lead to positive or negative asymmetry (Bailey and Brorsen, 1989).

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Secondly, price asymmetry can be to a certain extent credited to adjustment cost that arises when firms change their quantities and prices of inputs and outputs. Accordingly, positive or negative asymmetric price transmission results if these costs are symmetric with respect to increase or decrease in quantities or prices. In an analysis of the US beef market, Bailey and Brorsen (1989) argued that firms may face different adjustment cost depending on whether prices are rising or falling. Consequently, they noted that the competition between meat packers faced with a high fixed cost and excess capacity, for example, might result in farm prices that are bid up rapidly, in response to increased demand for meat products, but fall more slowly as demand weakens. Ward (1982) thought that retailers of perishable products may be cautious to raise prices for fear that they could end up holding spoiled stocks, leading to negative asymmetry. Heien (1980) disputed this assertion and noted that changing prices is less of a problem for perishable products than it is for those with a long shelf life, because for the latter, changing prices incur higher time cost and loss of good will. Thus, echoing the so called menu cost hypothesis proposed by Barro (1972), (i.e. a change in nominal price induces cost for example, the reprinting of price list or catalogues and the cost of informing market partners). Later in 1994, Ball and Mankim developed a model based on menu cost (the cost involved in changing nominal prices such as the cost of reprinting catalogues etc.) in combination with inflation that leads to asymmetry. In their model, positive nominal input price shocks are more likely to lead to output price adjustment than negative price shocks. This they argued was due to the fact that in the presence of inflation, some of the adjustment made necessary by an input price

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

reduction is automatically carried out by inflation, which reduces the real value of the margin. Thus, in situations where firms face menu cost and inflation, shocks that bring upward price adjustment are rapidly responded to than those that reduce it, as inflation in this respect would have automatically affected some of the adjustments made necessary by the downward adjustment shocks (Kuran, 1983). Alternative to Bailey and Brorsen (1989) idea, Peltzman (2000) makes a case for positive asymmetry affirming that it is easier for firms to disemploy inputs in the case of an output reduction than it is to recruit new inputs to increase output. This recruitment will lead to search cost and price premier increasing phases. They added that additional explanation for price asymmetry which has been proposed cannot be considered directly under market power or adjustment cost.

Kinnucan and Forker (1987) suggested that asymmetry could result from government intervention. This indicated that such political intervention can lead to asymmetric price transmission if it makes wholesalers or retailers to believe that a reduction in farm prices will only be temporary because it will only trigger government intervention, while an increase in farm prices is more likely to be permanent.

Modelling Asymmetric Price Transmission

Asymmetric price transmission modelling (See Von Cramon-Taubadel (1998) for detailed discussion) may be categorized into pre-cointegration and cointegration approaches (Meyer & Von Cramon, 2004). The pre-cointegration and the cointegration approaches draw heavily from Houck (1977) and Von Cramon-Taubadel (1998), respectively.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Deductions from Houck's (1977) pre-cointegration approaches lead numerous authors to develop a test for asymmetric price transmission that is based on the segmentation of prices into increasing and decreasing phases (Kinnucan & Forker, 1987; Bailey & Brorsen, 1989; Zhang, Fletcher & Carley, 1995; Mohanty, Peterson & Kruse, 1995; Boyd & Brorsen, 1998; Shin et al., 2014). These applications were considered as modifications of the Houck's model and denoted by Houck's approaches. The pre-cointegration methods require data to be stationary in order to avoid spurious regression. Thus, cointegration methods are fortified due to the fact that the Houck's approaches are not consistent with cointegration between the prices series involved. Fervidly, this sparked the motivation for the asymmetric error correction modelling (Von Cramon-Taubadel, 1998; Von Cramon-Taubadel & Loy, 1999).

Variants of Houck's Specification

A wide variety of agricultural markets have tested Asymmetric Price Transmission and a study conducted by Appel (1992) found that both speed and degree of price transmission from the producer to the retail level for broilers in Germany is asymmetric. However, Boyd and Brorsen (1998) also studied the US pork market and found no evidence of asymmetric price transmission. Then the result was challenged by Hahn (1990) who found that prices at all levels of the US pork and beef marketing chains are rather more sensitive to price increasing shocks than to price decreasing shocks. Subsequently, Hansmire and Willett's (1992) indicated that farm-retail price transmission for New York state apples is asymmetric and Kinnucan and Forker (1987) came to the same conclusion regarding dairy product

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

transmission in the United States. Pick et al. (1990) finds evidence that short-run but not long-run vertical price transmission on the US citrus market is asymmetric. Again, Ward (1982) found both short and long run asymmetries in vertical price transmission for fresh vegetable in the United States, while Zhang et al. (1995) noted that price transmission for peanut to peanut butter prices in the US is asymmetric in the short-run, but symmetric in the long-run.

In general, each of these studies used some different econometric technique for estimating irreversibility that was introduced by Wolfram (1971) which was a response to work on irreversible supply reaction by Tweeten and Quance (1969). Investigating the relationship between an output price P_A and input price P_B, Tweeten and Quance (1969) used an indicator variable to split the input price into two parts: one variable includes only increasing input prices P_B^+ and another includes only decreasing input prices P_B^-. From this, two input price adjustments coefficients (i.e. β_1^+ and β_1^-) can be estimated as

$$P_{A,t} = \beta_o + \beta_1^+ P_{B,t}^+ + \beta_1^- P_{B,t}^- + \varepsilon_t \tag{19}$$

Symmetric price transmission is rejected if the coefficients β_1^+ and β_1^- are significantly different from one another. Based on Tweeten and Quance's (1969) work, Wolffram (1971) proposed a variable splitting technique that unambiguously included first difference of prices in the equation to be estimated which was later modified by Houck (1977). Deductions from Wolfram-Houck's (W-H) method revealed that the response of price P_A to another price P_B is estimated with the

$$\Delta P_{A,t} = \beta_o + \beta_1^+ \Delta P_{B,t}^+ + \beta_1^- \Delta P_{B,t}^- + \varepsilon_t \tag{20}$$

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

where ΔP^{+} and ΔP^{-} are the positive and negative changes in P_B, respectively, $\beta_o, \beta_1^{+}, \beta_1^{-}$ are coefficients and t is the current period. Several studies estimated a dynamic variant of the Houck's static model. Again, some analysts distinguish between short-run and long-run asymmetries by introducing lagged terms in $\Delta P_{B,t}^{+}$ and, $\Delta P_{B,t}^{-}$ into Equation (20), in which case β^{+} and β^{-} become lag polynomials. Long-run symmetry was tested by determining whether the sums of the coefficients in these polynomials are identical. Ward (1982) extended the Houck's specification by including lags and Brorsen's group was also the first to use lags to differentiate between magnitude and speed of transmission. Similarly, Hahn (1990) attempted to generalize the methods discussed so far by referring to them as the pre-cointegration methods.

Further modification by Mohanty, Peterson and Kruse (1995) improved Equation (20) by taking the sum of both sides to derive the equation

$$\sum_{t+1}^{T} \Delta P_{A,t} = \alpha_o + \alpha_1 \sum_{t+1}^{T} \Delta P_{B,t}^{+} + \alpha_2 \sum_{t+1}^{T} \Delta P_{B,t}^{-} + \varepsilon_t \quad (21)$$

which can be rearranged as follows:

$$P_{A,t} - P_{A,0} = \alpha_o + \alpha_1 P_{B,t}^{UP} + \alpha_2 P_{B,t}^{DOWN} + \varepsilon_t \quad (22)$$

where P_B^{UP} is the sum of all positive changes in price B and P_B^{DOWN} is the sum of all negative changes in price B. Thus, a formal test for symmetry using an F-test or t-statistic is rejected when the coefficients α_1 and α_2 are unequal.

The Houck's model is sometimes used without sufficient regards to time series properties of the data but Von Cramon-Taubadel (1998) had demonstrated that the model is fundamentally incompatible with cointegration

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

between two price series. In order to demonstrate this point, Von Cramon-Taubadel (1998) considered two I (1) processes, P_A and P_B, in the model below as previously defined in Equations (19) and (20)

$$\sum \Delta P_{A,t} = \beta_o + \beta_1 \sum \Delta P^+_{B,t} + \beta_2 \sum \Delta P^-_{B,t} + \varepsilon_t \tag{23}$$

which can be reparametrized using the identity:

$$\sum \Delta P^+_{B,t} + \sum \Delta P^-_{B,t} \equiv P_{B,t} - P_{B,0} \tag{24}$$

to yield

$$P_{A,t} - P_{A,0} = \beta_o + \beta_1^+ \sum \Delta P^+_{B,t} + \beta_1^- \left(P_{B,t} - P_{B,0} - \sum \Delta P^+_{B,t} \right) + \varepsilon_t \tag{25}$$

or

$$P_{A,t} = \left(P_{A,0} + \beta_o + \beta_1^- P_{B,0} \right) + \beta_1^- P_{B,t} + \left(\beta_1^+ - \beta_1^- \right) \sum \Delta P^+_{B,t} + \varepsilon_t \tag{26}$$

This reparameterization of Equation (23) was proposed by Ward (1982) who tested whether the coefficient $\left(\beta_1^+ - \beta_1^-\right)$ differs from 0 in order to test whether price transmission was asymmetric. Von Cramon-Taubadel (1998) asserts that the estimation of Equation (26) can lead to four basic results depending on the significance of the term $\left(\beta_1^+ - \beta_1^-\right)$ and the stationarity of the error term ε_t:

Case 1: $\beta_1^+ - \beta_1^- \neq 0$ (asymmetry) and ε_t is I(0)

Case 2: $\beta_1^+ - \beta_1^- = 0$ (symmetry) and ε_t is I(1)

Case 3: $\beta_1^+ - \beta_1^- \neq 0$ (asymmetry) and ε_t is I(1)

Case 4: $\beta_1^+ - \beta_1^- = 0$ (symmetry) and ε_t is I(0)

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Case 1 implies that P_A, P_B, and $\sum \Delta P_{B,t}^{+}$ are cointegrated, which precludes cointegration between P_A and P_B alone. Case 2 and 3 are spurious regressions (Granger & Newbold, 1974), while case 4 implies that P_A and P_B are cointegrated. Notably, if Houck's method points to asymmetry, then either the results reflect spurious regression (Case 3), or the prices in question are not cointegrated (Case 1).

Presentation of the Asymmetric Error Correction Model

Primarily, the asymmetric error correction model (ECM) approach was motivated by the fact that all the variants of the aforementioned Houck's approach are not consistent with cointegration between the price series. If the prices P_A and P_B are cointegrated, then an error correction representation exists (Engle & Granger, 1987). Granger and Lee (1989) proposed a modification to reflect the error correction representation that makes it possible to test for asymmetric price transmission between cointegrated variables. This can be achieved by the segmentation of the error correction term into positive and negative components (Wolffram, 2005).

Von Cramon-Taubadel and Fahlbusch (1994) made the first attempt to draw on cointegration technique in testing for asymmetry in vertical price transmission and later improved by Von Cramon-Taubadel and Loy (1996) and Von Cramon-Taubadel (1998). They suggested that in the case of cointegration between the price series, an error correction model extended by the incorporation of asymmetric adjustment terms provides a more appropriate specification for testing for asymmetric price transmission. In effect, when Equation (27) was estimated, the test proved to be non-spurious regression and

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

then P_A and P_B was referred to as being cointegrated, and Equation (27) was now considered as an estimate of the long-run relationship between the prices.

$$P_{A\ ,\ t} = \beta_o + \beta_1 P_{B,\ t} + \varepsilon_t \tag{27}$$

The Error Correction Model (ECM) then relates changes in P_A to changes in P_B as well as the so called Error Correction Term (ECT); the lagged residuals derived from estimation of Equation (27). The ECT measures the deviation from the long-run equilibrium between the P_A and P_B, and including it in the ECM allows not only P_A to respond to changes in P_B but also to correct any deviations from the long-run equilibrium that may be left over from previous periods. Splitting the ECT into positive and negative component (i.e. positive and negative deviation from the long-run equilibrium; ECT^+ and ECT^-) makes it possible to test for asymmetric price transmission as follows:

$$\Delta P_{A\ ,\ t} = \beta_o + \beta_1 \Delta P_{B\ ,\ t} + \beta_2^+ ECT_{t-1}^+ + \beta_2^- ECT_{t-1}^- + \beta_3 \Delta P_{B\ ,\ t-1} + \beta_4 \Delta P_{A\ ,\ t-1} + \varepsilon \tag{28}$$

Von Cramon-Taubadel and Loy (1996) also segmented the concurrent response term in Equation (27). This led to the following specification in which concurrent and short–run responses to departures from the cointegrating relation are asymmetric if $\beta_1^+ \neq \beta_1^-$ and $\beta_2^+ \neq \beta_2^-$ respectively, that is:

$$\Delta P_{A\ ,\ t} = \beta_o + \beta_1^+ \Delta P_{B,t}^+ + \beta_1^- \Delta P_{B,t}^- + \beta_2^+ ECT_{t-1}^+ + \beta_2^- ECT_{t-1}^- + \beta_3 \Delta P_{B,\ t-1} + \beta_4 \Delta P_{A,\ t-1} + \varepsilon \tag{29}$$

Conspicuously, Equation (29) is equivalent to the Houck approach given by Equation (20), except that Equation (29) also contains $\beta_2^+ ECT_{t-1}^+ + \beta_2^- ECT_{t-1}^- + \beta_3 \Delta P_{B\ ,\ t-1} + \beta_4 \Delta P_{A\ ,\ t-1}$. In effect, the asymmetric

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

ECM with complex dynamics nests the Houck's model in first difference or has the structures of the Houck's model.

Submissions so far have emphasised the use of the Granger and Lee asymmetric error correction model by Von Cramon-Taubadel (1998). Subsequently, other researchers (Acquah & Von Cramon, 2009; Acquah, 2010; etc.) also implemented it. Specifically some studies employed variants of these approaches. For instance a recent approach proposed by Chavas and Mehta (2004) appears to nest a variety of earlier approaches.

Many price transmission studies implemented Von Cramon-Taubadel and Loy (1996) testing procedure for asymmetric price transmission or some variants of their proposed ECM approach. For example, Von Cramon-Taubadel and Loy (1996) used an ECM to study the spatial price transmission on world wheat markets. Similarly, Capps and Sherwell (2007) analysed the behaviour of spatial test of asymmetric price transmission according to the Von Cramon-Taubadel and Loy ECM's approach and the conventional Houck's approach. Scholnick (1996) also used an asymmetric ECM to test for asymmetric adjustment of interest rates, while Borenstein et al. (1997) employed an ECM specification where the error correction terms are not segmented. Von Cramon-Taubadel (1998) demonstrates that transmission between producer and wholesale pork prices in northern Germany is asymmetric. Balke et al. (1998) and Frost and Bowden (1999) also employ variants of the asymmetric ECM. FAO (2003) provided a review of the application of time series techniques (cointegration, ECM) in testing market integration and price transmission for a number of cash and food crop markets in developing countries.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Threshold Models

Other researchers continue to work on how to improve the already existing approaches. Drawing from the threshold approach introduced by Tong (1983), it was possible to consider an intuitively appealing type of ECM in which deviation from the long-run equilibrium between P_A and P_B will lead to a price response if they exceed a specific threshold level. Many studies measuring asymmetric price transmission using the threshold approach estimates variants of the following simplified equation.

Standard Threshold Cointegrated Model is given by

$$\Delta P_{A\ ,\ t} = \beta_o + \beta_1 \Delta P_{B,\ t} + \beta_2^+ ECT_{t-1}^+ + \beta_2^- ECT_{t-1}^- + \varepsilon_t \qquad (30)$$

Given a threshold (γ), where

$$ECT_{t-1}^+ > \gamma \text{ and } ECT_{t-1}^- \leq \gamma \qquad (31)$$

The Error Correction Term (ECT) is segmented into ECT_{t-1}^+ and ECT_{t-1}^- according to whether it is greater or less than a defined threshold value, respectively. Detailed discussion on the threshold modelling is provided in numerous studies including Enders (2004), Balke and Fomby (1997) and Tsay (1989). Balke and Fomby (1997) presented a model that allows for non-linear adjustment to equilibrium by introducing the concept of threshold cointegration. The relationship between symmetry and threshold is systematically developed in Meyer and Von Cramon-Taubadel (2004). The authors noted that thresholds allow for different types of asymmetry.

The first type refers to a two symmetric threshold model (i.e. three regime model) with asymmetric responses in the outside regimes, reflecting asymmetry with respect to speed of transmission. In contrast, a two symmetric

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

threshold model with three regimes and symmetric responses in the outside regimes need not be asymmetric. On the other hand, a one threshold model is asymmetric if the threshold parameter (γ) differs from zero.

The second type of asymmetry refers to the fact that the two thresholds need not be equal. If this type of asymmetry holds then the deviations in the positive and negative directions must reach a different magnitude before a price response is triggered.

Following the development of the threshold model, a number of applications have estimated asymmetric adjustments using threshold error correction models. Abudulai (2002) draws on Enders and Granger (1998) to test for asymmetric price transmission in a methodology in which the threshold parameter (γ) is set to zero. Alternatively, Goodwin and Harper (2000) and Goodwin and Piggott (2001) use a grid search to find optimal thresholds in price transmission analysis. Hansen and Seo (2002) developed a test for the significance of a single threshold in an error correction model where the ECT is segmented not according to whether it is greater or less than zero but rather according to whether it is greater or less than a threshold value that might differ from zero. In an empirical application, Aguero (2004) uses a threshold error correction model to estimate price adjustments under risk in the Peruvian agricultural markets.

Review of Empirical Applications

As discussed by Acquah (2010), the review of empirical application is based on thorough literature survey of asymmetric price transmission which draws heavily from Meyer and Von Cramon-Taubadel (2004). The survey

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

documentation by Meyer and Von Cramon-Taubadel (2004) consists of 40 publications estimating asymmetric price transmission in samples of different products including fuel and gasoline, interest rates, and agricultural products.

Most of these applications are based on monthly and weekly price data while daily, fortnightly and quarterly data are each used once. Nearly half of the test for asymmetric price transmission employs some type of pre-cointegration procedure (i.e. 19 out of 40). Post cointegration methods such as the error correction mechanisms and the threshold approaches are employed in 11 publications (i.e. 4 ECM and 7 threshold applications). The remaining studies implemented a variety of other approaches. Table1 presents the outcome of a Meta-analysis based on the results of all published individual tests derived by Meyer and Von Cramon-Taubadel (2004).

Table 1: Asymmetry Test of the Different Econometric Models

	Test Methods					
	All methods	Methods using first difference	Methods using summed difference	ECM methods	Threshold methods	Misc methods
Total cases	205	93	53	31	10	18
Symmetry maintained	106	30	40	17	2	17
Symmetry rejected	99	63	13	14	8	1
Symmetry rejected (%)	48	68	25	45	80	6

Source: Meyer and Von Cramon-Taubadel (2004, pp. 22)

Inferring from the reviewed literature (Meyer & Von Cramon-Taubadel 2004), it is increasingly evident that the different methods appear to lead to different rates of rejection of the null hypothesis of symmetry. The fact that the literature to date contains few rigorous comparison and analysis of the

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

strengths and weaknesses of the available methods is a hurdle to the advancement of asymmetric price transmission analysis. If researchers knew which method is appropriate, they could focus on this and gather insight and knowledge more rapidly in the future.

On the other hand, Meyer and Von Cramon-Taubadel (2004) noted that the available methods are not all simply reparameterization of one another and that they cannot all be equally appropriate in all cases. Emphatically, the survey by Meyer and Von Cramon-Taubadel (2004) showed that the different methods detect asymmetry at different rates (but this could be because different methods applied to different data in each case). Subsequently, a simple experiment with simulated data implemented in research conducted by Acquah and Von Cramon (2009) produced a similar result consistent with the findings of the survey by Meyer and Von Cramon-Taubadel (2004) that the different methods detect asymmetry at different rates.

Research and the discussions so far have established that a different econometric model of asymmetric price transmission detect asymmetry at different rates or culminates in differences in inferences and conclusion. It therefore remains indispensable to select one model from the set of competing models that best captures the true underlying asymmetric data generating process.

Chapter Summary

The methodology examines various techniques used for model selection. These include AIC, BIC and MDL. Various specifications of the MDL have been reviewed. It explains the notion of data compression. It has

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

explained the code compression presented by various authors. In line with this, it has explained the concepts of optimal codes, Normalized Maximum Likelihoods criterion and Renormalized Maximum Likelihood criterion.

Various measures of information criteria have been stated. These are AIC, BIC, rMDL, gMDL, SBIC, nMDL and eMDL. The concept of Asymmetric Price Transmission (APT) has been discussed quite extensively and illustrated. It points out the factors to APT. Various models for APT have also been reviewed that suggest the manner for testing for price asymmetry. It then presents Asymmetric Error Correction models and threshold models.

The review asserts that for rigorous comparisons and analysis of the strength and weaknesses of the methods pose a challenge of determining the most appropriate method that has optimal rates of correct identifications of asymmetry.

In the light of selecting one model from the set of econometric asymmetric price transmission models, this research however, introduces yet another criteria; the Minimum Description Length Principle (MDL) discussed earlier, which will help capture the true data generating process among others. The MDL principle represents a drastically different foundation for model selection and, in fact, statistical inference in general. It has a number of distinctive features: There is no need to assume anything about the data generation mechanism and in particular, unlike traditional criterion (AIC, BIC, etc.), it is not needed that the data form a sample from a population with some known probability law. Hence, the objective is not to estimate an assumed and "unknown" distribution, whether inside or outside the proposed class of models, but to find good models for the data. Most essentially, the principle

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

permits comparison of any two models, regardless of their type (Sund, 2001 Lecture series).

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

CHAPTER FOUR

RESULTS AND DISCUSSIONS

Introduction

In this chapter, the R-functions for the various formulations of MDL will be developed since no such functions have been built in available software yet. This will enable us apply them to the various formulations of APT models which have been in use in econometric modelling over the years. Further, artificial data will be used to simulate the behaviour of the models to find out their long term behaviour in determining the true data generating process of our APT econometric models. This will help to determine the best MDL criteria in selecting the best APT linear model even in several iterations and under various conditions.

Overview on Transforming MDL Equations into R Functions

This study has established four formulations of the MDL in our theoretical review in Chapter Three. These have to be converted into R-functions (since R is the statistical programming language employed throughout this study and no such functions exist yet in R) from scratch using the AIC criterion already developed in R. The reason is, it was established from our previous chapter that, most information-theoretic fit criteria can be considered as special cases of what Barron and Cover (1991) referred to as minimum complexity density estimators; $i\left((X|m(\theta)\right)+i\left(m(\theta)\right)$. It was explained that the first term is equivalent to the negative log-likelihood of the data calculated at the maximum likelihood estimates of the parameter and the second term is the penalty for model complexity and this second part differs

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

for different information-theoretic fit criteria and uniquely defines a give criterion. Therefore, one can use, say, AIC to find maximum log likelihood and then re-write MDL; such that the maximum likelihood is common to both criteria. To make derivation of the MDL's easier and understandable, the study employ the sum of square versions of the MDL as discussed in Chapter Three [also further explained by Hansen and Yu (2001) and further broken down with examples by Stine (2003)].

Statistical Programming Language: The R-software

R is an object oriented programming language. It is a command-line-based language that requires all commands are entered directly into the console. It has some inbuilt functions already pertaining to information criteria like AIC and BIC. The goal of this research is to re-write the four formulations of MDL (Rissanen, Normalized Maximum Likelihood, Exact Maximum Likelihood and G-Prior Mixture) into R-functions. In order to write all the four we must start with the very simple one which is similar to AIC; that is Rissanen's approach (see Appendix J1).

Derivation of MDL (Rissanen's approach) from AIC using R-Software

We start with setting or defining an arbitrary response variable, and its predictor variables from a set of random numbers. We also need to ascertain the same set of random numbers are generated each time we call for our parameters and response variable and this is achieved by the following command 'set.seed ()' (see Appendix J2 for an example and derivation of MDL from AIC using the R-software.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Also note that this research used the random normal distribution and the distribution can change depending on the researcher's interest. Arbitrary predictor and response variables for our model are defined using a sample size of 1000 which can also change. Let $x_{1,} x_{2,} x_3$ and x_4 be arbitrary predictor variables and y the dependent variable, and then the following equations can be defined in R console as

```
lm_fit1 <- lm(y ~ x1 + x2 + x3 + x4)
lm_fit2 <- lm(y ~ x1 + x2 + x4)
lm_fit3 <- lm(y ~ x2+ x4)
lm_fit4 <- lm(y ~ x1 + x2)
```

Calling AIC to determine which of the following models (fit1, fit2, fit3, and fit4) is the true data generating process is as follows:

```
AIC(lm_fit1)
AIC(lm_fit2)
AIC(lm_fit3)
AIC(lm_fit4)
```

As expected in Table 2, lm_fit4 yielded the least AIC since the least error is preferable. Thus, the better model. We now calculate AIC manually in order to find the maximum likelihood ratio and use that to re-write the formula for minimum description length (MDL). Along the way, BIC will be introduced for comparison.

Using the formula:

AIC = k*npar - 2ln(L)

where,

npar = number of parameters (which includes the constant)

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

L = Maximized Likelihood

k -> this is 2 for AIC and log(#data points) for BIC

Here, the Log-Likelihood value can be extracted using R function logLik(). Redefining the AIC formula, we have

```
AIC_gid <- function(model=fit){
  AIC_g <- 2*(length(model$coefficients)+1) - 2*as.numeric(logLik(model))
  return(AIC_g)
}
```

Redefining the BIC formula;

```
BIC_gid <- function(model=fit){
 BIC_g <- log(nrow(model$model))*(length(model$coefficients)+1) -
2*as.numeric(logLik(model))
 return(BIC_g)
}
```

Comparing AIC and BIC inbuilt functions to the manually calculated one will ensure the manual manipulations that were used to develop the MDL holds.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Table 2: Criteria Comparison of Inbuilt verses Manual Calculation

Functions		Value
Inbuilt	AIC(lm_fit4)	8538.276
	BIC(lm_fit4)	8557.907
Developed	AIC_gid(lm_fit4)	8538.276
	BIC_gid(lm_fit4)	8557.907

Table 2 confirms that manually calculating the criteria in R and using the already inbuilt function is the same. This paves the way for us to confidently calculate the maximum likelihood value from AIC formula and use this to develop the MDL function from scratch. No such inbuilt functions exist yet but other programming languages may have some work on MDL.

Developing the Minimum Description Length Criteria

As established, the AIC formula given as

$$AIC = k*npar - 2\ln(L)$$

is key in deriving the MDL formula in terms of maximum likelihood,

where

npar = number of parameters (this includes the constant),

L = Maximized Likelihood,

k = this is 2 for AIC and log(#data points) for BIC,

Thus, the Log-Likelihood value can be extracted using R function logLik().

Minimum Description Length

Rissanen's minimum description length (MDL) principle (Rissanen, 1978) formalizes the two-part codes which compresses data optimally by maximizing its likelihood and at the same time how data was encoded

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

(knowing the model associated with the data). To convey the model to the receiver, one must include the parameter estimates as part of the message to the receiver. Thus, the MDL of a fitted model is the sum of two parts; the first part of the description length represents the complexity of the model and this encodes the parameters of the model itself – it grows as the model becomes more complex. The second part of the description length represents the fit of the model to the data; as the model fits better, this term shrinks. The choice of a code for the rounded estimates is equivalent to the dilemma faced by a Bayesian who must choose a prior distribution. A large-sample approach leads to the association between MDL and BIC. Comparisons of these two criteria in Chapter Two showed that the complexity term of BIC may appear equivalent to MDL (they are equivalent under the spike-and-slab prior but for other priors it gives different results).

The concept of universal code as against the optimal priors in developing codes is the way forward now since it is tailored to specific problems and works well over a wide range of distributions. Analogous to robust estimates in statistics, universal priors encode as well as the best prior, but they do not require that one knows the "true" model and based on its robustness, universal priors are well suited for model selection than the former. Thus, the idealized length of the universal code is calculated (more suited for regression models) and it shows that the description length leads to a version of MDL that resembles AIC (Forster & Stine, 1999) in structure and this gives us a basis for comparison as suggested by Barron and Cover (1991).

Calculating the Idealized Length in R

```
t.values <- as.numeric(summary(lm_fit1)$coefficients[,3])
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Idealized length of universal code for an integer j != 0

l(j) = 2 + log2(abs(j)) + 2*log2log2(abs(j)) where the logs are the positive side such that log2(x)=0 for |x|<= 1

The function that calculates the Idealized length which is the penalty of using a complex model is based on the t values from the model.

```
IDl_len <- function(z){
  x <- round(z)
  frst_log <- ifelse(abs(x)<= 1, 0, log2(abs(x)))
  scnd_log <- ifelse(abs(frst_log)<=1, 0, 2*log2(frst_log))
  id_len <- 2 + frst_log + scnd_log
  return(id_len)
}
```

Example: IDl_len(3) is approximately 4.92

Sum of Idealized length for a vector

sum(sapply(t.values,IDl_len))

The function for the MDL using the Residual Sum of Squares (RSS) and Idealized Length is given as

MDL = (n/2)*log2(RSS/n) + sum(Idealized_length)

The formula holds after eliminating constants that do not affect selection.

Writing a function for MDL

MDL <- function(fit = model){

t.values <- as.numeric(summary(fit)$coefficients[,3])

That is to extract t-statistic value from the model including the intercept.

RSS <- sum((fit$residuals)^2) – Sum of Squares residual

n <- nrow(fit$model) – The number of data points

Digitized by Sam Jonah Library

```
 log_lkl <- (n/2)*log2(RSS/n) – calculating the log likelihood
 unv_cdlen <- sum(sapply(t.values,IDl_len)) – Calculates the universal codes
 Mdl <- log_lkl + unv_cdlen
 return(Mdl)
}
```

Functions for calculations of AIC and BIC (both full version and when constants are eliminated) can be found in Appendix H1. The value of MDL criteria for all our models (fit1, fit2, fit3 and fit4) is called for and compared. The least value (lm_fit4) is therefore the best model.

MDL(lm_fit1) = 4124.467

MDL(lm_fit2) = 4122.47

MDL(lm_fit3) = 4128.291

MDL(lm_fit4) = **4121.984**

Hence, model 4 is the best in describing the data generating process. This concept is extended to AIC and BIC since we have now developed an MDL function in R. This gives us a basis to write more program codes to calculate all test statistics to help us use big data through simulation. The algorithms will confirm or otherwise after the long run a model's ability to retrieve the data generating process of a particular distribution and find out if the former model will still be the best. After this, we now compare MDL to the other criteria and evaluate their efficiency.

Developing R-Functions for all four formulations of MDL

Our task now is to develop codes in R for some formulations of MDL used in this study. The sums of squares versions of these equations [the formulations or equations which are extensions of rMDL (see Chapter Two)]

Digitized by Sam Jonah Library

© University of Cape Coast   https://erl.ucc.edu.gh/jspui

were used. The study developed codes and idealized lengths based on our initial rMDL algorithm for the remaining extensions (gMDL, nMDL, eMDL) and each converted into R-functions. The house-style is to write all codes as simple as possible so that when they are introduced into the iteration loop, the computations will be fairly easy for the computer and less time will be used in arriving at the results (see Appendix H2 for the R – functions for the four MDL's).

Developing R-Functions for the Asymmetric Price Transmission (APT) Linear Models

The study looked at three APT models in theory; the Houck's Specification Model denoted as Houck's_L (fit_3), Standard Error Correction Model denoted as SECM (fit_2) and Complex Error Correction Model denoted as CECM (fit_1) which were discussed in the second chapter. R codes were also developed for the three price models basically based on the work of Acquah and Von Cramon-Taubadel (2009). Some modifications (optimized by making it shorter) were made to suit this current study (see Appendix H3). The study compares all criteria under study (MDL, AIC and BIC) for the three price models.

Simulation and Results of Computations

This research has found simulation as a very useful technique for building models that can be simulated and for solving modelling problems involving huge observations. This research specifically used Monte Carlo simulation to solve the problem of getting repeated samples and able to use huge data. Simulation allows analysts to easily create many samples of data in a computing environment which allows us to assess patterns that appear across those repeated samples. This research emphasizes the point that Monte Carlo

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

simulation is a powerful tool for assessing new methods and comparing competing methods. This was done with the R-programming language. Data was drawn from the random normal distribution in all our computations. One of the tasks for all the information theoretic fit criteria was to be able to identify the true data generating process.

All codes developed for the computations of the various criteria (AIC, BIC, MDLs) can be found at Appendix H4. The package 'Hmisc' was installed to help with R-codes used in calculating 'lags' or differencing of time series (our price models). In order to load package 'Hmisc' it is necessary to also load the other packages (see Appendix J3).

The 'actual' R-codes developed for computations in this research basically stem from the hypothetical scenario illustrated earlier in this section. All codes for Idealized length which was used in developing the MDL formulations were calculated and then the R-functions of rMDL, gMDL, nMDL and eMDL were developed. The four formulations of the MDLs were also combined together into one function named the 'combo', which is useful during simulations (for optimization).

The simulation was started by first of all creating empty matrices [fit_1 <- data. frame ( )] to hold output from calculations of averages and output of simulations. This function creates data frames, tightly coupled collections of variables which share many of the properties of matrices and of lists, used as the fundamental data structure by most of R modelling software framework.

The number of iterations (Niter) was set to 1000; this number can be varied depending on the purpose of study and in our case for purposes of comparison. The sample size n was set to 1000 (this can also change). A seed

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

was also set to maintain the exact random numbers generated for consistency. R-functions for both the predictor and response variables of interest were then developed based on the asymmetric price transmission linear models denoted as follows:

Complex Error Correction Model (CECM)

lm_fit1 <- lm(Dyt ~ dxxpos + dxxneg + LECTpos + LECTneg-1)

Standard Error Correction Model (SECM):

lm_fit2 <- lm(Dyt ~ Dxt + LECTpos + LECTneg-1)

Houck's Model in Summed Difference:

lm_fit3 <- lm(dyy ~ dxxpos + dxxneg-1)

The arbitrary model of interest (data generating process) whose parameters were based on Holly et al. (2003) and others is denoted in R-codes as follows:

dyt <- 0.7*dxt + 0.25*lECTpos + 0.75*lECTneg + rnorm(n-1, 0, 1)

(SECM)

dyt <- 0.95*dxxpos + 0.20*dxxneg - 0.25*lECTpos - 0.75*lECTneg + rnorm(n-1, 0, 1) **(CECM)**

It is recalled that:

$dy = dyt =$ change in output price = change in response variable

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

$dx = dxt$ = change in input price = change in predictor variable

$dxxpos$ = positive change in input price

$dxxneg$ = negative change in input price

$ECT = ICET$ = error correction term = residuals

$ICET_{pos} \quad = LECT_{pos}$ = positive lag residuals

= positive deviations from long − run equilibrium

$ICET_{neg} \quad = LECT_{neg}$ = negative lag residuals

= negative deviations from long − run equilibrium

The random normal distribution was chosen as the data generating process for this study (AIC and BIC were actually build on the random normal distribution too) hence it is just ideal to use it as the basis of comparison.

The following outputs are the R-functions for the three APT models the study employed:

```
> lm_fit1

Call:

lm(formula = Dyt ~ dxxpos + dxxneg + LECTpos + LECTneg - 1)

Coefficients:

 dxxpos   dxxneg  LECTpos  LECTneg

 0.6887   0.7989   0.2658   0.6746

> lm_fit2

Call:

lm(formula = Dyt ~ Dxt + LECTpos + LECTneg - 1)

Coefficients:

     Dxt  LECTpos  LECTneg

 0.7447   0.2323   0.7129
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

> lm_fit3

Call:

lm(formula = dyy ~ dxxpos + dxxneg - 1)

Coefficients:

dxxpos dxxneg

0.5634 0.9551

Therefore, our three models for consideration are generated and matrices are then created to receive the average calculated criteria using the data.frame code for the models.

Data Analysis and Results

The study now looks at the performance of the three APT linear models vis-à-vis the performance of all six criteria. Firstly, we will examine the performance of all criteria against the three models using the number of observations (*n*) to be equal to 50 (small), 150 (moderate), and 500 (large) which is consistent with previous researches. Also, the number of iteration (niter) is 1000. This study will also look at the performance of the model selection methods in selecting the true data generating process (CECM, SECM, HOUCK'S) within conditions of sample size, stochastic variance and level of asymmetry.

General performance of models and criteria

Previous studies, in econometric APT modelling, considered sample sizes of cases between 100 and 300 to be moderate, below 100 to be small and above 500 to be large (Holly et al., 2003; Acquah & Von Cramon, 2009, Von Cramon & Meyer, 2000). The study will first of all consider a general case

Digitized by Sam Jonah Library

© University of Cape Coast   https://erl.ucc.edu.gh/jspui

where we look at the mean performance of the models across all criteria (rMDL, nMDL, eMDL, gMDL, AIC, BIC) and the ability of the criteria to predict the models to select the Data Generating Process {standard random normal distribution = rnorm (0, 1)} for every 1000 Monte Carlo Simulations and a sample size of 1000.

This gives us a general overview of what is happening to both the criteria and the models. In all analysis, the average criteria for each model and the number of times (in percentages) each criterion was able to predict (select) the data generating process (DGP) were calculated and tabulated. The three models of APT were used in the analysis for purposes of comparison. In all simulations, a particular data generating process was used as the 'starting model' and all three models were explored but the Houck's model performed very poorly. Thus, due to non-competitive nature of the Houck's model the actual competition was between the Standard Error Correction Model and Complex Error Correction Model and so they were used as the starting models for each comparison across conditions of varying sample size, stochastic variance and difference in asymmetric adjustment parameters.

General overview of performance of criteria and models when data generating process is SECM and CECM

This section looked at how the various criteria performed and also how they are able to predict the number of times the original data generating process of all models are selected. The study employed 1000 Monte Carlo simulations throughout the entire work. Although, subsequent analysis looked at the effect of sample size on performance, which were varied based on previous works of some economists (Holly, Turner and Weeks, 2003; Meyer and Von Cramon, 2004; Acquah, 2010), and in some cases a moderate sample

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

size was used for purposes of comparison. A sample size of 1000 was used in this general overview of the performance of each criterion and how they are able to select the true DGP.

Tables 3 and 4 compare the grand averages across each model in the cases when the original DGP is SECM and CECM, respectively.

Table 3: Criteria Averages when DGP is SECM

	rMDL	gMDL	nMDL	eMDL	AIC	BIC
CECM	33.41405	18.13018	12.59736	21.06456	2830.934	2855.448
SECM	26.64921	15.89862	10.79782	16.89012	2829.915	2849.526
HOUCK'S	199.73897	138.44222	133.95012	262.38379	3080.173	3094.881

Sample size (1000) 1000 Monte Carlo Simulations

Table 4: Criteria Averages when DGP is CECM

	rMDL	gMDL	nMDL	eMDL	AIC	BIC
CECM	30.20129	18.35132	12.81850	21.50684	2830.934	2855.448
SECM	78.47769	55.12114	50.02034	95.33516	2908.553	2928.164
HOUCK'S	192.48686	137.68654	133.19444	260.87244	3078.316	3093.024

Sample size (1000) 1000 Monte Carlo Simulations

Table 3 revealed that, with a sample size of 1000 observations in a 1000 simulations and when the data generating process is SECM, all criteria recorded lower values for the standard model, then followed by complex model and Houck's model. Thus, when the true data generating process was SECM, all criteria were able to identify the true data generating process with lower average values for the standard model (SECM) than for any of the other two. Table K1 (see Appendix K) which provides information on the criteria and model rankings reveals that there are differences in the performances of

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

criteria regarding selected alternative models and this forms the basis for their different rankings indicated in Table K1. Evidently, all criteria selected the true data generating process (SECM) indicated by a ranking of one (1), followed by CECM with a ranking of two (2) and finally Houcks with a ranking of three (3). This implies that the MDLs (rMDL, gMDL, nMDL and eMDL) can be used as alternatives to AIC and BIC when choosing which set of models point to the true data generating process or which model among competing models is reliable.

In Table 4, given the same conditions but with the true data generating process being CECM (complex model), all criteria selected the complex model as the true data generating process with the lowest values. Similarly, Table K2 in Appendix K reveals that, all criteria performances are consistent in choosing the true data generating process with a ranking of one (1) for CECM, two for SECM and finally Houcks, a ranking of three (3). Also, Houcks model provided the poorest fit (Table 3 and Table 4) to the data and this is due to the fact that the Houcks model does not incorporate the equilibrium relationships of cointegration in price series like the standard and complex error correction models (SECM and CECM) and this affects the performance of selecting the true data generating process. This confirms the assertion that in the presence of cointegration of the data, an error correction mechanism exists (see discussion provided in Chapter Three).

In general, irrespective of which model (SECM or CECM) is used as DGP and given the same conditions (sample size of 1000 and 1000 simulations), the criteria values seems to be almost the same for the Complex Error Correction Model and the Houck's Model. But the criteria values for the

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Standard Error Correction Model are dependent on what the initial DGP is. Subsequently, given the six criteria, the normalized maximum likelihood MDL (nMDL) recorded the minimum means value across the three models for the two simulations. Therefore, the nMDL tends to be doing better compared to all criteria followed by G-prior mixture MDL (gMDL). AIC and BIC recorded very high averages.

The study continued to verify after 1000 simulations what percentage of time a particular criteria is able to select each model (under consideration) to follow a particular DGP. The SECM and CECM data generating processes were chosen for the first and second experiment, respectively, and are shown in Table 5 and Table 6.

Table 5: Percentage of Time Criteria Correctly Select Model to Predict the SECM-DGP

(%)	CECM	SECM	HOUCK'S
rMDL	0.2	99.8	0
gMDL	2.0	98.0	0
nMDL	3.2	96.8	0
eMDL	2.4	97.6	0
AIC	14.7	85.3	0
BIC	0.9	99.1	0

Sample size (1000) 1000 Monte Carlo Simulations

Table 6: Percentage of Time Criteria Correctly Select Model to Predict the CECM- DGP

(%)	CECM	SECM	HOUCK'S
rMDL	100	0	0
gMDL	100	0	0
nMDL	100	0	0
eMDL	100	0	0
AIC	100	0	0
BIC	100	0	0

Sample size (1000) 1000 Monte Carlo Simulations

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

The results showed that each time the DGP is SECM or CECM, all criteria are able to select the true data generating process to a very large extent. Consistently, if the DGP is CECM, all criteria selected the CECM to be the true DGP 100% of the time although, for SECM, percentage of prediction was above 90% except for AIC which recorded 85% recovery rate. In both cases, all criteria consistently do not identify the Houck's model 100% of the time. In other words, all criteria identifies 0% of the time the Houck's model as it is rightly not the DGP. Thus, it is very easy for all ITFC to rightly detect the Houck's model as the wrong model. Specifically, rMDL outperformed (99.8) all criteria with BIC being comparable (99.1%). The remaining MDLs (gMDL, eMDL and nMDL) outperformed AIC. The Houck's model records 0% rate of recovery all the time by all criteria as expected.

As a matter of emphasis, data simulation employed in this research means generating data from asymmetric data generating process and exploring the analysis in repeated sample. This research designed a Monte Carlo experiment using 1000 simulations to test whether Minimum Description Length criteria on the average identifies a true asymmetric DGP out of competing models. Also, the Monte Carlo simulation emphasizes the relative performance of the model selection procedures in a price transmission modelling framework of which no studies have been undertaken so far.

To make comparisons realistic and to be able to generalize in our subsequent analysis, this research draws on the experimental designs of Cook et al. (1999), Cook and Holly (2002), Holly et al. (2003) and Acquah and Von Cramon-Taubadel, (2009). Estimates of the coefficients are used to generate artificial data for the Granger and Lee (1989) asymmetric error correction

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

model (SECM) and the Von Cramon-Taubadel and Loy (1996) asymmetric error correction models. A moderate sample size of 150 observations was used for purposes of comparison regarding discussions on level of asymmetry and stochastic variance. Further, varying sample sizes of 50, 150 and 500 was employed to determine its effect on the performance of both criteria averages and the number of times they could predict a model as the specified DGP as demonstrated in later results.

Comparison of Criteria Model Recovery Rates of the Tests of Asymmetry when True Model is SECM

In this section, the relative performance of the model selection criteria in recovering the true DGP is evaluated by simulating the effects of sample size, the amount of noise in the model (stochastic variance) and level of asymmetry on model selection. That is, the three competing models are fitted to the simulated data and the criteria ability to recover the true model was measured (Model Recovery Rate). The model recovery rates define the percentage of time samples in each competing model provides a better model fit than the other competing models. The derivation of recovery rates is done using 1000 Monte Carlo simulations. In other words, the number of samples for which each model fits better among competing models, is measured out of 1000 samples and expressed as a proportion/percentage. The proportions derived from each model by selection methods are the arithmetic mean based on 1000 samples.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Monte Carlo Simulation of Effect of Sample Size on Model Selection-SECM

The study compares the relative performance of model selection across sample sizes of 50 (small), 150 (moderate) and 500 (large) in Table 7. Generally, all model selection criteria correctly recovered the true DGP (SECM) and lower percentages were recorded for CECM but the Houck's model which is included in all simulations performed poorly for all model selection methods (rMDL, gMDL, nMDL, eMDL, AIC, BIC) and was never or was least (0% for large and moderate samples and below 10% for small samples) recovered due to poor fit. This may be due to the fact that Houck's model is not consistent with cointegration between price series.

The recovery rate for SECM in the case of small sample size for all model selection criteria ranged between 64% and 86%. Two MDL criteria (gMDL- 86%, and eMDL- 84 %) competed very well with BIC whose rate of recovery was 86% (highest). Recovery rate in the case of moderate sample size, increased to 91-98% except for AIC that recorded 86%. Also, the performance of rMDL (97%) again was similar to BIC (98%).

Although, BIC did better than rMDL in small to moderate samples, rMDL approximately reached full recovery (100%) in large sample and outperformed BIC (99%) and BIC in turn outperformed the remaining MDLs. rMDL and BIC are comparable in moderate to large samples whereas gMDL and BIC are comparable in small samples.

Digitized by Sam Jonah Library

© University of Cape Coast   https://erl.ucc.edu.gh/jspui

Table 7: Relative Performance of Model Selection across Sample Size

Sample Size		Model Fitted (%)		
50		CECM	SECM (DGP)	HOUCK'S
	Methods			
	rMDL	29.9	63.5	6.6
	gMDL	9.0	85.8	5.2
	nMDL	20.6	78.0	1.4
	eMDL	12.2	84.2	3.6
	AIC	17.7	79.5	2.8
	BIC	5.0	86.3	8.7
150		CECM	SECM (DGP)	HOUCK'S
	Methods			
	rMDL	3.3	96.7	0
	gMDL	4.6	95.4	0
	nMDL	8.3	91.7	0
	eMDL	5.8	94.2	0
	AIC	13.9	86.1	0
	BIC	2.1	97.9	0
500		CECM	SECM (DGP)	HOUCK'S
	Methods			
	rMDL	0.5	99.5	0
	gMDL	3.3	96.7	0
	nMDL	5.2	94.8	0
	eMDL	3.6	96.4	0
	AIC	16.4	83.6	0
	BIC	1.3	98.7	0

Based on 1000 Monte Carlo Simulations

This may be suggestive that although the MDLs do very well, the various formulations' performances are also affected by change in sample size. Specifically, gMDL and eMDL do well with small samples whiles rMDL (highest), nMDL and gMDL have strength in moderate to large samples. Houck's model was never recovered (0%) for moderate to large samples. There was an improvement in recovery rate for all model selection criteria in choosing the true DGP in all sample categories except for AIC that decreased in the large samples.

The general pattern of effect of sample size on recovery rate of the model selection criteria when the DGP is SECM is clearly depicted in Figure

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

2; Appendix L. Irrespective of the sample size being small or large, MDLs perform very well. In general, all criteria recovery rate improved with increasing sample size although AIC's performance decreased in large samples. Specifically, with a sample size of 50, all criteria (except rMDL and nMDL) tend to do better than the AIC, though not consistent, it is known to perform relatively well with small sample sizes. BIC's performance is also consistent with literature where it does very well with large samples. Thus, the performances of the formulations of MDL in selecting the true data generating process has a wide range of uses whether we are dealing with small or large sample sizes. Also, the similar results of MDLs to BIC and sometimes to AIC stem from our discussions (Chapter Three) in the derivation of information criteria having a basic structure of what Barron described as minimum complexity density estimators. Their differences only come as a result of their penalty for complexity and MDLs specifically are moderations of BIC.

Monte Carlo Simulation of Effect of Stochastic Variance on Model Selection-SECM

In order to simulate the effect of noise level on model selection, this study considered three different standard deviations from the mean (noise levels) and categorized as small, moderate and large which corresponds to 1.0, 2.0 and 3.0, respectively.

A sample size of 150 was used to generate the models with the different error sizes. Essentially, the data fitting abilities of alternative models are compared in relation to the true model as the error in the data generating process was increased systematically. The results of 1000 Monte Carlo

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

simulations comparing the performance of the model selection methods as error size (σ) increased are displayed in Table 8.

Generally, recovery rates of the true asymmetric data generating process (i.e. SECM) declined for all model selection methods as the error increased. A moderate sample size of 150 and an error size (σ) of 3 and 2 recorded between 1 – 52% of ability to select the DGP for Houck's model for all criteria, whereas nothing was recorded for error size of 1.0. However, for the criteria recovery rates for the CECM when the true DGP is SECM range between 0.5 – 43% for noise levels of 2 – 3 and between 2.1 – 1.4% for lower noise levels.

The percentage of the simulated data in which the correct model (i.e. SECM) was selected or recovered among competing models by the model selection criteria as the amount of noise in the DGP increased. The recovery rate for the SECM model, at lower noise level (1), revealed that BIC outperformed all criteria and all the MDLs outperformed AIC. rMDL (97%) can be used alongside BIC (98%) whilst gMDL (95%) was comparable to nMDL (94%) Alternatively, when noise increased to 2, gMDL (85%) and eMDL (84%) outperformed all criteria but AIC (82%) did slightly better than BIC (81%), thus comparable to each other. Subsequently, with much higher noise (3), AIC (68%) outperformed all criteria but comparable to eMDL (68%) and this is followed by gMDL (67%). Thus, eMDL and gMDL show stronger recovery rate in noisy data. All MDLs except rMDL outperformed BIC.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Table 8: Relative Performance of Model Selection across Noise Levels Based on Sample Size of 150-SECM

Error Size		Model Fitted (%)		
3		CECM	SECM (DGP)	HOUCK'S
	Methods			
	rMDL	21.2	38.9	39.9
	gMDL	11.9	66.9	24.1
	nMDL	42.8	53.7	3.5
	eMDL	18.8	67.5	13.7
	AIC	10.0	68.1	21.9
	BIC	0.5	47.5	52.0
2		CECM	SECM (DGP)	HOUCK'S
	Methods			
	rMDL	29.2	60.3	10.5
	gMDL	8.8	84.9	6.3
	nMDL	21.0	77.1	1.9
	eMDL	11.5	84.2	4.3
	AIC	13.1	82.0	4.9
	BIC	1.0	80.6	18.4
1		CECM	SECM (DGP)	HOUCK'S
	Methods			
	rMDL	3.3	96.7	0
	gMDL	4.6	95.4	0
	nMDL	8.3	91.7	0
	eMDL	5.8	94.2	0
	AIC	13.9	86.1	0
	BIC	2.1	97.9	0

Based on 1000 Monte Carlo Simulations

Notably (see Figure 3; Appendix L), the MDLs (gMDL, eMDL, nMDL) performed similarly to one another with their recovery rates decreasing substantially as noise levels increased. Interestingly, when error size increases, rMDL and BIC drastically decreased but gMDL, AIC, eMDL and nMDL were steady.

The MDLs continue to perform comparatively whether the noise is high or low thus making them to be robust given any situation. This means whatever the noise level; rMDL performs well (alongside BIC) with lower

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

noise levels whiles eMDL perform very well (alongside AIC) at higher noise level. Moderate noise levels sees gMDL and eMDL outperforming the traditional AIC and BIC.

Comparison of the different selection methods showed a general trend in which recovery rates decreased with increasing error sizes. In effect, the performance of all model selection algorithms analysed deteriorated with increasing amount of noise in the true asymmetric price transmission data generating process.

Concurrent effect of sample size and stochastic variance on model selection-SECM

It will also be of interest if we now take a look at the concurrent effect of sample size and stochastic variance on the model selection criteria ability to effectively recover the data generating process of the true distribution.

Simulating the effects of sample size and stochastic variance concurrently affirms that a small error and large sample improves recovery of the true asymmetric data generating process and vice.

Thus, with a small sample of 50 and an error size of 2.0 (unstable condition), the true data generating process was recovered at least 26 percent to 49 percent of the time by all the model selection criteria as illustrated in Table 9. On the other hand, with a relatively large sample of 150 and error size of 0.5 (stable condition), at least 86 percent to 100 percent of the true data generating process was recovered across all the model selection methods. Additional information is provided in Appendix I 1.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Specifically under stable conditions, the rMDL (99.5%) outperformed all criteria with BIC (97.9%), eMDL (97.3%) and gMDL (97.7%) in turn performing similarly.

Table 9: Stable and Unstable Conditions of Asymmetry-SECM

Difference	Selection Criteria	Model Fitted (%)
Stable		
	Rmdl	99.5
	gMDL	97.7
	Nmdl	96.4
	Emdl	97.3
	AIC	86.1
	BIC	97.9
Unstable		
	rMDL	26.1
	gMDL	49.1
	Nmdl	27.1
	Emdl	44.7
	AIC	45.2
	BIC	30.7

Based on 1000 Monte Carlo Simulations

Evidently, all MDLs outperformed AIC in recovering the true data generating process. On the other hand, with unstable conditions, AIC (45.2%) naturally improved in recovery rate but the gMDL (49.1%) outperformed AIC and all other criteria. Generally, all criteria recovery rates were generally low (under 49%) when study conditions are unstable.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Monte Carlo Simulation of the Effects of Difference in Asymmetric Adjustment Parameters on Model Selection-SECM

This research also took into consideration the probable effect of difference in asymmetric adjustment of parameters on the ability of the model selection criteria to recover the true data generating process.

Simulated data of sample size 150 with an error size of 1 from the standard asymmetric price transmission model and asymmetry values $(\beta_1^+, \beta_2^+) \in (-0.25, -0.50)$ or $(-0.25, -0.75)$ were considered for the coefficients of the asymmetric error correction terms. Subsequently, examination of the effect of the increase in difference of asymmetric adjustment parameters on model recovery was investigated. Different model selection methods exhibit different relative performance in recovering the true model at different levels of asymmetry.

Table 10: Varying Levels of Asymmetry When n=150

Difference	Selection Criteria	Model Fitted (%)
0.50 (Strong)		
	rMDL	96.7
	gMDL	95.4
	nMDL	91.7
	eMDL	94.2
	AIC	86.1
	BIC	97.9
0.25 (Weak)		
	rMDL	96.2
	gMDL	94.1
	nMDL	90.8
	eMDL	93.5
	AIC	86.1
	BIC	96.3

Based on 1000 Monte Carlo Simulations

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

In the moderate samples, an increase in the difference between the asymmetric adjustments parameters from 0.25 to 0.5 led to improvement in the model recovery rates of the model selection methods (see Table 10).

Specifically, when the level of asymmetry was weak, BIC (96.3%) and rMDL (96.2%) were comparable and outperformed all other criteria. Though the pattern was similar for strong asymmetry, BIC (97.9%) recovered slightly better than rMDL (96.7%). Meaning, all criteria performance increased when the difference in asymmetric adjustment on model selection also increased. Comparatively, BIC responded stronger than all criteria to increase in asymmetric adjustment. This is followed by the MDLs and lastly AIC.

The experiment was repeated for small sample size and large sample size to ascertain if there might be a concurrent effect of different asymmetric adjustment levels across sample sizes (n= 50 and 500) on model selection. The difference in sample size had an effect on the level of asymmetry across the different sample sizes and performance of all criteria increased with increasing sample sizes. That is, an increase in the difference between the asymmetric adjustments parameters from 0.25 to 0.5 led to improvement in the model recovery rates of the model selection methods but this improvement becomes almost comparable as the sample size increases for some criteria.

Specifically, for a small sample size, gMDL (73.4%) and eMDL (73.3%) outperform all criteria at weak levels of asymmetry and AIC and nMDL are comparable. Then at the same sample size but a stronger level of asymmetry, BIC out performed all criteria but comparable to gMDL with rMDL recording the least. The MDLs and AIC increased steadily with increased in asymmetric adjustment but although BIC did not fare well in

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

weak levels of asymmetry (small samples) it responded stronger (68.8% to 86.3%) than all criteria in strong levels of asymmetry. Also, gMDL (85.8%) recovered similarly well in stronger asymmetry.

Lastly, for large samples and a weak level of asymmetry, rMDL outperformed all criteria but BIC did better than remaining MDLs and AIC. The same pattern was revealed for stronger level of asymmetry for large samples but the performance of BIC and AIC remained the same for both weak and strong levels of asymmetry. Thus, the MDLs improved in their ability to recover the true DGP with increased asymmetric adjustment but that of BIC and AIC did not improve. Evidently, rMDL recovery was the strongest as asymmetric adjustment increased by 0.25.

The MDLs continue to make a huge contribution even in the examination of the effect of increase in asymmetric adjustments across sample sizes. Thus, although all MDLs improved with increase in asymmetric level of adjustment across all sample sizes, gMDL improves with small samples and rMDL improves in both moderate to large samples.

Overview of the Performance Analysis of the Different Asymmetry Test using Simulated Data generated from the Complex Asymmetric ECM

This research empirically evaluated and compared the performance of the model selection methods in an asymmetric price transmission modelling context when the true data generating process was complex.

Previous research (Acquah & Von-Cramon-Taubadel, 2009; Gagne & Dayton, 2002; Markon & Krueger, 2004) continue to emphasize that the performance of model selection methods improve when the true model is

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

complex and this study will want to find out if that will be the same for variant of the minimum description length principle. As a matter of comparison this study repeats all analysis when the DGP is SECM for the CECM as well so as to have a common ground for objective inferences and validate literature.

Monte Carlo Simulation of the Effects of Sample Size on Model Recovery

The overall trends in performance across the different model selection criteria as the sample size increases are similar to those observed when the data was simulated from the SECM.

Table 11: Relative Performance of Model Selection across Sample Size- CECM

Sample Size	Model Fitted (%)			
50		CECM(DGP)	SECM	HOUCK'S
	Methods			
	rMDL	63.9	19.2	16.9
	gMDL	49.1	40.5	10.5
	nMDL	69.8	26.7	3.5
	eMDL	55.1	37.0	7.9
	AIC	62.6	31.0	6.4
	BIC	34.0	48.0	18.0
150		CECM (DGP)	SECM	HOUCK'S
	Methods			
	rMDL	90.1	9.9	0.0
	gMDL	92.3	7.6	0.1
	nMDL	95.0	5.0	0.0
	eMDL	93.3	6.6	0.1
	AIC	96.8	3.1	0.0
	BIC	87.0	12.8	0.2
500		CECM (DGP)	SECM	HOUCK'S
	Methods			
	rMDL	100	0	0
	gMDL	100	0	0
	nMDL	100	0	0
	eMDL	100	0	0
	AIC	100	0	0
	BIC	100	0	0

Based on 1000 Monte Carlo Simulations

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

In that, all model selection criteria were able to recover the true DGP (CECM). The recovery rate increased with increasing sample size (see Table 11) such that, for large samples, all criteria recorded a recovery rate of 100 percent. However, performance is poor for all ITFC in small samples.

Furthermore, for small samples, nMDL (70%) outperforms all criteria and followed by rMDL (64%), AIC (62%) with BIC being the least. Interestingly, in the moderate sample size case scenario, the AIC (97%) outperformed all MDLs and BIC given the complex asymmetric data generating process but nMDL (95%) performed similarly. The percentage of the simulated data in which the correct model (i.e. CECM) was selected or recovered by the model selection criteria across different sample sizes is examined. Notably, all criteria recovered the CECM fully (100%) for large samples and nMDL and AIC tend to select complex model for small to moderate samples. Thus, the ability of selection methods to recover the true DGP improves with complexity for APT models. This is because, complex models usually have many parameters and the combination of parameters is powerfully connected by nonlinear equations and this enables the data structure (not single) to change as a function of the parameter values of the model. Hence, the data structure is finely tuned so that the model fits a wide range of data patterns making it able to be easily recovered as the true data generating process.

Monte Carlo Simulation of the Effects of Noise levels on Model Recovery-CECM

The effects of noise on the recovery of the true model and model fit are detailed in this section. The data fitting abilities of alternative models are compared in relation to the true model as the noise level in the data was

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

decreased systematically from 3.0 to 1.0. The results of 1000 Monte Carlo simulations using a moderate sample size of 150 was used to compare the performance of the model selection methods as error size (σ) increased are displayed in Table 12. Generally, recovery rates of the true asymmetric data generating process (i.e. CECM) declined for all model selection methods as the noise level increased as expected. Thus, the SECM and Houck's model recorded an inverse relationship instead of an increase in recovery rate as the noise level increases.

Table 12: Relative Performance of Model Selection across Noise Levels –CECM When n = 150

Error Size	Model Fitted (%)			
3		CECM (DGP)	SECM	HOUCK'S
	Methods			
	rMDL	33.3	24.8	41.9
	gMDL	23.3	50.8	27.2
	nMDL	60.2	33.6	6.2
	eMDL	32.9	47.4	19.7
	AIC	23.1	50.8	26.1
	BIC	2.0	39.8	58.2
2		CECM (DGP)	SECM	HOUCK'S
	Methods			
	rMDL	62.7	19.6	17.7
	gMDL	44.1	44.2	11.7
	nMDL	63.8	32.1	4.1
	eMDL	51.5	40.9	7.6
	AIC	53.0	39.0	8.0
	BIC	16.0	55.7	28.3
1		CECM (DGP)	SECM	HOUCK'S
	Methods			
	rMDL	90.1	9.9	0.0
	gMDL	92.3	7.6	0.1
	nMDL	95.0	5.0	0.0
	eMDL	93.3	6.6	0.1
	AIC	96.9	3.1	0.0
	BIC	87.0	12.8	0.2

Based on 1000 Monte Carlo Simulations

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Notably, with an error size of 1, AIC (97%) outperformed all other criteria and nMDL (95%) did very well and the remaining MDLs outperformed BIC. On the other hand, as the error size increased (2-3), the recovery rate of AIC drastically declined (53%-23%) but the nMDL consistently outperformed all criteria for higher noise levels (error sizes of 2-3) and nMDL was comparable to rMDL at noise level of 2.

Furthermore, an important point one needs to note is that at lower noise levels, for the complex model, AIC performs the best and however for moderate noise levels, the nMDL is most reliable. On the other hand, as the noise level increases, the performance of AIC decreased as compared to nMDL and rMDL. Also, BIC consistently performs poorly with noise levels for complex models.

Thus, MDLs (in this particular case rMDL and nMDL) are affected by the various conditions regarding selecting the true model since whatever the situation; you will always find a formulation of MDL to give desired results consistently. Specifically nMDL has the propensity to select complex models with a lot of noise.

Concurrent effect of sample size and stochastic variance on model selection

Model recovery rates of the model selection methods are derived under combined conditions of a small sample (50) and large error size (2); known as unstable conditions and a relatively large sample (150) and a small error (0.5); which is also referred to as stable condition.

Deduction on the unstable conditions revealed that nMDL outperformed all criteria and generally recovery rate was relatively poor. Thus,

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

the true data generating process was recovered 66.9 percent of the time by the nMDL model selection criterion and below 31% for the rest of the criteria with BIC recording the least (3.8%) given the true DGP as illustrated in Table 13.

Table 13: Stable and Unstable Conditions of Asymmetry-CECM

Difference	Selection Criteria	Model Fitted (%)
Stable		
	rMDL	100
	gMDL	100
	nMDL	100
	eMDL	100
	AIC	100
	BIC	100
Unstable		
	rMDL	18.6
	gMDL	18.8
	nMDL	66.9
	eMDL	31.3
	AIC	15.2
	BIC	3.8

Based on 1000 Monte Carlo Simulations

This confirms from our earlier deduction that nMDL has the tendency to select unstable complex models. On the other hand, exploration of stable conditions revealed that all model selection criteria fully (100%) recovered the true DGP. Thus, under stable conditions, all criteria naturally select complex models. Notably, one may say that, all selection criteria tend to select complex models of asymmetric price transmission under large samples and minimal noise levels.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Monte Carlo Simulation of the Effects of Difference in Asymmetric Adjustment Parameters on Model Selection

In order to assess the effect of difference in the asymmetric adjustment parameters on the ability of the different selection methods to recover the true data generating process (CECM), this study simulated data of sample size 150 (50 and 500 were considered for more emphasis in Appendix I); and error size 1(Holy et al., 2003) and asymmetry values of $(\beta_2^+, \beta_2^+) \in (-0.25, -0.50)$ or $(-0.25, -0.75)$ which were considered for the coefficients of the asymmetric error correction terms. Subsequently, the effect of the increase in difference of asymmetric adjustment parameters on model recovery was examined.

Table 14: Varying Levels of Asymmetry when n=150-CECM

Difference	Selection Criteria	Model Fitted (%)
0.50 (Strong)		
	rMDL	90.1
	gMDL	92.3
	nMDL	95.0
	eMDL	93.3
	AIC	96.9
	BIC	87.0
0.25 (Weak)		
	rMDL	88.3
	gMDL	90.5
	nMDL	94.4
	eMDL	92.3
	AIC	96.4
	BIC	80.9

Based on 1000 Monte Carlo Simulations

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Regarding the moderate sample size, an increase in the difference in the asymmetric adjustment parameters from weak (0.25) to strong (0.5) asymmetric adjustment revealed an increase in model recovery of the true asymmetric data generating process as depicted in Table 14.

Also, graphical representation of results depicted in Figure 9 (Appendix L) shows that there is not much difference in performance of all criteria under both weak and strong levels of asymmetry. Specifically, the change in asymmetric adjustment in selecting the true DGP by AIC increased slightly (96.4% to 96.9%) from weak to strong level of adjustments (see Table 14).

Also, AIC outperformed all model selection criteria when the true DGP was CECM. The MDLs performed similarly but nMDL (95%), whose performance was similar to AIC, outperformed the rest of MDLs and BIC.

This study stated earlier that the experiment was repeated for small and large samples as well (see Appendix I 3). The general trend for the CECM data generating process saw an increase in model recovery rate as the asymmetric adjustment increased from weak to strong in the case of small sample size (n = 50). Especially, eMDL (84%) outperformed all criteria followed by AIC (80%) and AIC did better than the rest. In the large sample size scenario, all criteria recovered fully (100%) in both the weak and strong levels of asymmetry when the true DGP was complex. This makes them have an equal chance in selecting the original model completely. Deductively, the influence of large samples is necessary to achieve high recovery of asymmetric adjustments for linear econometric models of APT with complex dynamics.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Discussion

Firstly, we take a look at inferences made when the true data generating process is the Standard Error Correction Model whilst the second section looks at the Complex Error Correction Model.

Effect of Sample Size on Model Selection-SECM

The Monte Carlo simulation on the effect of sample size on model selection when the data generating process is the standard error correction model has so far been consistent with past research in econometric modelling framework of asymmetric price transmissions.

Generally, inspection of the recovery rates for the different model selection criteria illustrated the extent to which the true model (SECM) was recovered by each selection criteria across the different sample sizes. In comparison, a model's recovery rate of the true model improved significantly with increasing sample size (Acquah, 2017; Acquah and Von Cramon-Taubadel, 2009; Bickel and Zhang, 1992). Despite differences in performance among the model selection criteria, trends holding across the different criteria were evident in the simulation results (Myung, 2000; Markon and Krueger, 2004). In effect, the performance of the model selection methods to select the true model (i.e. recovery rates of SECM) generally increased with increased in sample size from small to large (50 to 500).

However, some distinct patterns were identified. The rMDL, gMDL, eMDL and nMDL performed better than AIC in their ability to recover the true GDP across all sample sizes except for sample size of 50 where AIC actually performed better than rMDL and nMDL by a small margin. Thus, gMDL and eMDL also outperformed AIC. Secondly, the MDLs'

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

performances among each other were not the same. That is, rMDL outperforms the remaining MDLs in moderate to large samples whilst gMDL outperforms the rest of the MDLs but was comparable to eMDL. This implies the rMDL, gMDL and eMDL performed extremely well. The BIC on the other hand, outperformed all criteria in small to moderate samples but did better than all MDLs except rMDL in large samples. All criteria recovery rates improved with increasing sample size although AIC's performance does not improve for moderate to large samples.

Therefore, the gMDL and eMDL can be used instead of AIC and rMDL can be used alongside BIC in moderate samples whilst rMDL can be used instead of BIC in large samples for recovering and selecting Standard Error Correction Models of APT. This is consistent with some work in structural equation modelling and even in Markon and Krueger's work (2004) that stated that although AIC performs relatively well in small samples (so is gMDL or eMDL and they do better) it is inconsistent and does not improve in performance in large samples. BIC in contrast appears to improve in performance relatively from small to large samples and is consistent and some variant of MDL (gMDL, rMDL and nMDL) also behave similarly. Thus, depending on a particular variant of MDL, they boast of strengths, whether the sample is small or large and hence this research confidently presents the various formulations of MDL as alternative and sometimes better model selection criteria as compared to AIC and BIC (these tend to have one-sided performance behaviours as stipulated in Hurvish and Tsai, 1990; Bickel and Zhan, 1992).

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

The model selection methods performed reasonably well in identifying the true model and their ability to recover the true asymmetric data generating process increased with increase in sample size. Intuitively, the results point to the fact that the sample size is important in the selection of the true asymmetric data generating process during price transmission analysis. Generally, larger sample sizes do improve the ability of the model selection methods to make correct inferences about asymmetric price transmission models.

Effect of stochastic variance on model selection-SECM

In the 1000 Monte Carlo simulation of the effect of stochastic variance on model selection, some distinct observations were made. Generally, comparison of the different selection methods shows a general trend in which recovery rates decreased with increasing error sizes on all information theoretic fit criteria with emphasis on the MDLs which have not been explored yet in APT model selection framework. In effect, the performance of all model selection algorithms analysed deteriorated with increasing amount of noise in the true asymmetric price transmission data generating process. This is consistent with works (Acquah, 2017; Roos et al., 2005; Hui et al., 2011; Yang, 2003; Gheissari and Bab-Hadiashar, 2004, Myung, 2000; Rissanen et al., 2010) on comparisons using other information criteria (AIC, BIC, RIC, KIC, GAIC, CAIC, GBIC, CP, SSD, G-CP, MAIC, SSC).

Also, rMDL, gMDL, nMDL and eMDL performed similarly to one another with their recovery rates decreasing substantially as noise levels increased. Although, BIC and rMDL decreased drastically when noise level increased rMDL decline faster than BIC. Whereas, gMDL and eMDL

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

performed creditably with increasing noise and recovers the true model better than rMDL and nMDL in small to moderate samples, AIC outperformed all criteria in small samples but did better than only rMDL, nMDL and BIC in moderate samples. AIC performance is consistent with previous linear regression analysis framework when MDL was not included (Chen et al., 2007; Yang, 2003).

However, as error size decreased from 3 to 1, the Minimum Description Length criteria (i.e. rMDL, gMDL, nMDL and eMDL) and BIC outperforms the AIC in recovering the true model with recovery rates of near to 100 percent.

Thus, MDLs also behave like the Bayesian criteria and in some cases better for some of the MDL formulations. gMDL and eMDL also behave like the AIC. The MDLs continue to be a better alternative for model selection with its numerous advantages. Intuitively, higher noise levels make it difficult for the model selection methods to identify the true asymmetric model or alternatively the performance of the model selection methods deteriorates with high levels of noise in the asymmetric price transmission modelling framework.

These results are generally consistent with those obtained by experts who studied the effects of noise levels on model selection in other applications such as linear regression models and computer vision applications (See Myung, 2000; Gheissari and Bab-Hadiashar, 2004; Yang, 2003). Importantly, Yang (2003) finds that the recovery rates of the true data generating process decreases with increasing noise levels in linear regression models. In conclusion Yang notes that selection can yield the wrong model at higher

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

noise levels. Additionally, this study using MDLs has confirmed that recovery rates of model selection criteria in recovering the true DGP decreases with higher noise levels.

Concurrent effect of sample size and stochastic variance on model selection-SECM

Again, the effect of noise and varying sample size was explored on the information criteria's ability to recover the true DGP. Thus, small sample sizes with a big error create an unstable condition while large sample sizes with a small error create a stable condition.

Stable conditions boost the performances of model selection criteria in recovering the true data generating processes as compared to unstable conditions. This is supported by other researches on AIC, BIC, KIC (Acquah, 2017, 2010; Hui et al., 2011) and this research has proved that all the MDLs behaved in similar manner. Specifically, under stable conditions rMDL with a near full recovery (99.5%), outperformed all criteria and BIC in turn did better than the remaining MDLs with above 97% recovery rates whilst AIC was the least. Under unstable conditions, though the performance of model selection criteria dropped drastically (below 49% to 26%), gMDL outperformed AIC (which is known generally to have strength when it comes to unstable conditions – Hui et al., 2011; Acquah, 2010) and AIC (although comparable to eMDL) in turn outperformed remaining criteria. The results indicated that it is always desirable to go for stable conditions (large samples with lower noise) when comparing standard APT models and lower noise levels for all criteria (except for AIC) but gMDL (best) and AIC are favourable in unstable conditions.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Finally, this research notes that model selection methods may have difficulty in identifying the true asymmetric model at higher noise levels and lower samples. Alternatively, the performances of all model selection methods in recovering the true model deteriorated with increasing noise levels within the Asymmetric Price Transmission modelling framework.

Monte Carlo Simulation of the Effects of Difference in Asymmetric Adjustment Parameters on Model Selection-SECM

This study has revealed that an increase of the asymmetric adjustments parameters from 0.25 to 0.5 led to improvement in the recovery rates of the model selection methods. Importantly, another factor which influences model selection or the recovery of the true data generating process is the difference in asymmetric adjustment parameters (Acquah 2010, 2017; Cook et al., 1999). Remarkably, difference in asymmetric adjustment parameters or speeds is important in the performance of the model selection methods in recovering the true DGP.

Generally, recovery rates of the BIC and Minimum Description Length criteria respond more strongly to increases in the difference between the asymmetric adjustments parameters (see also Han et al., 2014) for the true model than AIC. Concurrent effect of sample size on asymmetry was also explored. Within a small sample size, gMDL and eMDL outperformed all criteria with AIC and nMDL being comparable and rMDL and BIC performing poorly for weak asymmetry. Although, all criteria improved in terms of strong asymmetry, BIC did better than all but comparable to gMDL and eMDL in turn outperformed AIC. For moderate samples, although all criteria improved from weak to strong levels of asymmetry, BIC though

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

comparable to rMdL was the best and AIC the least. Subsequently, as sample size became relatively large (500 and above) rMDL consistently outperformed all criteria and was comparable to BIC. On the other hand, rMDL's performance reached approximately 100 percent in large samples whilst the performance of AIC remained unchanged from weak to strong asymmetry across moderate to large sample sizes.

Hence the improvement in recovery rate due to a strong or weak level of adjustment increased as sample size also increases. In other words, with small samples the difference in improvement in model recovery rate due to a strong level of adjustment was higher than the weak level of asymmetry. But as sample size increases, this difference was small. Thus, large samples can nullify the effect of the difference between weak and strong level of asymmetry in the selection criteria's ability to recover the true data generating process.

One can confidently say that, within the asymmetric price transmission modelling framework, this study has not only confirmed empirical work on the relative performance of the model selection (AIC, BIC) algorithms but has also established that the Minimum Description Length Criteria (MDLs) correctly identifies the true asymmetric data generating process similarly or sometimes better than AIC and BIC given that the data generating process is the standard error correction model.

The study now takes a look at what inferences one can draw if we change the true data generating process to be the complex error correction model (CECM).

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Monte Carlo Simulation of the Effects of Sample Size on Model Recovery-CECM

On the whole, model recovery rates by all criteria, increased with increasing sample size when the data generating process was complex. Thus, the MDLs (nMDL, rMDL) persistently outperform AIC and BIC criteria across small samples whiles across moderate sample sizes, AIC outperformed MDLs and BIC but was comparable to nMDL. In large samples the behaviour of all model selection criteria drastically changed and recovered the true DGP 100 percent which is in contrast to the pattern exhibited given the standard asymmetric data generating process.

The aforementioned pattern of increase in performance being affected by increase in sample size for all model selection criteria is consistent with literature (Markon and Krueger, 2004; Acquah, 2010, 2017; etc.), yet the 100 percent recovery by all information criteria in selecting the complex APT model in this study (given 1000 monte Carlo simulations) is of particular importance. Markon and Kruger in 2004 had a few cases of 100% recovery in large samples (1000 and above) for some criteria (BIC and DIC).

Obviously, the discussions so far point to the fact that another factor that may influence the performance of the model selection methods is model complexity (Harremoes and Brock, 2018; Fade, 2015; Han et al., 2014; Wagenmakers et al., 2006; Stine, 2003; Markon and Krueger, 2004; Sund, 2001; Pitt et al., 2002; Myung et al., 2005; Myung, 2000; Rissanen et al., 2010). Specifically, the MDL criteria achieve full recovery in a large sample size of 500 when the true model is complex. Overall this study notes that when the true model is complex, nMDL and rMDL performs better than AIC when

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

sample size is small. However, AIC does better than BIC and some formulations of the MDL only in moderate sample sizes (150) but all criteria have similar performance in large sample sizes of 500 and above. Similarly, previous studies (Lin and Dayton, 1997) found that AIC was superior to BIC when the true model was complex in mixed models. Gagne and Dayton (2002) also observed that AIC are more successful when the true model was relatively complex in multiple regression analysis. In small samples, the nMDL and rMDL outperforms AIC, and thus, making them more successful when the model is complex and the sample is small. The various formulations of the MDL can suit any condition within the econometric asymmetric prices transmission modelling framework. Thus, emphasizing the fact that the MDL criteria (rMDL, nMDL, eMDL and gMDL) are a very useful alternative. Results from this study support Li et al. (2012) emphasis on the ability of some variants of MDL to mimic the behaviour of AIC or BIC. Also, this study adds on to say that some variants of MDLs sometimes perform better and this is consistent with other researches (Haddadi et al., 2010; Fade et al., 2011; Sund 2001; Costa et al., 2012; Jiao et al., 2011; Velampalli and Jonnalagedda, 2017).

An important point is that comparatively, the model selection methods performed similarly or better in moderate to large samples when the true asymmetric data generating is relatively complex (CECM) than when the standard asymmetric data generating process (SECM) was used under the same conditions with AIC and MDLs having recovery strengths. Then with small sample size the criteria recovery rate was better for SECM with BIC, gMDL and eMDL holding strengths.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Generally, larger sample sizes improve the ability to make correct inferences about the true asymmetric price transmission model. This research notes that an additional factor that may influence the performance of the model selection criteria in addition to sample size is model complexity (i.e. number of asymmetric adjustment parameters included) or the number of informative variables in the model (Myung, 2000; Stine, 2003; Wagenmakers et al., 2006; Acquah, 2010; Markon and Krueger, 2004; Acquah, 2017; etc.).

Monte Carlo Simulation of the Effects of Noise levels on Model Recovery-CECM

Comparison of the different selection methods in Figure 8 shows a general trend in which recovery rates decreased with increasing error sizes. In effect, the performance of the model selection algorithms in recovering the true model deteriorates with increasing amount of noise in the true asymmetric price transmission data generating process (CECM). Alternatively, the risk of selecting the false asymmetric model increased at higher noise levels.

Findings in this research echo the results of previous studies by Jiao et al., 2011 which emphasized the fact that not only did all criteria do better at lower noise but gMDL outperformed AIC (see also Acquah, 2010 and 2017 where criteria performed better at lower noise) with regards to the relationship between the noise levels and the recovery of the true asymmetric data generating process. Yang (2003) also found that the recovery rates of the true data generating process decreases with increasing noise levels in linear regression models and concludes by noting that selection can yield the wrong model at higher noise levels.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Another interesting discovery for this study was that whereas AIC is susceptible to performing well when noise level is increasing (noise = 2 or 3), some formulations of MDL even did better although generally the overall performance decreased with increasing noise levels. In contrast, under the same conditions, when the true data generating process was the SECM, the BIC and MDLs outperformed AIC at noise level of 1 but for CECM, AIC and MDLs outperformed BIC. Chen et al. (2007) in factorial data analysis notes the tendency of BIC to perform worse than AIC at high noise levels. Also, in a comparison of model selection methods, Yang (2003) demonstrated that AIC outperforms BIC as noise levels increased in linear regression models. In the context of comparing mixture models, Lin and Dayton (1997) found that AIC was superior to BIC when the true model was relatively complex (i.e. complexity is based on the number of parameters included). Yet all the formulations of MDL (rMDL, gMDL, nMDL and eMDL) outperformed AIC when noise level is very high ($\sigma = 3$). MDL continues to prove as a useful alternative for AIC and BIC no matter the conditions presented.

Importantly, this research notes that the performances of the model selection methods in selecting the true asymmetric price transmission model deteriorates with high levels of noise (as found in Roos et al., 2005; worked on denoising where good performance is associated with low variance). Also, the performances of some formulations of the MDLs (gMDL, nMDL and eMDL) are similar to the popular AIC when noise level increases for recovering SECM but for recovery of CECM, the MDLs outperformed both AIC and BIC at higher noise levels. Intuitively, complexity and higher noise levels deteriorates the performance of the model selection methods in an asymmetric

Digitized by Sam Jonah Library

price transmission modelling framework (but interestingly nMDL was robust to this effect for CECM).

Concurrent effect of sample size and stochastic variance on model selection-CECM

Simulating the effects of sample size and noise levels concurrently affirms that a small error and large sample improve recovery of the true asymmetric data generating process and vice versa.

Obviously the improved performance of the model selection methods can be partly attributed to the fact that the true model is complex (Wagenmakers et al., 2006, Stine, 2003; Hansen and Yu, 2001; Sund, 2001; Markon and Krueger, 2004; etc.). When the true model is complex, all the model selection methods recovered fully (100%) the true data generating process under stable conditions. However, under the same stable conditions when the true model is not complex (i.e. SECM) all the model selection methods recover at least 86 percent and at most 99.5 percent of the true data generating process.

This observation suggests that increase in the number of asymmetric adjustment parameters or variables used to model asymmetry (i.e. complexity of the true model) may have influenced the improved performance or model recovery rate. Similarly, Markon and Krueger (2004) noted that the number of variables used to model a phenomenon generally improves the ability to make correct inferences in structural equation modelling. Gagne and Dayton (2002) asserted that the performance of model selection methods improve when the true model is complex in multiple regression analysis (see others- Han et al., 2014; Fade, 2015; Harremoes and Brock, 2018; Stine 2003; Fade et al., 2011).

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Additionally, AIC has an edge over BIC but not the MDLs, in their ability to select the true asymmetric data generating process in complex models with stable conditions having a recovery rate of 100 percent. Furthermore, for unstable conditions all criteria recovered between 4 and 70 percent and nMDL recovered the strongest. In standard models BIC and MDLs (rMDL almost recovered fully) far outperformed AIC under stable conditions but under unstable conditions, although AIC and eMDL was comparable, gMDL outperformed all criteria. Overall, MDLs are doing very well and some cases (gMDL and rMDL) outperform both AIC and BIC in both the standard and complex model scenarios. This study also emphasizes that some variants of MDLs are good alternatives to both BIC and AIC under both stable and unstable conditions.

Monte Carlo Simulation of the Effects of Difference in Asymmetric Adjustment Parameters on Model Selection-CECM

This study implements a modest modification in the error correction terms of the error correction model with complex dynamics. Generally, recovery rates improved from weak to strong level of asymmetry across all model selection criteria for increasing sample size (consistent with literature when the DGP was SECM and the model selection criteria are AIC, BIC, KIC, RIC). Recovery rates of the AIC, for the moderate sample size (n = 150), responded better to increases in the difference between the asymmetric adjustments parameters although the difference between that of SECM and CECM is very small or insignificant. The MDLs on the other hand outperformed BIC but among themselves, their performance was somewhat similar.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Similarly, Cook et al. (1999) in their study on the concept of the marginal likelihood and information criteria noted that the increases in the difference in asymmetric adjustments parameters (that is the adjustment speeds or positive and negative component of the error correction terms) from 0.25 to 0.50 have positive effects on the test for asymmetry. Subsequently, this study revealed similar trends. The ability of all criteria to recover the complex asymmetric error correction model (CECM) was higher than the recovery rate for the standard asymmetric error correction model. The asymmetric adjustment parameters for small sample sizes (n = 50), saw a general decrease in performance for all model selection criteria in selecting the true DGP. Unlike, in the moderate (n = 150) case scenario, eMDL outperformed all criteria and AIC in turn outperformed remaining MDLs and BIC for strong asymmetric level. Also, a similar pattern was observed in weak asymmetric adjustment but nMDL rather outperformed all criteria.

Thus MDLs do well (even in small samples) generally as compared to the traditional AIC and BIC irrespective of the sample size and this improvement in their ability to select the true DGP increase with increasing sample size. For all model selection criteria, a full recovery rate was achieved for large samples (n = 500 and above). Thus, for large samples, all criteria have the same chance in selecting the true DGP and the recovery rate was 100 percent when the original model was CECM. See Appendix I 3 for outputs. Thus, an increase in the level of asymmetric adjustment saw an improvement in criteria recovery rate for large samples. Although, the rate of recovery was the same when true DGP was CECM for all criteria that of SECM improved but varied in their performances across increasing sample size.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Importantly, the performance of the model selection methods in recovering the true data generating process depends on the difference in asymmetric adjustment parameters or speeds. Also, the effect of sample size is paramount in determining the recovery rates of model selection criteria in APT models although this effect is insignificant (full recovery rate for all criteria) when dealing with large sample sizes of 500 and above when the true DGP is complex.

Chapter Summary

The chapter developed R-functions for various formulations of the MDL in R-software. It uses Monte Carlo simulation to obtain repeated samples that are used to assess the ability of the ITFC to identify the true data generating process. The random normal distribution was used as the data generating process which generated the APT models for the study. The performances of the models were examined for all ITFC using samples of size 50, 150 and 500 with 1000 iterations.

It is found that among three DGP, all ITFC recorded lowest average values for SECM if the DGP is SECM. Over all DGP, the MDLs performed much better than the AIC and BIC. Criteria values for SECM are versatile to underlying DGP. However, if the underlying DGP is CECM, all ITFC are able to select it 100% of the time. If the DGP is SECM, the percentage of correct identification is generally above 90% and all ITFC surely detect the Houck’s model poorest.

Assessment of model selection performance is also determined based on the percentage samples (1000) for which the model fits better than any

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

competing model. Generally, for all ITFC, recovery rate is higher in large samples. Traditional criteria performed equally well as the MDLs irrespective of the sample size. In large samples, the rMDL is particularly highest in performance if the underlying DGP is SECM. If the underlying DGP is CECM, then for large samples, all criteria surely identifies the model. In this case, performance is poor for small samples for all IFTC. Assessment of model selection performance was studied along three dimensions of average criteria values, the percentage of correct selection of the underlying DGP and recovery rate. The recovery rate was assessed across conditions of effect of varying samples size, stochastics variance, combined effect of sample size and noise and difference in asymmetric adjustment parameter on model selection.

Generally, recovery rate is constrained by larger error in DGP for all criteria. No single ITFC is consistent in performance in model recovery in the presence of error (noise). However, over small error (1.0), the BIC performs the least whilst AIC was most reliable with presence of large noise (3.0). The study also assessed the model recovery in the presence of noise levels for all criteria with a specific underlying DGP. With CECM as DGP, all ITFC performed abysmally in the presence of substantial noise. For large noise (2 and 3) of noise, the BIC was most affected in performance with AIC the best performing. However, the nMDL is most reliable under moderate to large noise.

A more complex model with combined effect of sample size and level of noise has been studied. The combined effect is categorized under stable (large sample with small error) and unstable (small sample with moderately large error). It is found that under unstable condition for complex models,

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

nMDL is most reliable whilst all others perform rather abysmally. Generally, under these conditions, performances of selection criteria are poor. Under stable conditions, all criteria fully records 100% recovery. Thus, high recovery for complex models can be ensured under large sample size and small error, for any criteria. Effect of difference in Asymmetric Adjustment parameter was not too visible for all criteria for both weak and strong asymmetry under moderately large samples. There was even full recovery for all criteria under complex models. However, the AIC is most reliable (97%) under this condition although all the other criteria also performed quite well (>82%).

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Overview

This chapter summarizes all the various chapters and conclusions are drawn from the findings of the research. Based on the conclusions, relevant recommendations were provided.

Summary

This study, in general, sought to introduce and demonstrate the usefulness of the MDL principle in model selection for econometric models of asymmetric price transmission. This involved the evaluation of the performance of the formulations of MDL in comparison to AIC and BIC across sample sizes, stochastic variance and difference in asymmetric adjustment parameters.

An extensive review of literature showed the emergence of a less known but effective and competitive information criteria; the Minimum Description Length (MDL), was making great strides in model selection in the midst of already existing and widely used criteria (AIC and BIC). The MDL has seen some work in structural equation modelling, latent class models, method of inductive interference, regression analysis, etc. Several studies have concluded that neither AIC or BIC is completely satisfactory and that other fits exhibit greater power to select the true models. Interestingly, Monte Carlo evidence suggested that early MDL approximations, essentially corrected forms of BIC, perform well in selecting latent class models. Hence, this

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

research sought to introduce and demonstrate the usefulness of MDL in APT modelling framework and evaluate its performance against the popular AIC and BIC.

Subsequently, the underlying concepts that guided this study, namely information theoretic fit criteria, linear models and asymmetric price transmission was introduced and explained. More importantly, the concept behind the MDL principle was explained and emphasis put on how the algorithms were developed in connection with the price models under consideration using Monte Carlo simulation.

Basically, the R-functions for the various formulations of MDL were developed since no such functions have been built in R yet. This enabled its application to the various formulations of APT models which have been in use in econometric modelling over the years. Further, artificial data were used to simulate the behaviour of the models to find out their long term behaviour in recovering the true data generating process of our APT econometric models. Subsequently, the selected econometric models of asymmetric price transmission were submitted to MDL model selection technique and their performance was evaluated and compared to the traditional AIC and BIC in identifying the true data generating process (CECM and SECM).

This helped to recommend the best (MDL) criteria in selecting the best APT linear model under various conditions of varying sample size, stochastic variance and difference in asymmetric adjustment parameters.

Finally, the findings of our 1000 Monte Carlo simulations on the ability of all model selection criteria to recover the DGP, especially MDL, were discussed and conclusions drawn based on the objectives of this study.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Conclusions

In the asymmetric price transmission framework, the present study has empirically developed and implemented the Minimum Description Length (MDL) principle to Asymmetric Price Transmission (APT) linear models and evaluated the relative performance of the model selection algorithms of which little is known. The MDLs, (rMDL, gMDL, nMDL and eMDL) were developed using R-functions (built from scratch) and all data analysis and Monte Carlo simulations were also performed using R programming language.

The study showed that, for the selected econometric models of Asymmetric Price Transmission (APT) that is, Complex Error Correction Model (CECM) and Standard Error Correction Model (SECM), overall, the MDL criteria generally do point to the true asymmetric data generating process. Comparatively, some variants of MDL (rMDL, nMDL) criteria generally outperformed, or were comparable (depending on which condition was being examined) to the traditional AIC and BIC in recovering the true data generating process (DGP) across some sample sizes, stochastic variance and difference in asymmetric adjustment parameters. In some cases, AIC and BIC also outperformed some variants of MDLs if not all. In a general case where the average performance of all criteria was examined in their ability to select the true GDP using a sample size of 1000, all MDLs outperformed BIC and AIC in both the standard and complex case scenarios.

Alternatively, performance of all model selection criteria in their ability to recover the true DGP improved with increasing sample size. BIC outperformed all criteria in small to moderate samples whilst gMDL and eMDL outperformed AIC which in turn did better than nMDL and rMDL in

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

small samples (SECM). All MDLs in the moderate samples outperformed AIC whilst in large samples rMDL outperforms all criteria. Also, generally, the performance of the criteria was better when the true data generating process is CECM (most times full recovery) than that of the SECM. In small samples, nMDL recovered strongest followed by AIC with BIC being the least. In moderate samples, AIC rather performed better than all criteria and the MDLs in turn performed better than BIC. Finally, all criteria recorded a 100 percent recovery when the DGP is complex and sample size was large.

Regarding the effect of stochastic variance on model recovery rates, the general performance of all model selection algorithms analysed deteriorated with increasing amount of noise in the true asymmetric price transmission data generating process. This general trend was similar for both the CECM and SECM asymmetric price models but some criteria (gMDL, AIC, and eMDL) performed better under CECM and the others (rMDL, nMDL, gMDL and eMDL) performed better under SECM. The MDLs have a wide performance range, in that, depending on the true data generating process and the condition under investigation, one or more variants of the MDL can be used alongside the traditional AIC and BIC. Specifically, though BIC outperformed all criteria in recovering the standard model, rMDL can be used alongside at lower noise levels. At higher noise levels AIC was comparable to eMDL and at moderate noise levels gMDL and eMDL were comparable and outperformed all other criteria. Also, all criteria ability to recover complex models generally decreased as compared to standard models. At higher and moderate noise levels, the nMDL outperformed all criteria while at lower

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

noise levels AIC outperformed all criteria and all MDLs in turn outperformed BIC. Hence the robustness of MDLs is emphasized.

Furthermore, simulating the effect of sample size and stochastic variance concurrently affirm that stable conditions improve model recovery rate of the true asymmetric data generating process and vice versa. The MDLs continue to do well under both unstable (gMDL, eMDL, nMDL) and stable (rMDL) conditions but of course rMDL outperformed AIC and BIC under stable and gMDL outperformed AIC and BIC under unstable conditions in the SECM and nMDL under unstable for the CECM. Relatively speaking, some variants of MDL can be used as alternatives or used alongside AIC and BIC given the same conditions under study.

Subsequently, an increase in the difference in the asymmetric adjustment parameters from weak (0.25) to strong (0.5) culminates in an increase in model recovery of the true asymmetric data generating process (CECM or SECM) across all sample sizes. More specifically, the performances improved for the SECM steadily but CECM recorded a full recovery rate when the sample size was large. In small samples, BIC which was comparable to gMDL selected the standard model strongest whilst eMDL selected the complex models strongest. In moderate samples, BIC again outperformed all criteria for standard models while AIC outperformed all criteria for complex models. Lastly, in large samples, rMDL was the only criteria that fully recovered (approximately) the standard model when the asymmetric adjustment level was increased while all criteria fully recovered the complex model respectively even in weak and strong levels of asymmetry.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

The major contribution to knowledge in this study is the development of the minimum description length using R-functions derived from scratch and its evaluation of econometric models of asymmetric price transmission for the first time ever. The formulations of the MDL principles were compared to the traditional AIC and BIC. Notable under all study conditions the MDLs have proved to be very useful in model selection and model recovery and should be used as alternatives to both AIC and BIC. This is because though comparable among themselves and within their own strength, some variants sometimes outperformed these traditional and popular information criteria within the asymmetric price transmission linear models.

Recommendations

The MDL criteria (rMDL, nMDL, gMDL and eMDL), first of all, are very useful (robust, wide performance range, no distributional assumption required) in the model selection especially with issues dealing with econometric price modelling. The performance of rMDL is consistent with standard error correction model of APT while nMDL is consistent with complex error correction models of APT therefore, they can be a good substitute to BIC and AIC, respectively. The rest of the formulations of MDL (like predictive MDL) not discussed may be further studied.

Importantly, the procedures implemented in this study may apply in a broader sense to a wide range of applications within agricultural sciences, mathematical or statistical sciences, etc., when the researcher is confronted with a problem of model selection and comparison, rate of recovery of the true

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

data generating process and model performance under standard and complex conditions in asymmetric price transmission modelling framework.

In matters of choosing appropriate samples for research in the light of financial constraint, time, and non-availability of other resources, a moderate sample size of 150 (or better) is adequate in making sound decisions on asymmetric price models. Researchers should also aim at higher asymmetric adjustment levels, lower noise levels and stable conditions to achieve the best results in asymmetric price transmission linear modelling.

Probably in the future, a study will be conducted to examine MDL's performance regarding other distributions as well as to buttress the criteria's wide range of application.

Digitized by Sam Jonah Library

© University of Cape Coast   https://erl.ucc.edu.gh/jspui

REFERENCES

Abdulai, A. (2002). Using Threshold Cointegration to estimate asymmetric Price Transmission in the Swiss Pork Market. *Applied Economics* 34, 679-687.

Acquah H. D. (2010). Asymmetry in retail-wholesale price transmission for maize: evidence from Ghana. *American-Eurasian J. Agric. & Environ.Sci.* 7(4), 453-456.

Acquah, H. D., & Von Cramon-Taubadel, S. (2009). A Monte Carlo Comparison of the Power of the Test for Non-Linearity in Asymmetric Price Transmission Models. *International Journal of Computational Intelligence Research and Applications.* 3(2), 281-28.

Acquah H. D. (2010). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *Journal of Development and Agricultural Economics,* 2(1), 001-006.

Acquah, H. D. (2017a). The effect of outliers on the performance of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *RJOAS,* 5(65).

Acquah, H. D. (2017b). Criteria for asymmetric price transmission model selection based on Kullback's symmetric divergence. *RJOAS*, 11(71).

Aguero, J. (2004). *Asymmetric Price Adjustments and Behaviour Under Risk: Evidence from Peruvian Agricultural Markets*. American Agricultural Economics Association. Denver, Colorado.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki(eds.), Proceedings of the Second International Symposium on Information Theory. *Budapest: Akademiai Kiado*, 267–281.

Akaike, H. (1974). A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control*, AC-19, KO.6, December.

Alpaydin E. (2004). *Introduction to machine learning*. MIT Press, London.

Appel, V. (1992). Asymmetrie in der Preistransmission. *Agrarwirtschaft Sonderheft*, 135,178-213.

Arnoldsson, G. (2011). Arnoldsson's Conjecture, unpublished work.

Bailey, D., & Brorsen, B.W. (1989). Price Asymmetry in Spatial Fed Cattle Market. *Western Journal of Agricultural Economics*, 14 (2), 246-252.

Ball, L., & Mankim, N.G. (1994). Asymmetric Price Adjustments and Economic Fluctuations. *The Economic Journal*, 104, 247-261.

Balke, N.S, & Fomby, T. B. (1997). Threshold Co-integration. *International Economic Review*, 38, 627-645.

Balke, N.S., Brown, S.P.A., & Yücel, M.K. (1998). Crude Oil and Gasoline Prices: An Asymmetric Relationship. Federal Reserve Bank of Dallas, *Economic Review*, First Quarter, 2-11.

Barro, R. J. (1972). A Theory of Monopolistic Price Adjustment. *Review of Economic Studies*, 39, 17- 26.

Barron, A., Rissanen, J., & Yu, B (1998). The minimum description length principle in coding and modelling, *IEEE Transactions on Information Theory*, *44*, 2743–2760.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Barron, A. R., & Cover, T. M. (1991). Minimum complexity density estimation. *IEEE Trans. Inform. Theory*, 37, 1034–1054.

Bickel, P., & Zhang, P. (1992). Variable selection in nonparametric regression with categorical covariates. *J. Am. Stat. Assoc*, 87, 90–97.

Bornat, R. (2008). Engineering and information science. Retrieved from http://www.eis.mdx.ac.uk/staffpages/r_bornat/books/compiling.pdf

Borenstein, S., Cameron, A.C. and Gilbert, R. (1997). DO Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes? *Quarterly Journal of Economics*, 112, 305- 339.

Boyd, M. S., & Brorsen, B. W. (1998). Price Asymmetry in the US Pork Marketing Channel. *North Central Journal of Agricultural Economics* 10, 103- 109.

Burgin, M. (2007), *Machines, Computations, and Universality*, 5th International Conference, MCU 2007, Orléans, France, Proceedings, 24-38, Springer.

Burhnam, K. P., & Anderson, D. R. (1998). *Model selection and inference: a practical information-theoretic approach*, New York: Springer.

Breiman L. (2001). Statistical Modelling: two cultures. *Statistical Science*, 16(3), 199–231.

Britannica online encyclopedia (accessed November, 2016). Retrieved from http://www.britannica.com/EBchecked/topic/424706/Ockhams-razor

Capps, O., & Sherwell, P. (2007). Alternative Approaches in Detecting Asymmetry in Farm-Retail Prices Transmission of Fluid Milk. *Journal of Agribusiness*, 23(3), 313-331.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Carsey T. M., & Harden J. J. (2014). *Monte Carlo simulation and resampling methods for social science*. SAGE, Los Angeles.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *J. Classif*, 13, 195–212.

Chavas, J. P., & Mehta, A. (2004). Price Dynamics in a Vertical Sector: the Case of Butter. *American Journal of Agricultural Economics*, 86, 1078-1093.

Chen, L., Giannakouros, P., & Yang, Y. (2007). Model Combining in Factorial Data Analysis. *Journal of Statistical Planning and Inference* 137 (9), 2920-293.

Christianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press. London.

Cook, S., & Holly, S. (2002). Threshold Specification for Asymmetric Error Correction Models. *Applied Economics Letters*. 9, 711–13.

Cook, S., Holly, S., & Turner, P. (1999). The Power of Tests for Non-Linearity: The Case of Granger–Lee Asymmetry. *Economics Letters*. 62, 155–159.

Cook, S. (2003). A Sensitivity Analysis of Threshold Determination for Asymmetric Error Correction Models. *Applied Economic Letters*. 10, 611-616.

Costa, M., Richter, A., & Koivnen, V. (2012). Model order selection in sensor array response modelling. *IEEE Xplore*, 26, doi: 10.1109/ICSPCC.2011.6061746

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: Wiley.

Cover, T., & Thomas, J. (2006). *Elements of Information Technology* (Second Edition), John Wiley & Sons, Inc.

Cox, D. R., & Hinkley, D. V. (1974). *Theoretical Statistics*. 1st Ed. Chapman $ Hall /CRC: USA

Draper, D. (1995). Assessment and propagation of model uncertainty. J. *Roy. Stat. Soc.: Ser. B (Methodol.)*, 57, 45–97.

Duda, R. O., Hart P. E., & Stork D. G. (2001). *Pattern classification*. John Wiley & Sons, USA. pp 461-463.

Enders, W. (2004). *Applied Econometric Time Series*. New Jersey: John Wiley and Sons Ltd.

Enders, W., & Granger, C.W.J. (1998). Unit Root Test and Asymmetric Adjustments with an Example Using the Term Structure of Interest Rates. *Journal of Business and Economic Statistics*, 16, 304-311.

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and Error Correction: Representation, Estimation and Testing. *Econometrica*, 55, 251-276.

Fade, J. (2015). Stochastic complexity-based model selection with false alarm rate control in optical spectroscopy. *Pattern Rocognition Letters*, 65, 152-156.

Fade, J., Lefebvre, S., & Cezard, N. (2011). Minimum Description Length approach for unsupervised spectral unmixing of multiple interfering gas species. *Opt. Express*, 19, 13862-13872.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

FAO (2003). *Market Integration and Price Transmission in Selected Food and Cash Crop Markets of developing Countries:* Review and Applications, by Rapsomanikis, G., Hallam, D., and P. Conforti, in Commodity Market Review 2003-2004. Commodities and Trade Division, FAO, Rome.

Fishler, E., Grosmann, M., & Messer, H. (2002). Detection of signals by information theoretic criteria: general asymptotic performance analysis. *IEEE Trans. Signal Process*, 50, 1027–1036.

Forster, D. P., & Stine, R. A. (1999). Local asymptotic coding. *IEEE Trans. On Info. Theory, 45,* 1289 – 1293.

Fowler, P., & Lindblad, P. (2011). The Minimum Description Length Principle in model selection (An evaluation of the renormalized maximum likelihood criterion in linear and logistic regression analysis. Original manuscript (thesis).

Frost, D., & Bowden, R. (1999). An asymmetry Generator for Error Correction Mechanism, with Application to Bank Mortgage- Rate Dynamics. *Journal of Business and Economic Statistics*, 17(2), 253-263.

Fujikoshi, Y., & Satoh, K. (1997). Modified AIC and Cp in multivariate linear regression. *Biometrika*, 84, 707–716.

GAO (1993). *Energy security and policy: Analysis of the pricing of crude oil and petroleum products: Report by General Accounting Office.* GAO, Washington, DC.

Gagne, P., & Dayton, C.M. (2002). Best Regression Model Using Information Criteria. *Journal of Modern Applied Statistical Method*, 1, 497-488.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Gheissari, N., & Bab-Hadiasher, A. (2004). Effect of Noise on Model Selection Criteria in Visual Applications. *Pattern Recognition* 2(23-26), 229-232.

Gauthier, W. M., & Zapata, H. (2001). Testing Symmetry in Price Transmission Models, Louisiana State University, Department of Agricultural Economics and Agribusiness, Working Paper.

Goodwin, B.K., & Holt, M.T. (1999). Asymmetric Adjustment and Price Transmission in the US Beef Sector. *American Journal of Agricultural Economics*, 81, 630-637.

Goodwin, B. K., & Harper, D. C. (2000). Price Transmission, Threshold Behaviour and Asymmetric Adjustment in the U.S. Pork Sector. *Journal of Agricultural and Applied Economics*, 32, 543-553.

Goodwin, B. K., & Piggott, N.E. (2001). Spatial Market Integration in the Presence of Threshold Effects. *American Journal of Agricultural Economics* 83(2), 302- 317.

Goodwin, B.K., & Serra, T. (2003). Price transmission and asymmetric adjustment in the Spanish dairy sector. *Applied Economics*, 35, 1889–1899.

Granger, C.W. J., & Newbold, P. (1974). Spurious Regressions in Econometrics. *Journal of Econometrics*, 2, 111-120.

Granger, C.W. J., & Lee, T.H. (1989). Investigation of Production, Sales and Inventory Relationships using Multicointegration and non-symmetric Error Correction Models. *Journal of Applied Econometrics*, 4, 135- 159.

Grünwald, P. D. (2007). *The Minimum Description Length principle*, Massachusetts Institute of Technology.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Haddadi, F., Malek-Mohammadi, M., Nayebi, M. M., & Aref, M. R. (2010). Statistical performance analysis of MDL source enumeration in array processing. *IEEE Transactions on Signal Processing*, 58(1), 452-457.

Hahn, W.F. (1990). Price Transmission Asymmetry in Pork and Beef Markets. *The Journal of Agricultural Economics Research,* 42(4), 21-30.

Han, M., Wang, Z., & Liu, Y. (2014). Selecting one dependency estimators in Bayesian Network using different MDL scores and overfitting criterion. *Journal of Information Science and Engineering*, 30, 371-385.

Hansen, M., & Yu, B. (2001), Minimum Description Length Model Selection Criteria for Generalized Linear Models, *Journal of the American Statistical Association*, 96(454), 746-774.

Hansen, B.E., & Seo, B. (2002). Testing for Two-Regime Threshold Cointegration in Vector Error Correction Models. *Journal of Econometrics,* 110, 293-318.

Hansmire, M. R., & Willet, L. S. (1992). *Price Transmission Processes: A Study of Price Lags and Asymmetric Price Response Behaviour for New York Red Delicious and McIntosh Apples.* Cornell University.

Harremoes, P., & Brock, N. (2018). Horizont Independent MDL. Retrieved from https://doi.org/10.3929/ethz-b-000245070

Heien, D.M. (1980). Markup Pricing in a Dynamic Model of Food Industry. *American Journal of Agricultural Economics,* 62, pp. 10-18.

Holly, S., Turner P., & Weeks, M. (2003). Asymmetric Adjustment and Bias in Estimation of an Equilibrium Relationship from a Co-integrating Regression. *Computational Economics,* 21, 195-202.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Houck, J.P. (1977). An Approach to Specifying and Estimating Nonreversible Functions. *American Journal of Agricultural Economics*, 59, 570-572.

Hui, M., Li J., Wen, X., Yao, L., & Long, Z. (2011). An Empirical Comparison of Information-Theoretic Criteria in Estimating the Number of Independent Components of fMRI Data. retrived from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3246467/

Hurvich, C. M., & Tsai, C-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76, 297–307.

Hurvich, C. M., & Tsai, C-L. (1990). The impact of model selection on inference in linear regression. *Am. Stat.* 44, 214–217.

Ichikawa, M. (1988). Empirical assessment of AIC procedure for model selection in factor analysis. *Behaviormetrika*, 24, 33 – 40.

Ichikawa, M., & Konishi, S. (1999). Model evaluation and information criteria in covariance structure analysis. *Brit. J. Math. Stat. Psychol*, 52, 285–302.

Jerome, R., Budemeyer, A., & Diederich, A. (2014). *Neuroeconomic: Estimation and Testing of Computational Psychological Models.* Elvsevier. Retrieved from https://doi.org/10.1016/B978-0-12-416008-8.00004-8

Jiao Y., Huang, J., He, C., & Wang, J. (2011). Source number estimator using the peak-to-average power ratio modified by Gerschgorin radii. *IEEE Xplore*. doi: 10.1109/ICSPCC.2011.6061746.

Kraft, L. (1949). A *device for quantizing, grouping and coding amplitude-modulated pulse*. Massachusetts Institute of Technology. USA.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Kinnucan, H.W., & Forker, O.D. (1987). Asymmetry in Farm-Retail Price Transmission for major Dairy Products. *American Journal of Agricultural Economics*, 69, 285-292.

Kuran, T. (1983). Asymmetric Price Rigidity and Inflationary Bias. *The American Economic Review*, 73(3), 373-382.

Li, M., & Vitányi, P. (1997), *An Introduction to Kolmogorov Complexity and Its Applications* (Second Edition ed.), Springer-Verlag.

Li, L., Yao, F., & Craiu, R. V. (2012). Model selection via Minimum Description Length. Retrieved from http://hdl.handle.net/1807/31834

Lin, T. H., & Dayton, C. M. (1997). Model selection information criteria for non-nested latent class models. *J. Educat. Behav. Stat.*, 22, 249–264.

Lloyd, T., McCorriston, S., Morgan, C. W., & Rayner, A. J. (2003). The Impact of Food Scares on Price Transmission in Inter-Related Markets. Paper presented to the *XXVth IAAE Conference in Durban/South Africa*.

Mackay, J. C. (2003). *Inference Theory, Inference and Learning Algorithms.* Angus and Robertson, Cambridge University Press. Cambridge, UK.

Manera, M., & Frey, G. (2007). Econometric Models of Asymmetric Price Transmission. *Journal of Economic Survey*, 67, 349- 415.

Markon, K. E., & Krueger, R. F. (2004). An Empirical Comparison of Information- Theoretic Selection Criteria for Multivariate Behavior Genetic Models. *Behavior Genetics*, 34 (6), 593- 609.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Meyer, J. (2003). Measuring Market Integration in the Presence of Transaction costs – A Threshold Vector Error Correction Approach. Contributed paper, XXV International Conference of Agricultural Economists, Durban, South Africa.

Meyer, J., & Von Cramon-Taubadel, S. (2004). Asymmetric Price Transmission: A survey. Journal of Agricultural Economics, 55 (3), 581-611.

McCorriston, S. (2002). Why Should Imperfect Competition Matter to Agricultural economists? *European Review of Agricultural Economics,* 29(3), 49-371.

Miller, D. J., & Hayenga, M.L. (2001). Price Cycles and Asymmetric Price Transmission in the U.S. Pork Market. *American Journal of Agricultural Economics*, 83, 551-562.

Mitchell, T. (1997). *Machine Learning.* Retrieved from www.academia.edu/32214523/Machine_Learning_-_Tom_Mitchell

Mohanty, S., Peterson, E. W. F., & Kruse, N. C. (1995). Price Asymmetry in the International Wheat Market. *Canadian Journal of Agricultural Economics*, 43, pp. 355-366.

Mohammed, E. A., Naugler, C., & Far, B. H. (2015). *Emerging Trends in Computational Biology, Bioinformatics and Systems Biology.* Elsevier https://doi.org/10.1016/B978-0-12-802508-6.00032-6

Myung, I. J., Balasubramanian, V., & Pitt, M. A. (2000). Counting probability distributions: Differential geometry and model selection. *Proc. Natl. Acad. Sci.* 97, 11170–11175.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Myung, J. I. (2000). The Importance of Complexity in Model Selection. *Journal of Mathematical Psychology*, 44, 190-204.

Myung, J. I., Navarro, D. J., & Pitt, M. A. (2005). Model selection by normalized maximum likelihood. *Journal of Mathematical Psychology*. 50(2006), 167-179.

Nelson, M. (2014). Data compression with arithmetic encoding. Retrieved from http://www.drdobbs.com/cpp/data-compression-with-arithmetic-encodin/240169251

Pauler, D. K. (1998). Schwarz Criterion and Related Methods for Normal Linear Models. *Biometrika*, 85, 13–27.

Peltzman, S. (2000). Prices Rise Faster Than They Fall. *Journal of Political Economy*, 108(3), 466-502.

Pick, D. H., Karrenbrock, J. D., & Carman, H. F. (1990). "Price Asymmetry and Marketing Margin Behaviour: an Example for California - Arizona Citrus". *Agribusiness*, 6(1), 75-84.

Pitt, M. A., Myung, J. I., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, 109(3).

Poland, J., & Hutter M. (2005). Asymptotic of discrete MDL for online prediction. IEEE Transactions on Information Theory, 51 (11), 3780 – 3795.

Raftery, A. E. (1993). Bayesian model selection in structural equation models. In K. A. Bollen and J. S. Long (eds.), *Testing structural equation models*. Newbury Park, CA: Sage, 163–180.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Rissanen, J. (1978). Modelling by shortest data description. *Automatica*, 14, 465–471.

Rissanen, J. (1983). A Universal Prior for Integers and Estimation by Minimum Description Length, *Annals of Statistics*, 11(2), 416-431.

Rissanen, J. (1986). Stochastic Complexity and Modelling. *Annals of Statistics,* 14(3), 1080-1100.

Rissanen, J. (1989). *Stochastic complexity and statistical inquiry*. Singapore: World Scientific.

Rissanen, J. (1996). Fisher information and stochastic complexity. *IEEE Trans. Inform. Theory* 42, 40–47.

Rissanen, J. (2001). *Lectures on Statistical Modelling Theory: Lecture notes*. Department of Computer Science. University of Helsinki.

Rissanen, J. (2001). Strong optimality of the normalized ML models as universal codes and information in data. *IEEE Trans. Inform. Theory* 47, 1712–1717.

Rissanen, J., & Tabus, I. (2005). Kolmongorov's structure function in MDL theory and lossy data compression. In P. D. Grünwald, I. J. Myung, and M. A. Pitt (Eds), *Advances in Minimum Description Length: Theory and Applications*. Cambridge, MA:MIT Press.

Rissanen, J. (2007), *Information and Complexity in Statistical Modelling*, Springer.

Rissanen, J., Roos, T., & Myllymäki, P, (2010). Model selection by sequentially normalized least squares, *Journal of Multivariate Analysis*, 101(4), 839-849.

Roos, T. (2004). MDL regression and denoising. Unpublished manuscript.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Roos, T. (2009a). Lecture 7: Information –theoretic modelling. Retrived from http://video.helsinki.fi/tero/flashvideo.php?m=tktl&id=itm2909.mp4

Roos, T. (2009b). Lecture 8: Information –theoretic modelling. Retrieved from http://video.helsinki.fi/tero/flashvideo.php?m=tktl&id=itm0210.mp4

Roos, T., Wettig, H., Grünwald, P., Myllymäki, P., & Tirri, H. (2005). On Discriminative Bayesian Network Classifiers and Logistic Regression, *Machine Learning*, 59(3), 267-296.

Roos, T., Myllymaki, P., & Tirri, H. (2005). On the behaviour of MDL Denoising, *AISTATS*. Retrived from http://cosco.hiit.fi/Articles/aistats05.pdf

Sewell, M. (2007). Model selection. Retrieved from https://pdfs.semanticscholar.org/b449/82b8446bef24d1b24104e50a7c84b10030e0.pdf

Scholnick, B. (1996). Asymmetric Adjustment of Commercial Bank Interest Rates: Evidence from Malaysia and Singapore. *Journal of International Money and Finance*, 15, 485-496.

Schwarz, G. (1978). Estimating the Dimension of a Model, *The Annals of Statistics*, 6(2), 461-464.

Shannon, C. (1948). A Mathematical Theory of Communication, *The Bell System Technical Journal*, 27, 379–423, 623–656

Shin Y., Yu B., & Greenwood-Nimmo M. (2014). Modelling Asymmetric Cointegration and Dynamic Multipliers in a Nonlinear ARDL Framework. In: Sickles R., Horrace W. (eds) Festschrift in Honor of Peter Schmidt. *Springer*, New York, NY.

Sipser, M. (2006). *Introduction to the theory of computation.* (second edition,

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

International edition ed.), Thomson Course Technology.

Solomonoff, R. (1964). A Formal Theory of Inductive Inference, Part I, *Information and Control*, 7, 1-22.

Stine, R. A. (2003). Model selection using information theory and the MDL principle. Retrieved from https://www-stat.wharton.upenn.edu/~stine/

Stone, M. (1979). Comments on Model Selection Criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society*, 41, 276–278. Ser B.

Sugiura, N. (1978). "Further Analysis of the Data by Akaike's Information Criterion and the Finite Corrections," *Communications in Statistics*, A7, 13–26.

Sund, R. (2001). Minimum Description Length based model selection in linear regression. Retrieved from www.helsinki.fi/~sund/pdf/sund_mdl.pdf

Takeuchi, K. (1976). Distribution of information statistics and criteria for adequacy of models. *Math. Sci*. 153, 12–18.

Tappata, M. (2008). Rockets and Feathers. Understanding Asymmetric Pricing. *American Economic Review*, 79(4), 700–712.

Tong, H. (1983). *Threshold Models in Non-Linear Time Series Analysis*. New York, Springer Verlag.

Tsay, R. S. (1989). Testing and Modelling Threshold Autoregressive Processes. *Journal of the American Statistical Association*, 84, 231-240.

Turing, A. (1937). On Computable Numbers With an Application to the Entscheidungsproblem, *Proceedings London Mathematical Society*, 42, 230–265.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Tweeten, L.G., & Quance, C.L. (1969). Positivistic Measures of Aggregate Supply Elasticities: Some new Approaches. *American Journal of Agricultural Economics*, 51, 342-352.

Velampalli S., & Murthy Jonnalagedda V.R. (2017). Minimum Description Length (MDL) Based Graph Analytics. In: Satapathy S., Prasad V., Rani B., Udgata S., Raju K. (eds). Proceedings of the First International Conference on Computational Intelligence and Informatics. *Advances in Intelligent Systems and Computing*, 507, *Springer*, Singapore.

Von. Cramon-Taubadel, S., & Fahlbusch, S. (1994). Identifying Asymmetric Price Transmission with Error Correction models. Poster Session EAAE European Seminar in Reading. *Economics*, 25, 1-18.

Von. Cramon-Taubadel, S., & Loy, J. P. (1996). Price Asymmetry in the International Wheat Market: Comment. *Canadian Journal of Agricultural Economics*, 44, 311-317.

Von. Cramon-Taubadel, S. (1996). An Investigation of Non-Linearity in Error Correction Representations of Agricultural Price Transmission. Contributed Paper. *VIII Congress of the European Association of Agricultural Economists*, Edinburgh.

Von. Cramon-Taubadel, S. (1998). Estimating Asymmetric Price Transmission with the Error Correction Representation: An Application to the German Pork Market". *European Review of Agricultural Economics*, 25, 1-18.

Von. Cramon-Taubadel, S., & Loy, J. P. (1999). The Identification of Asymmetric Price Transmission Processes with Integrated Time Series. *Jahrbücher für Nationalökonomie und Statistik*, 218(1-2), 85-106.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Von. Cramon-Taubadel, S., & Meyer, J. (2000). Asymmetric Price Transmission: Fact or Artefact? University Göttingen, Institute for Agricultural Economics, Working Paper.

Wagenmakers, E., Grunwald, P., & Steyvers, M. (2006). Accumulative prediction error and the selection of time series models. Journal of Mathematical Psychology, 50(2006).

Wallace, C. S., & Boulton, D. M. (1968). An information measure for classification. *Computer Journal*, 11, 185-194.

Ward, R.W. (1982). Asymmetry in Retail, Wholesale and Shipping Point Pricing for Fresh Vegetables. *American Journal of Agricultural Economics*, 62, 205-212.

Wlazlowski, S. S. (2003). Petrol and crude oil prices: Asymmetric price transmission. *Ekonomia / Economics*, 11, 1-25.

Wolffram, R. (1971). Positivistic Measures of Aggregate Supply Elasticities: Some New Approaches, Some Critical Notes. *American Journal of Agricultural Economics*, 53, 356-356.

Wolffram, R. (2005). Estimation of Short- and Long-term Irreversible Relations, Institute of Agricultural Policy working paper, University of Bonn.

Yang, Y. (2003). Regression with Multiple Candidate Models: Selecting or Mixing? *Statistica Sinica* 13, 783-809.

Yang, Y., & Barron, A. (1998). An asymptotic property of model selection criteria. *IEEE Transactions on information theory*. 44, 117 – 133.

Zhang, P. (1993). On the convergence of model selection criteria. *Commun. Stat.—Theory Meth*. 22, 2765–2775.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Zhang, P., Fletcher, S., & Carley, D (1995). Peanut Price Transmission Asymmetry in Peanut Butter. *Agribusiness*, 11(1), 13-20.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

APPENDICES

APPENDIX A

READING A TREE DIAGRAM

Reading a tree diagram

The easiest way to visualize how a code is formed is by constructing a so-called tree diagram. Consider the code (a = 0, b = 10, c = 11). It is presented in a tree diagram (Figure 1).

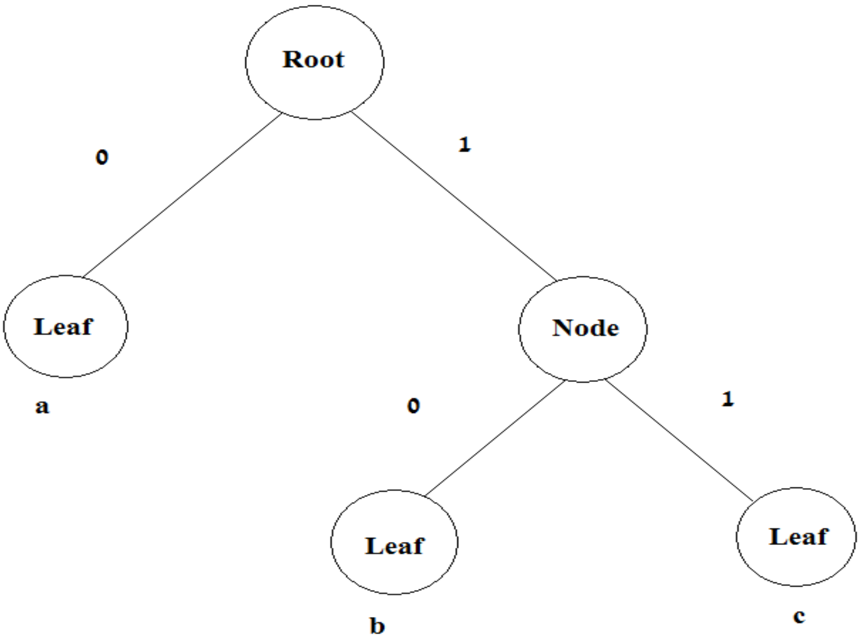


Figure 1: A tree diagram over a prefix-free code for (a, b, c)

The start of a tree diagram is called "the root". From the root two paths run, both ending up in a so-called node. A node can be two things. Either it is a junction that splits the path in two, or it is an endpoint of the path that leads down to it. In the latter case, the node is called a "leaf". It should be noted that all leaves are nodes, though not all nodes are leaves. The process of splitting paths in two is continued until there is as many leaves in the tree as there are symbols, x_i , in the code. Once this is accomplished, each symbol is assigned to a unique leaf. Each path below the root or a node is given the symbol 0 or 1, making sure that for each pair of paths, 0 or 1 appears no more than once.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

The code word assigned to each symbol x_i can then be read by tracing the paths from the root, down through all the nodes leading to the leaf corresponding to the code word of interest. By reading the junction of the binary symbols at each node and joining them together in chronological order, one gets the code word for symbol x_i.

For example, consider the symbol "b" in figure 1. If one is interested in finding out which code word is assigned to it, one starts at the root and work one‟s way downwards to leaf "b". One starts by going right and thus add a 1 to the code. At the first node, one turns left and adds a 0 to the code. It can now be seen that the final node, or leaf, has been reached, and the process halts. Thus the code is 10 for "b", which is in agreement with what was seen earlier.

Note that it does not matter how the 0‟s and 1‟s are assigned to the various paths in the tree diagram, one would still end up with a prefix-free code. It is common practice, however, to be consequent in assigning all zeros to either the left or the right path connected to each node.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

APPENDIX B

HUFFMAN'S ALGORITHM

Huffman's algorithm

Huffman‟s algorithm works by sorting symbols, x_i , that are to be encoded after their probabilities in increasing order. The two symbols with least probabilities are then defined as leaves of a, for now, disjoint node; a root if you wish. One of the branches between that node and the leaf gets a 0 attached to it, while the other gets a 1. (Cover and Thomas, 1999. p. 118)

Next, the node is assigned a probability equal to the sum of the two symbols' respective probabilities, and then put back into the sorted list where it now is treated as an element that can be selected and grouped. The procedure is then repeated until there is only one element left. That element forms the prefix-free binary tree for the symbols. Code word for symbol x_i is then traced from the root down to the corresponding leaf as usual (Cover and Thomas, 1999; p. 118).

It can be shown that Huffman‟s algorithm results in an optimal expected code length. However, the proof is beyond the scope of this paper, but can be found in (Cover and Thomas, 2006, pp. 123-127). For a more visual representation of how the algorithm works, see (Roos, 2009a).

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

APPENDIX C

KRAFT INEQUALITY

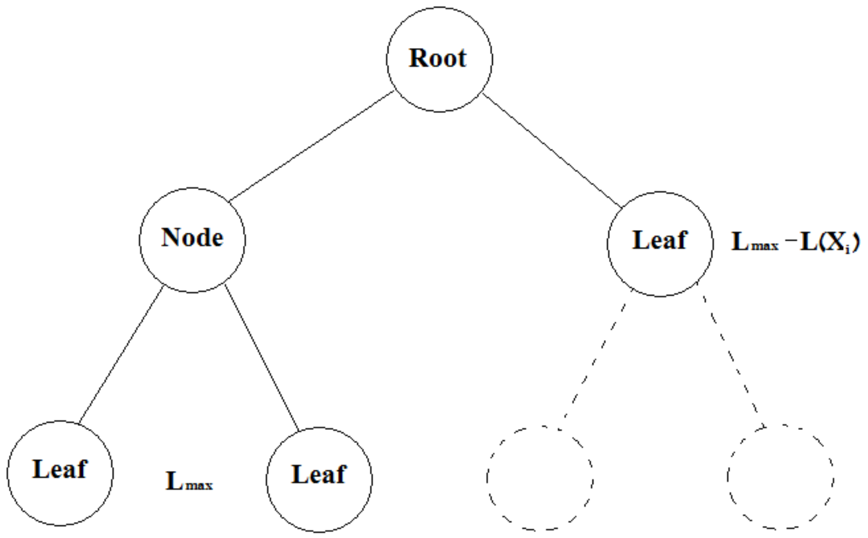
Proof of Kraft inequality

The proof is given only for the binomial case, though it can easily be extended to the general case by simply changing the base 2, to the base $|Y|$.

Assume a tree diagram representing a prefix-free code. Note that the length of each code word is equal to the number of nodes, including the leaf that connects the corresponding leaf to the root. In other words, the further down in the tree a leaf is, the longer a code word is associated to it.

The length of the longest code word can be defined as $L_{\max}$. Note that there can be at most $2^{L_{\max}}$ code words of that length, which occurs in the special case where all code words have equal length.

At each leaf that is not at $L_{\max}$ level, the number of possible code words that could have been generated from it, if the leaf had been a non-leaf node instead is, for each leaf, $2^{L_{\max}-L(x_i)}$.



Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Figure 2: An illustrative tree-diagram over the Kraft inequality

This gives us that the sum of all leaves in the tree diagram is the following:

$$\sum_{i=1}^{h} 2^{L_{\max} - L(x_i)}$$

Clearly this sum cannot be greater than in the case where all the code words are of maximum length.

$$\sum_{i=1}^{h} 2^{L_{\max} - L(x_i)} \leq 2^{L_{\max}}$$

It can then be seen that this is equivalent to

$$2^{L_{\max}} \sum_{i=1}^{h} 2^{-L(x_i)} \leq 2^{L_{\max}}$$

$$\sum_{i=1}^{h} 2^{-L(x_i)} \leq 1$$

which proves the Kraft inequality (Cover and Thomas, 2006, p. 108).

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

APPENDIX D

THE SHANNON CODE

Proof that of the Shannon code fulfils the Kraft inequality
The Kraft inequality states that

$$\sum_{i=1}^{h} 2^{-L(x_i)} \leq 1$$

$$\{L_{Shannon}(x_i) := \lceil -\log_2[P(x_i)] \rceil\}$$

It can be seen that,

$$= \sum_{i=1}^{h} 2^{-\lceil(-\log_2[p(x_i)])\rceil}$$

$$\leq \sum_{i=1}^{h} 2^{-(-\log_2[P(x_i)])}$$

$$= \sum_{i=1}^{h} 2^{\log_2[P(x_i)]}$$

$$= \sum_{i=1}^{h} P(x_i) = 1$$

$$\therefore \sum_{i=1}^{h} 2^{-L_{Shannon}(x_i)} \leq 1$$

Q.E.D

In other words, the Shannon code fulfils the Kraft inequality (Cover & Thomas, 2006, pp 112-113)

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

APPENDIX E
THE TURING MACHINE

Definition of a Turing machine

A Turing machine is a theoretical machine that, given a logical problem, either runs until it solves it and then stops, or never stops at all. (Turing, 1937, pp. 240-241) All logical problems that can be solved at all can be solved by a Turing machine. (Sipser, 2006, p. 139) This means that if a Turing machine cannot solve the problem, the problem cannot be solved with logic.

A problem with the Turing machine is that there is no way to know if the problem has a solution or not since some problems might just take a very, very long time to be computed. The problem is, of course, that one in advance cannot know if the problem is unsolvable since it is impossible to tell if the calculations ever will halt. In fact, Alan Turing proved in 1936 that it is impossible to know in advance if a problem is solvable or not. (Turing, 1937, p.247). This is known as the "Halting problem". (Sipser, 2006) and (Cover and Thomas, 2006, pp. 482-483).

A Universal Turing Machine is "[…] a single machine which can be used to compute any computable sequence." (Turing, 1937, p. 241). In other words, such a machine can do everything any specific Turing machine can do.

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

APPENDIX F

THE STIRLING APPROXIMATION

The Stirling approximation of the gamma function

The Stirling approximation states that:

$$\ln \Gamma(n+1) = \ln(n!) \approx \left(n+\frac{1}{2}\right)\ln(n) - n + \frac{1}{2}\ln(2\pi)$$

(Roos 2004, p.8)
A quick look at the Stirling approximation indicates that it is a good one, with an error very quickly approaching zero as *n* grows.

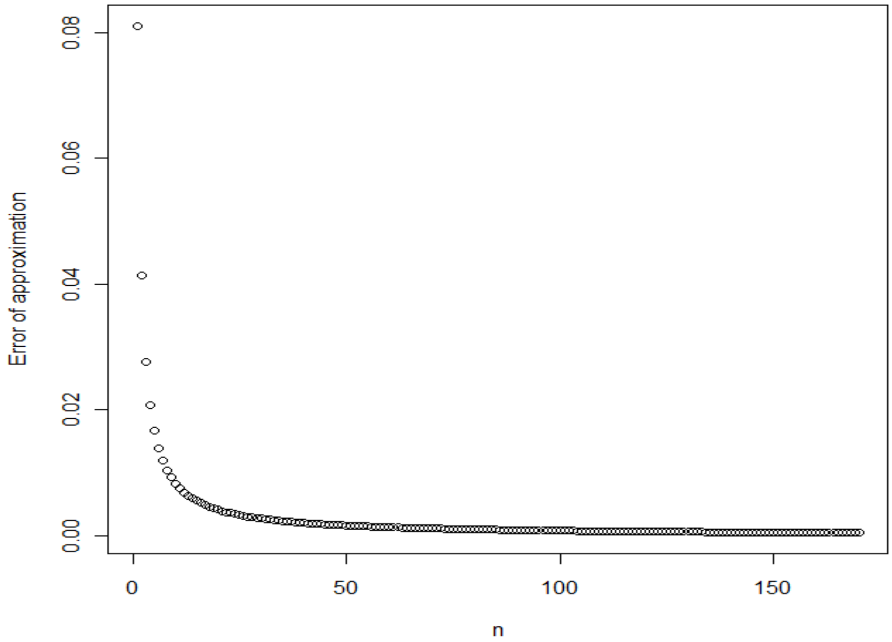


Figure 4: The error caused by the Stirling approximation for values of n=1 to 170

For the convenience of the reader, the R code used to produce figure 4 is presented below.

R code for the illustration of the Stirling approximation

```
# The following script generates the Stirling approximation of
# ln(n!)
# It is used by typing "Stirling(n)" for any given positive value of n.
Stirling <- function(n) { (n+1/2)*log(n)-n+(1/2)*log(2*pi) }
y <- c(1:170)
# The reason that we let y range from 1 to 170 is that 170
# is that 170 is the highest number that R lets us calculate
# the factorial of, at least on the computer we used.
truval <- log(factorial(y))
appval <- Stirling(y)
plot(truval-appval, ylab="Error of approximation", xlab="n")
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

APPENDIX G

ALTERNATIVE RNML

Alternative form of RNML

$$= (n-k)\ln\hat{\sigma}^2 + k\ln\hat{R} + (n-k-1)\ln\left(\frac{1}{n-k}\right) - (k-1)\ln k$$

$$= (n-k)\ln\hat{\sigma}^2 + k\ln\hat{R} + (n-k-1)\ln\left(\frac{1}{n-k}\right) - \ln\left(\frac{1}{n-1}\right) - (k-1)\ln k$$

$$= (n-k)\ln\left(\frac{\hat{\sigma}^2}{n-k}\right) + k\ln\hat{R} - \ln\left(\frac{1}{n-k}\right) - (k-1)\ln k$$

$$= (n-k)\ln\left(\frac{\hat{\sigma}^2}{n-k}\right) + k\ln\hat{R} - \ln\left(\frac{1}{n-k}\right) - k\ln k + \ln k$$

$$= (n-k)\ln\left(\frac{\hat{\sigma}^2}{n-k}\right) + k\ln\left(\frac{\hat{R}}{k}\right) - \ln\left(\frac{1}{n-k}\right) + \ln k$$

$$= (n-k)\ln\left(\frac{\hat{\sigma}^2}{n-k}\right) + k\ln\left(\frac{\hat{R}}{k}\right) + \ln(n-k) + \ln k$$

$$= (n-k)\ln\left(\frac{\hat{\sigma}^2}{n-k}\right) + k\ln\left(\frac{\hat{R}}{k}\right) + \ln[k(n-k)]$$

Q.E.D

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

APPENDIX H

DEVELOPMENT OF CODES IN R

Appendix H1: Full version and when constants are eliminated

```
#### AIC -- Removing constants that does not affect selection
AIC_gid <- function(fit = model){
  RSS <- sum((fit$residuals)^2) # Sum of Squares residual
  n <- nrow(fit$model) # The number of data points
  log_lkl <- n*log(RSS/n)
  npar <- length(fit$coefficients)
  calc <- 2*npar + log_lkl
  return(calc)
}
AIC_gid(lm_fit1)
AIC_gid(lm_fit2)
AIC_gid(lm_fit3)
AIC_gid(lm_fit4)

#### BIC -- Removing constants that does not affect selection
BIC_gid <- function(fit=model){
  RSS <- sum((fit$residuals)^2) # Sum of Squares residual
  n <- nrow(fit$model) # The number of data points
  log_lkl <- n*log(RSS/n)
  npar <- length(fit$coefficients)
  calc <- log(n)*npar + log_lkl
  return(calc)
}

BIC_gid(lm_fit1)
BIC_gid(lm_fit2)
BIC_gid(lm_fit3)
BIC_gid(lm_fit4)

################## AIC full Versions ################
AIC_full <- function(fit=model){
  RSS <- sum((fit$residuals)^2) # Sum of Squares residual
  n <- nrow(fit$model) # The number of data points
  log_lkl <- n*(log(2*pi)+1+ log(RSS/n))
  npar <- length(fit$coefficients)
  calc <- 2*(npar+1) + log_lkl
  return(calc)
}
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

```
AIC_full(lm_fit1)
AIC_full(lm_fit2)

### R built in AIC

AIC(lm_fit1)
AIC(lm_fit2)

## We notice the formula I wrote gives the same result as R's built in AIC

### Now let's simplify by removing constants
AIC_gid <- function(fit = model){
  RSS <- sum((fit$residuals)^2) # Sum of Squares residual
  n <- nrow(fit$model) # The number of data points
  log_lkl <- n*log(RSS/n)  # Here we removed log(2*pi)+1     --------------------
----------------------
  npar <- length(fit$coefficients)
  calc <- 2*npar + log_lkl # Also removed the +1 by changing 2*(npar+1) to
2*npar ------------------
  return(calc)
}

AIC_gid(lm_fit1)
AIC_gid(lm_fit2)

# Now we realize the numbers are about ~3000 less than R's Built in AIC
# However, we will notice that when comparing, the differences are the same
AIC(lm_fit1) - AIC(lm_fit2)
AIC_gid(lm_fit1) - AIC_gid(lm_fit2)
```

Appendix H2: R – functions for the four MDL's

```
# Minimum Description Length (Rissanen's method = rMDL)

IDl_len <- function(z){
  x <- round(z)
  frst_log <- ifelse(abs(x)<= 1, 0, log2(abs(x)))
  scnd_log <- ifelse(abs(frst_log)<=1, 0, 2*log2(frst_log))
  id_len <- 2 + frst_log + scnd_log
  return(id_len)
}
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

```
### R MDL
rMDL <- function(fit = model){
  t.values <- as.numeric(summary(fit)$coefficients[,3]) # Extract t-statistic
value from the model including the intercept
  RSS <- sum((fit$residuals)^2) # Sum of Squares residual
  n <- nrow(fit$model) # The number of data points
  log_lkl <- (n/2)*log2(RSS/n)
  unv_cdlen <- sum(sapply(t.values,IDl_len)) # Calculates the universal codes
  Mdl <- log_lkl + unv_cdlen
  return(Mdl)
}


### G MDL
gMDL <- function(fit = model){
  smmry <- summary(fit)                        # Summary of the lm function
or the equation
  X <- as.matrix(fit$model[, -1])               # Matrix of predictors
  td <- data.frame(smmry$coefficients)              # matrix with coefficients
  b <- as.matrix(td[rownames(td) != "(Intercept)", 1])    # vector of coefficients
excluding intercept
  k <- nrow(b)                           # Number of parameters excluding
the intercept
  SSE <- sum((fit$residuals)^2)                  # Sum of square errors
  n <- nrow(X)                            # Number of data points used in the
model
  S <- (SSE/(n - k) )                      # mean square error or standard
error
  dep <- fit$model[ ,1]
  F <- as.numeric(( dep%*%dep - SSE)/(k*S))           # F statistic
  Rsq <- smmry$r.squared                      # R squared value

  vl <- ifelse(Rsq >= k/n,(n/2)*log(S) + (k/2)*log(F) + log(n),
(n/2)*log((dep%*%dep)/n) + .5*log(n) )
  return(vl)
}


## Normalized Maximum Likelihood MDL
nMDL <- function(fit=model){
  smmry <- summary(fit)                        # Summary of the lm function
or the equation
  X <- as.matrix(fit$model[, -1])               # Matrix of predictors
  td <- data.frame(smmry$coefficients)              # matrix with coefficients
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

```
 b <- as.matrix(td[rownames(td) != "(Intercept)", 1])    # vector of coefficients
excluding intercept
 k <- nrow(b)                                # Number of parameters excluding
the intercept
 SSE <- sum((fit$residuals)^2)               # Sum of square errors
 n <- nrow(X)                                # Number of data points used in the
model
 S <- (SSE/(n - k) )
 dep <- fit$model[ ,1]
 F <- as.numeric(( dep%*%dep - SSE)/(k*S))            # F statistic
 Rsq <- smmry$r.squared                      # R squared value

 vl1 <- (n/2)*log(S) + (k/2)*log(F) + .5*log(n-k) - 1.5*log(k)
 return(vl1)
}


## Exact Normalised Maximum Likelihood MDL
eMDL <- function(fit=model){
 smmry <- summary(fit)                       # Summary of the lm function
or the equation
 X <- as.matrix(fit$model[, -1])             # Matrix of predictors
 td <- data.frame(smmry$coefficients)        # matrix with coefficients
 b <- as.matrix(td[rownames(td) != "(Intercept)", 1])    # vector of coefficients
excluding intercept
 k <- nrow(b)                                # Number of parameters excluding
the intercept
 SSE <- sum((fit$residuals)^2)               # Sum of square errors
 n <- nrow(X)                                # Number of data points used in the
model
 S <- (SSE/(n - k) )
 dep <- fit$model[ ,1]
 F <- as.numeric(( dep%*%dep - SSE)/(k*S))            # F statistic
 Rsq <- smmry$r.squared                      # R squared value

 vl2 <- (n-k)*log(SSE/n) + k*log(as.numeric(t(b)%*%crossprod(X)%*%b)) +
(n-k-1)*log(n/(n-k)) - (k+1)*log(k)
 return(vl2)

}
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

```
## Combine gMDL, nMDL and the Exact MDL into one function. This will
be useful during simulations

MDL <- function(fit=model){
  smmry <- summary(fit)                    # Summary of the lm function
or the equation
  X <- as.matrix(fit$model[, -1])                # Matrix of predictors
  td <- data.frame(smmry$coefficients)              # matrix with coefficients
  b <- as.matrix(td[rownames(td) != "(Intercept)", 1])    # vector of coefficients
excluding intercept
  k <- nrow(b)                           # Number of parameters excluding
the intercept
  SSE <- sum((fit$residuals)^2)                   # Sum of square errors
  n <- nrow(X)                           # Number of data points used in the
model
  S <- (SSE/(n - k) )
  dep <- fit$model[ ,1]
  F <- as.numeric(( dep%*%dep - SSE)/(k*S))            # F statistic
  Rsq <- smmry$r.squared                      # R squared value
  t.values <- as.numeric(summary(fit)$coefficients[,3]) # Extract t-statistic
value from the model including the intercept
  RSS <- sum((fit$residuals)^2) # Sum of Squares residual
  n <- nrow(fit$model) # The number of data points
  log_lkl <- (n/2)*log2(RSS/n)
  unv_cdlen <- sum(sapply(t.values,IDl_len)) # Calculates the universal codes

  rMDL <- log_lkl + unv_cdlen       #Rissanen MDL
  eMDL <- (n-k)*log(SSE/n) +
k*log(as.numeric(t(b)%*%crossprod(X)%*%b)) + (n-k-1)*log(n/(n-k)) -
(k+1)*log(k)
  gMDL <- ifelse(Rsq >= k/n,(n/2)*log(S) + (k/2)*log(F) + log(n),
(n/2)*log((dep%*%dep)/n) + .5*log(n) )
  nMDL <- (n/2)*log(S) + (k/2)*log(F) + .5*log(n-k) - 1.5*log(k)

  aic <- AIC(fit)  # AIC
  bic <- BIC(fit)  # BIC
  combo <- data.frame(Rissanen.MDL = rMDL, G.Prior.MDL = gMDL,
Norm.MDL = nMDL, Exact.MDL = eMDL, AIC = aic, BIC = bic)
  return(combo)
}
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Appendix H3: R-codes for econometric asymmetric price models

```
xt <- cumsum(rnorm(n))
 et <- rnorm(n)
 yt <- xt + et

 dxt <- diff(xt)
 dxtpos <- (dxt + abs(dxt))/2
 dxtneg <- dxt - dxtpos

 reg <- lm(yt ~ xt)

 ECT<-reg$res
 ECTpos<-(ECT+abs(ECT))/2
 ECTneg<-ECT-ECTpos
 lECTpos<-(Lag(ECTpos,1))
 lECTneg<-(Lag(ECTneg,1))

 lECTpos<-lECTpos[-c(1,1)]
 lECTneg<-lECTneg[-c(1,1)]

 dyt <- 0.7*dxt + 0.25*lECTpos + 0.75*lECTneg + rnorm(n-1)

 dyy <- dyt[c(6:n-5)]
 dxxpos <- dxtpos[c(6:n-5)]
 dxxneg <- dxtneg[c(6:n-5)]

 Dyt <- dyt[c(6:n-5)]
 Dxt <- dxt[c(6:n-5)]

 LECTpos <- lECTpos[c(6:n-5)]

 LECTneg <- lECTneg[c(6:n-5)]
```

Appendix H4: Final script for analysis (changes are made per objective of study)

```
packages <- c("Hmisc")
new.packages <- packages[!(packages %in%
installed.packages()[,"Package"])]
if(length(new.packages) > 0) {install.packages(new.packages)}
```


Digitized by Sam Jonah Library

© University of Cape Coast   https://erl.ucc.edu.gh/jspui

```
for(i in packages){
  require(i, character.only = TRUE )
}

# Minimum Description Length (Rissanen's method = rMDL)

IDl_len <- function(z){
  x <- round(z)
  frst_log <- ifelse(abs(x)<= 1, 0, log2(abs(x)))
  scnd_log <- ifelse(abs(frst_log)<=1, 0, 2*log2(frst_log))
  id_len <- 2 + frst_log + scnd_log
  return(id_len)
}

### R MDL

rMDL <- function(fit = model){
  t.values  <-  as.numeric(summary(fit)$coefficients[,3])  #  Extract  t-statistic
value from the model including the intercept
  RSS <- sum((fit$residuals)^2) # Sum of Squares residual
  n <- nrow(fit$model) # The number of data points
  log_lkl <- (n/2)*log2(RSS/n)
  unv_cdlen <- sum(sapply(t.values,IDl_len)) # Calculates the universal codes
  Mdl <- log_lkl + unv_cdlen
  return(Mdl)
}

### G MDL
gMDL <- function(fit = model){
  smmry <- summary(fit)                          # Summary of the lm function
or the equation
  X <- as.matrix(fit$model[, -1])                # Matrix of predictors
  td <- data.frame(smmry$coefficients)               # matrix with coefficients
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

b <- as.matrix(td[rownames(td) != "(Intercept)", 1]) # vector of coefficients excluding intercept
k <- nrow(b) # Number of parameters excluding the intercept
SSE <- sum((fit$residuals)^2) # Sum of square errors
n <- nrow(X) # Number of data points used in the model
S <- (SSE/(n - k)) # mean square error or standard error
dep <- fit$model[,1]
F <- as.numeric((dep%*%dep - SSE)/(k*S)) # F statistic
Rsq <- smmry$r.squared # R squared value

vl <- ifelse(Rsq >= k/n,(n/2)*log(S) + (k/2)*log(F) + log(n), (n/2)*log((dep%*%dep)/n) + .5*log(n))
return(vl)
}

Normalized Maximum Likelihood MDL

nMDL <- function(fit=model){
smmry <- summary(fit) # Summary of the lm function or the equation
X <- as.matrix(fit$model[, -1]) # Matrix of predictors
td <- data.frame(smmry$coefficients) # matrix with coefficients
b <- as.matrix(td[rownames(td) != "(Intercept)", 1]) # vector of coefficients excluding intercept
k <- nrow(b) # Number of parameters excluding the intercept
SSE <- sum((fit$residuals)^2) # Sum of square errors
n <- nrow(X) # Number of data points used in the model
S <- (SSE/(n - k))
dep <- fit$model[,1]

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

```
  F <- as.numeric(( dep%*%dep - SSE)/(k*S))            # F statistic
  Rsq <- smmry$r.squared                               # R squared value

  vl1 <- (n/2)*log(S) + (k/2)*log(F) + .5*log(n-k) - 1.5*log(k)
  return(vl1)
}

## Exact Normalised Maximum Likelihood MDL

eMDL <- function(fit=model){
  smmry <- summary(fit)                                # Summary of the lm function
or the equation
  X <- as.matrix(fit$model[, -1])                      # Matrix of predictors
  td <- data.frame(smmry$coefficients)                 # matrix with coefficients
  b <- as.matrix(td[rownames(td) != "(Intercept)", 1])   # vector of coefficients
excluding intercept
  k <- nrow(b)                                         # Number of parameters excluding
the intercept
  SSE <- sum((fit$residuals)^2)                        # Sum of square errors
  n <- nrow(X)                                         # Number of data points used in the
model
  S <- (SSE/(n - k) )
  dep <- fit$model[ ,1]
  F <- as.numeric(( dep%*%dep - SSE)/(k*S))            # F statistic
  Rsq <- smmry$r.squared                               # R squared value

  vl2 <- (n-k)*log(SSE/n) + k*log(as.numeric(t(b)%*%crossprod(X)%*%b)) +
(n-k-1)*log(n/(n-k)) - (k+1)*log(k)
  return(vl2)

}
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Combine gMDL, nMDL and the Exact MDL into one function. This will be useful during simulations

```
MDL <- function(fit=model){
  smmry <- summary(fit)                    # Summary of the lm function or the equation
  X <- as.matrix(fit$model[, -1])          # Matrix of predictors
  td <- data.frame(smmry$coefficients)     # matrix with coefficients
  b <- as.matrix(td[rownames(td) != "(Intercept)", 1])   # vector of coefficients excluding intercept
  k <- nrow(b)                             # Number of parameters excluding the intercept
  SSE <- sum((fit$residuals)^2)            # Sum of square errors
  n <- nrow(X)                             # Number of data points used in the model
  S <- (SSE/(n - k) )
  dep <- fit$model[ ,1]
  F <- as.numeric(( dep%*%dep - SSE)/(k*S))        # F statistic
  Rsq <- smmry$r.squared                   # R squared value
  t.values <- as.numeric(summary(fit)$coefficients[,3]) # Extract t-statistic value from the model including the intercept
  RSS <- sum((fit$residuals)^2) # Sum of Squares residual
  n <- nrow(fit$model) # The number of data points
  log_lkl <- (n/2)*log2(RSS/n)
  unv_cdlen <- sum(sapply(t.values,IDl_len)) # Calculates the universal codes

  rMDL <- log_lkl + unv_cdlen      #Rissanen MDL
  eMDL <- (n-k)*log(SSE/n) + k*log(as.numeric(t(b)%*%crossprod(X)%*%b)) + (n-k-1)*log(n/(n-k)) - (k+1)*log(k)
  gMDL <- ifelse(Rsq >= k/n,(n/2)*log(S) + (k/2)*log(F) + log(n), (n/2)*log((dep%*%dep)/n) + .5*log(n) )
  nMDL <- (n/2)*log(S) + (k/2)*log(F) + .5*log(n-k) - 1.5*log(k)
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

```
  aic <- AIC(fit)  # AIC
  bic <- BIC(fit)  # BIC
  combo <- data.frame(Rissanen.MDL = rMDL, G.Prior.MDL = gMDL,
Norm.MDL = nMDL, Exact.MDL = eMDL, AIC = aic, BIC = bic)
  return(combo)
}

## --------------------------------------------------------------------------------------
----------------------------------

## The Simulation starts here


# Create Empty Matrices to hold output from simulation

fit_1 <- data.frame(Rissanen.MDL = numeric() , G.Prior.MDL = numeric(),
Norm.MDL = numeric(), Exact.MDL = numeric(), AIC= numeric(), BIC =
numeric())
fit_2 <- data.frame(Rissanen.MDL = numeric() , G.Prior.MDL = numeric(),
Norm.MDL = numeric(), Exact.MDL = numeric(), AIC= numeric(), BIC =
numeric())
fit_3 <- data.frame(Rissanen.MDL = numeric() , G.Prior.MDL = numeric(),
Norm.MDL = numeric(), Exact.MDL = numeric(), AIC= numeric(), BIC =
numeric())

## Set the number of Iterations

niter <- 1000        ## 1000 Iterations

## Set the size of the data

n <- 50 ## Number of rows in sample generated
# if we take lag or difference of 1 then this may affect the sample size. take
note
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

```
### Loop begins here
set.seed(1231)

for(i in 1:niter){

  xt <- cumsum(rnorm(n))
  et <- rnorm(n)
  yt <- xt + et

  dxt <- diff(xt)
  dxtpos <- (dxt + abs(dxt))/2
  dxtneg <- dxt - dxtpos

  reg <- lm(yt ~ xt)

  ECT<-reg$res
  ECTpos<-(ECT+abs(ECT))/2
  ECTneg<-ECT-ECTpos
  lECTpos<-(Lag(ECTpos,1))
  lECTneg<-(Lag(ECTneg,1))

  lECTpos<-lECTpos[-c(1,1)]
  lECTneg<-lECTneg[-c(1,1)]

  dyt <- 0.95*dxxpos + 0.20*dxxneg - 0.25*lECTpos - 0.75*lECTneg +
rnorm(n-1, 0, 1)## can't change bcos these are the parameters in research.

  dyy <- dyt[c(6:n-5)]
  dxxpos <- dxtpos[c(6:n-5)]
  dxxneg <- dxtneg[c(6:n-5)]

  Dyt <- dyt[c(6:n-5)]
  Dxt <- dxt[c(6:n-5)]
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

```
  LECTpos <- lECTpos[c(6:n-5)]

  LECTneg <- lECTneg[c(6:n-5)]

   ## Build Linear Regression Model

  lm_fit1 <- lm(Dyt ~ dxxpos + dxxneg + LECTpos + LECTneg-1 ) ##Complex Error Correction Model = CECM

  lm_fit2 <- lm( Dyt ~ Dxt + LECTpos + LECTneg-1)        ## Standard Error Correction Model=SECM

  lm_fit3 <- lm(dyy ~ dxxpos + dxxneg-1 )        ##Houck's Model in Summed Difference= HOUCKS_L

  ## Append the 1-line output to the Matrix created above

  fit_1 <- rbind(fit_1, MDL(lm_fit1))
  fit_2 <- rbind(fit_2, MDL(lm_fit2))
  fit_3 <- rbind(fit_3, MDL(lm_fit3))


}


## Average of the Matrices

Criteria_Averages <- rbind(sapply(fit_1, mean ),
                           sapply(fit_2, mean ),
                           sapply(fit_3, mean )
)

# Rename the rows of the matrix
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

```
rownames(Criteria_Averages) <- c("lm_fit1", "lm_fit2", "lm_fit3")  # Rename rows to indicate the model

rownames(Criteria_Averages) <- c("CECM", "SECM", "HOUCKS_L")

### Data Frames for the models -- A matrix for each MDL prediction (how often each type of MDL is able to predict the data generating process)

Rissanen_df <- data.frame(fit_1 [,1], fit_2[,1], fit_3[,1])
colnames(Rissanen_df) <- c("CECM", "SECM", "HOUCKS_L")

G.Prior_df <- data.frame( fit_1[,2], fit_2[,2], fit_3[,2])
colnames(G.Prior_df) <- c("CECM", "SECM", "HOUCKS_L")

Norm_df <- data.frame( fit_1[,3], fit_2[,3], fit_3[,3])
colnames(Norm_df) <- c("CECM", "SECM", "HOUCKS_L")

Exact_df <- data.frame( fit_1[,4], fit_2[,4], fit_3[,4])
colnames(Exact_df) <- c("CECM", "SECM", "HOUCKS_L")

AIC_df <- data.frame( fit_1[,5], fit_2[,5], fit_3[,5])
colnames(AIC_df) <- c("CECM", "SECM", "HOUCKS_L")

BIC_df <- data.frame( fit_1[,6], fit_2[,6], fit_3[,6])
colnames(BIC_df) <- c("CECM", "SECM", "HOUCKS_L")

## Calculate  the Minimum MDL for each iteration for each MDL for the four models

Rissanen_df$Minimum <- apply(Rissanen_df, 1, min)
G.Prior_df$Minimum <- apply(G.Prior_df, 1, min)
Norm_df$Minimum <- apply(Norm_df, 1, min)
Exact_df$Minimum <- apply(Exact_df, 1, min)
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

```
AIC_df$Minimum <- apply(AIC_df, 1 , min)
BIC_df$Minimum <- apply(BIC_df, 1, min)

### Test Each iteration to the Minimum of the MDLs for each iteration -- This will be true or false

Rissanen_df_boolean <- data.frame(Rissanen_df[ ,-ncol(Rissanen_df)] == Rissanen_df[ ,ncol(Rissanen_df)] )
G.Prior_boolean <- data.frame(G.Prior_df[ ,-ncol(G.Prior_df)] == G.Prior_df[ ,ncol(G.Prior_df)] )
Norm_df_boolean <- data.frame(Norm_df[ ,-ncol(Norm_df)] == Norm_df[ ,ncol(Norm_df)] )
Exact_df_boolean <- data.frame(Exact_df[ ,-ncol(Exact_df)] == Exact_df[ ,ncol(Exact_df)] )
AIC_df_boolean <- data.frame(AIC_df[ ,-ncol(AIC_df)] == AIC_df[ ,ncol(AIC_df)])
BIC_df_boolean <- data.frame(BIC_df[ ,-ncol(BIC_df)] == BIC_df[ ,ncol(BIC_df)])

## Percentage of the time an MDL chooses a model
Criteria_Choose_Pcntage <- rbind(sapply(Rissanen_df_boolean, mean),
                    sapply(G.Prior_boolean, mean),
                    sapply(Norm_df_boolean, mean),
                    sapply(Exact_df_boolean, mean),
                    sapply(AIC_df_boolean, mean),
                    sapply(BIC_df_boolean, mean)
)

rownames(Criteria_Choose_Pcntage) <- c("Rissanen.MDL", "G.Prior.MDL", "Norm.MDL", "Exact.MDL", "AIC", "BIC")
#### # Print the MDL Averages
Criteria_Averages
### Print the number of times an MDL chooses model
Criteria_Choose_Pcntage
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

APPENDIX I

DATA ANALYSIS AND SIMULATION RESULTS

I 1. Results for SECM DGP

Comparison of all six criteria across the following dynamics:

1. Data generating process (DGP) is SECM {dyt <- 0.7*dxt + 0.25*lECTpos + 0.75*lECTneg + rnorm(n-1, 0, 1)}

Varing Sample Sizes of n = 50, 150, and 500. The number of Monte Carlo simulations are 1000 throughout the comparison.

Start with n= 50

Criteria_Averages

	Rissanen.MDL	G.Prior.MDL	Norm.MDL	Exact.MDL	AIC	BIC
CECM	12.16307	7.874018	3.844700	6.655327	132.5745	141.6078
SECM	11.30761	7.142498	3.556752	5.504065	131.6881	138.9148
HOUCKS_L	15.29472	11.328300	8.362517	14.304666	142.2828	147.7028

```
>
>
> ### Print the number of times an MDL chooses model
> Criteria_Choose_Pcntage
```

	CECM	SECM	HOUCKS_L
Rissanen.MDL	0.299	0.635	0.066
G.Prior.MDL	0.090	0.858	0.052
Norm.MDL	0.206	0.780	0.014
Exact.MDL	0.122	0.842	0.036
AIC	0.177	0.795	0.028
BIC	0.050	0.863	0.087

##Table 3a: Criteria Averages Per Model

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

##Table 3b: Percentage of Time Criteria Predicts Data Generating Process

n = 150

```
#### # Print the MDL Averages
> Criteria_Averages
       Rissanen.MDL G.Prior.MDL  Norm.MDL Exact.MDL      AIC      BIC
CECM        19.99147    11.35449  6.772699 11.341253 416.6468 431.5305
SECM        16.33678    10.03636  5.889626  8.999743 415.5806 427.4875
HOUCKS_L         37.37689       26.01201  22.476973  41.363506  450.7932
459.7234
>
>
> ### Print the number of times an MDL chooses model
> Criteria_Choose_Pcntage
            CECM  SECM HOUCKS_L
Rissanen.MDL 0.033 0.967        0
G.Prior.MDL  0.046 0.954        0
Norm.MDL     0.083 0.917        0
Exact.MDL    0.058 0.942        0
AIC          0.139 0.861        0
BIC          0.021 0.979        0
```

n =500

```
#### # Print the MDL Averages
> Criteria_Averages
       Rissanen.MDL G.Prior.MDL  Norm.MDL Exact.MDL      AIC      BIC
CECM        27.99848    15.25561 10.069836  16.70770 1410.462 1431.485
SECM        22.41974    13.39023  8.636989  13.26664 1409.476 1426.295
HOUCKS_L       104.92220       72.62824  68.484218  132.15017  1532.417
1545.030
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

```
>
>
> ### Print the number of times an MDL chooses model
> Criteria_Choose_Pcntage
             CECM  SECM HOUCKS_L
Rissanen.MDL 0.005 0.995        0
G.Prior.MDL  0.033 0.967        0
Norm.MDL     0.052 0.948        0
Exact.MDL    0.036 0.964        0
AIC          0.164 0.836        0
BIC          0.013 0.987        0
```

2. When stochastic variance (noise) in data increases as against the standardized conditions

compared at n = 150

rnorm (n-1, 0, 1)

```
#### # Print the MDL Averages
> Criteria_Averages
         Rissanen.MDL G.Prior.MDL  Norm.MDL Exact.MDL      AIC      BIC
CECM         19.99147    11.35449  6.772699 11.341253 416.6468 431.5305
SECM         16.33678    10.03636  5.889626  8.999743 415.5806 427.4875
HOUCKS_L     37.37689    26.01201 22.476973 41.363506 450.7932 459.7234
>
>
> ### Print the number of times an MDL chooses model
> Criteria_Choose_Pcntage
             CECM  SECM HOUCKS_L
Rissanen.MDL 0.033 0.967        0
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

```
G.Prior.MDL  0.046 0.954       0
Norm.MDL     0.083 0.917       0
Exact.MDL    0.058 0.942       0
AIC          0.139 0.861       0
BIC          0.021 0.979       0
```

rnorm (n-1, 0, 2)

```
 #### # Print the MDL Averages
> Criteria_Averages
        Rissanen.MDL G.Prior.MDL Norm.MDL Exact.MDL      AIC      BIC
CECM        156.3295    109.1940 104.6122  207.0203 617.6595 632.5432
SECM        155.5390    108.5067 104.3600  205.9404 616.5933 628.5002
HOUCKS_L    159.0446    112.1920 108.6569  213.7235 625.5506 634.4808
>
>
> ### Print the number of times an MDL chooses model
> Criteria_Choose_Pcntage
             CECM  SECM HOUCKS_L
Rissanen.MDL 0.292 0.603    0.105
G.Prior.MDL  0.088 0.849    0.063
Norm.MDL     0.210 0.771    0.019
Exact.MDL    0.115 0.842    0.043
AIC          0.131 0.820    0.049
BIC          0.010 0.806    0.184
```

rnorm (n-1, 0, 3)

```
#### # Print the MDL Averages
> Criteria_Averages
        Rissanen.MDL G.Prior.MDL Norm.MDL Exact.MDL      AIC      BIC
CECM        237.8408    166.4827 161.9558  321.7075 735.2444 750.1280
SECM        237.2629    166.1235 162.0072  321.2348 734.1782 746.0851
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

```
HOUCKS_L    237.7474   167.4120 163.9777  324.3650 737.7201 746.6503
>
>
> ### Print the number of times an MDL chooses model
> Criteria_Choose_Pcntage
             CECM  SECM HOUCKS_L
Rissanen.MDL 0.212 0.389    0.399
G.Prior.MDL  0.119 0.669    0.241
Norm.MDL     0.428 0.537    0.035
Exact.MDL    0.188 0.675    0.137
AIC          0.100 0.681    0.219
BIC          0.005 0.475    0.520
```

3. Varing Levels of asymmetry of n (50, 150, 500)

-Strong level of asymmetry (0.25, 0.75) when n = 50

```
 #### # Print the MDL Averages
> Criteria_Averages
         Rissanen.MDL G.Prior.MDL Norm.MDL Exact.MDL      AIC      BIC
CECM        12.16307    7.874018 3.844700  6.655327 132.5745 141.6078
SECM        11.30761    7.142498 3.556752  5.504065 131.6881 138.9148
HOUCKS_L          15.29472      11.328300  8.362517  14.304666  142.2828
147.7028
>
>
> ### Print the number of times an MDL chooses model
> Criteria_Choose_Pcntage
             CECM  SECM HOUCKS_L
Rissanen.MDL 0.299 0.635    0.066
G.Prior.MDL  0.090 0.858    0.052
Norm.MDL     0.206 0.780    0.014
Exact.MDL    0.122 0.842    0.036
AIC          0.177 0.795    0.028
BIC          0.050 0.863    0.087
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

-Strong level of asymmetry (0.25, 0.75) when n = 150

Print the MDL Averages

> Criteria_Averages

	Rissanen.MDL	G.Prior.MDL	Norm.MDL	Exact.MDL	AIC	BIC
CECM	19.99147	11.35449	6.772699	11.341253	416.6468	431.5305
SECM	16.33678	10.03636	5.889626	8.999743	415.5806	427.4875
HOUCKS_L	37.37689	26.01201	22.476973	41.363506	450.7932	459.7234

>

>

> ### Print the number of times an MDL chooses model

> Criteria_Choose_Pcntage

	CECM	SECM	HOUCKS_L
Rissanen.MDL	0.033	0.967	0
G.Prior.MDL	0.046	0.954	0
Norm.MDL	0.083	0.917	0
Exact.MDL	0.058	0.942	0
AIC	0.139	0.861	0
BIC	0.021	0.979	0

-Strong level of asymmetry (0.25, 0.75) when n = 500

Print the MDL Averages

> Criteria_Averages

	Rissanen.MDL	G.Prior.MDL	Norm.MDL	Exact.MDL	AIC	BIC
CECM	27.99848	15.25561	10.069836	16.70770	1410.462	1431.485
SECM	22.41974	13.39023	8.636989	13.26664	1409.476	1426.295
HOUCKS_L	104.92220	72.62824	68.484218	132.15017	1532.417	1545.030

>

>

> ### Print the number of times an MDL chooses model

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

```
> Criteria_Choose_Pcntage
             CECM  SECM HOUCKS_L
Rissanen.MDL 0.005 0.995        0
G.Prior.MDL  0.033 0.967        0
Norm.MDL     0.052 0.948        0
Exact.MDL    0.036 0.964        0
AIC          0.164 0.836        0
BIC          0.013 0.987        0
```

-Weak level of asymmetry (0.25, 0.50) when n = 50

```
#### # Print the MDL Averages
> Criteria_Averages
         Rissanen.MDL G.Prior.MDL Norm.MDL Exact.MDL      AIC      BIC
CECM        10.994543    7.475542 3.448125  5.862177 132.5745 141.6078
SECM         9.870371    6.835296 3.249550  4.889662 131.6881 138.9148
HOUCKS_L    12.099759    8.885587 5.921708  9.423047 137.2567 142.6767
>
>
> ### Print the number of times an MDL chooses model
> Criteria_Choose_Pcntage
             CECM  SECM HOUCKS_L
Rissanen.MDL 0.217 0.634    0.149
G.Prior.MDL  0.086 0.734    0.180
Norm.MDL     0.229 0.701    0.070
Exact.MDL    0.124 0.733    0.143
AIC          0.158 0.710    0.132
BIC          0.039 0.688    0.273
```

-Weak level of asymmetry (0.25, 0.50) when n = 150

```
#### # Print the MDL Averages
> Criteria_Averages
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

```
       Rissanen.MDL G.Prior.MDL  Norm.MDL Exact.MDL      AIC      BIC
CECM        18.45466   10.937248  6.355453 10.506760 416.6468 431.5305
SECM        14.91338    9.720577  5.573839  8.368168 415.5806 427.4875
HOUCKS_L          26.30927      18.000260  14.465227  25.340016  434.6182
443.5484
>
>
> ### Print the number of times an MDL chooses model
> Criteria_Choose_Pcntage
             CECM  SECM HOUCKS_L
Rissanen.MDL 0.037 0.962    0.001
G.Prior.MDL  0.055 0.941    0.004
Norm.MDL     0.092 0.908    0.000
Exact.MDL    0.062 0.935    0.003
AIC          0.139 0.861    0.000
BIC          0.017 0.963    0.020
```

-Weak level of asymmetry (0.25, 0.50) when n = 500

```
#### # Print the MDL Averages
> Criteria_Averages
       Rissanen.MDL G.Prior.MDL  Norm.MDL Exact.MDL      AIC      BIC
CECM        26.82988    14.82394  9.638165  15.84436 1410.462 1431.485
SECM        21.29252    13.06554  8.312300  12.61727 1409.476 1426.295
HOUCKS_L          65.28723        44.92147  40.777448    76.73663  1476.853
1489.467
>
>
> ### Print the number of times an MDL chooses model
> Criteria_Choose_Pcntage
             CECM  SECM HOUCKS_L
Rissanen.MDL 0.007 0.993        0
G.Prior.MDL  0.036 0.964        0
Norm.MDL     0.058 0.942        0
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Exact.MDL 0.042 0.958 0

AIC 0.164 0.836 0

BIC 0.013 0.987 0

4. Stable and unstable conditions of asymmetry (small sample size vrs large noice and large sample size vrs small noise)

Stable (n = 150, noise = 0.5)

Print the MDL Averages

> Criteria_Averages

	Rissanen.MDL	G.Prior.MDL	Norm.MDL	Exact.MDL	AIC	BIC
CECM	-117.62811	-86.83157	-91.41336	-185.0309	215.6341	230.5178
SECM	-123.45086	-88.72146	-92.86820	-188.5159	214.5679	226.4748
HOUCKS_L	-81.25335	-57.72025	-61.25528	-126.1010	281.0656	289.9958

>

>

> ### Print the number of times an MDL chooses model

> Criteria_Choose_Pcntage

	CECM	SECM	HOUCKS_L
Rissanen.MDL	0.005	0.995	0
G.Prior.MDL	0.023	0.977	0
Norm.MDL	0.036	0.964	0
Exact.MDL	0.027	0.973	0
AIC	0.139	0.861	0
BIC	0.021	0.979	0

Unstable (n = 50, noise = 2)

Print the MDL Averages

> Criteria_Averages

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

```
        Rissanen.MDL G.Prior.MDL Norm.MDL Exact.MDL      AIC      BIC
CECM        51.62390    36.18907 32.31139  63.58870 194.9578 203.9911
SECM        50.78374    35.99887 32.56002  63.51060 194.0714 201.2980
HOUCKS_L    50.26096    36.28984 33.51287  64.60537 194.9734 200.3934
>
>
> ### Print the number of times an MDL chooses model
> Criteria_Choose_Pcntage
             CECM  SECM HOUCKS_L
Rissanen.MDL 0.112 0.261    0.627
G.Prior.MDL  0.194 0.491    0.451
Norm.MDL     0.656 0.271    0.073
Exact.MDL    0.306 0.447    0.247
AIC          0.112 0.452    0.436
BIC          0.016 0.307    0.677
```

General case senerio for n = 1000, niter = 1000

```
#### # Print the MDL Averages
> Criteria_Averages
        Rissanen.MDL G.Prior.MDL  Norm.MDL Exact.MDL      AIC      BIC
CECM        33.41405    18.13018  12.59736  21.06456 2830.934 2855.448
SECM        26.64921    15.89862  10.79782  16.89012 2829.915 2849.526
HOUCKS_L       199.73897      138.44222  133.95012  262.38379  3080.173
3094.881
>
>
> ### Print the number of times an MDL chooses model
> Criteria_Choose_Pcntage
             CECM  SECM HOUCKS_L
Rissanen.MDL 0.002 0.998        0
G.Prior.MDL  0.020 0.980        0
Norm.MDL     0.032 0.968        0
```


Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Exact.MDL	0.024	0.976	0
AIC	0.147	0.853	0
BIC	0.009	0.991	0

I 2. Results for SECM DGP (other tables and charts)

1. Varying Sample Size of n=50, 150, 500

Sample Size	Model Fitted	
50		SECM
	Selection Criteria	
	Rissanen. MDL	0.635 (63.5%)
	G. Prior. MDL	0.858 (85.8%)
	Norm. MDL	0.780 (78.0%)
	Exact. MDL	0.842 (84.2%)
	AIC	0.795 (79.5%)
	BIC	0.863 (86.3%)
150		SECM
	Selection Criteria	
	Rissanen. MDL	0.967 (96.7%)
	G. Prior. MDL	0.954 (95.4%)
	Norm. MDL	0.917 (91.7%)
	Exact. MDL	0.942 (94.2%)
	AIC	0.861 (86.1%)
	BIC	0.979 (97.9%)
500		SECM
	Selection Criteria	
	Rissanen. MDL	0.995 (99.5%)
	G. Prior. MDL	0.967 (96.7%)
	Norm. MDL	0.948 (94.8%)
	Exact. MDL	0.964 (96.4%)
	AIC	0.836 (83.6%)
	BIC	0.987 (98.7%)

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

2. When Stochastic Variance (noise) in the data increasing as against the standardized conditions

Compared at n=150

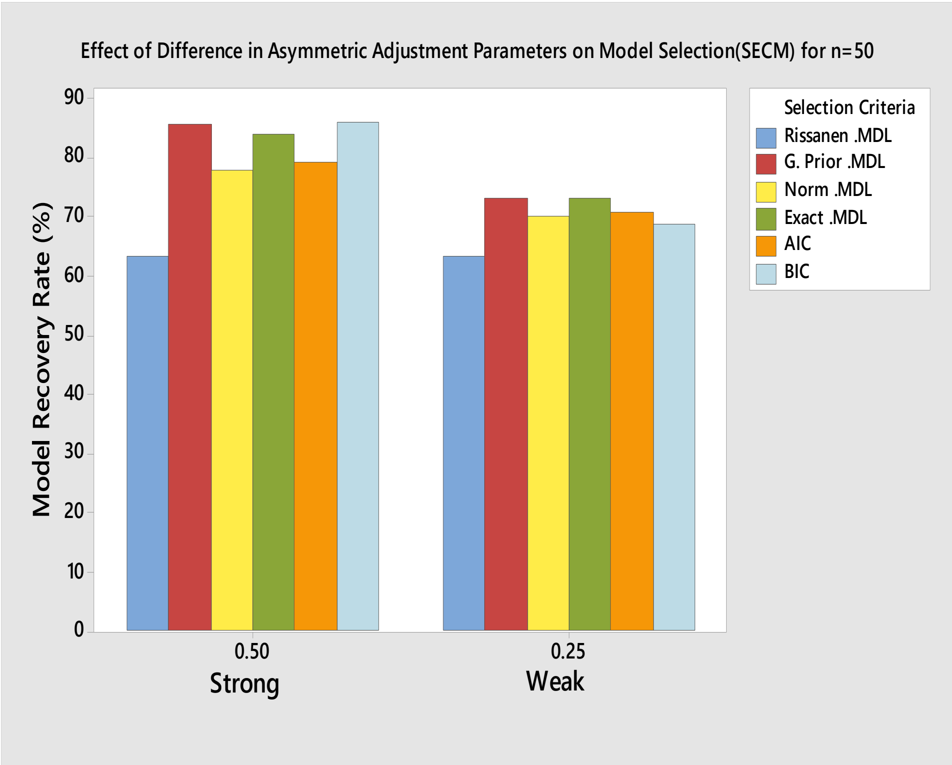
Stochastic Variance(noise)	Model Fitted	
1		SECM
	Selection Criteria	
	Rissanen. MDL	0.967 (96.7%)
	G. Prior. MDL	0.954 (95.4%)
	Norm. MDL	0.917 (91.7%)
	Exact. MDL	0.942 (94.2%)
	AIC	0.861 (86.1%)
	BIC	0.979 (97.9%)
2		SECM
	Selection Criteria	
	Rissanen. MDL	0.603 (60.3%)
	G. Prior. MDL	0.849 (84.9%)
	Norm. MDL	0.771 (77.1%)
	Exact. MDL	0.842 (84.2%)
	AIC	0.820 (82.0%)
	BIC	0.806 (80.6%)
3		SECM
	Selection Criteria	
	Rissanen. MDL	0.389 (38.9%)
	G. Prior. MDL	0.669 (66.9%)
	Norm. MDL	0.537 (53.7%)
	Exact. MDL	0.675 (67.5%)
	AIC	0.681 (68.1%)
	BIC	0.475 (47.5%)

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

3. Varing Levels of asymmetry of n=50

Difference		Model Fitted
0.50 (Strong)		SECM
	Selection Criteria	
	Rissanen. MDL	0.635 (63.5%)
	G. Prior. MDL	0.858 (85.8%)
	Norm. MDL	0.780 (78.0%)
	Exact. MDL	0.842 (84.2%)
	AIC	0.795 (79.5%)
	BIC	0.863 (86.3%)
0.25 (Weak)		SECM
	Selection Criteria	
	Rissanen. MDL	0.634(63.4%)
	G. Prior. MDL	0.734 (73.4%)
	Norm. MDL	0.701(70.1%)
	Exact. MDL	0.733 (73.3%)
	AIC	0.710 (71.0%)
	BIC	0.688(68.8%)



4. Varing Levels of asymmetry of n=150

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

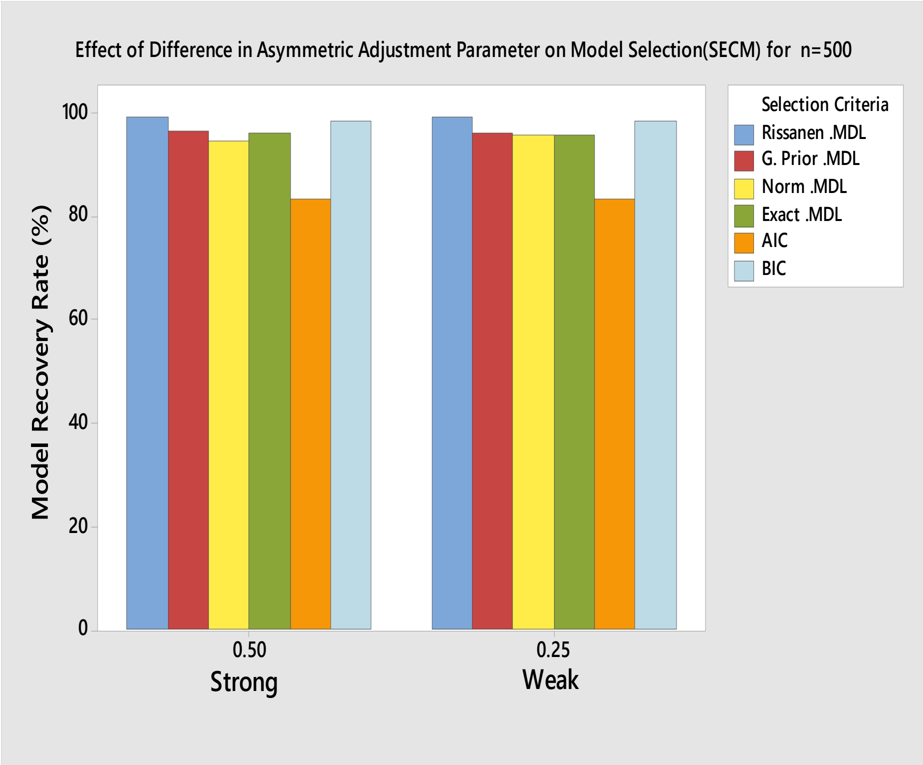
Difference		Model Fitted
0.50 (Strong)		SECM
	Selection Criteria	
	Rissanen. MDL	0.967 (96.7%)
	G. Prior. MDL	0.954 (95.4%)
	Norm. MDL	0.917 (91.7%)
	Exact. MDL	0.942 (94.2%)
	AIC	0.861 (86.1%)
	BIC	0.979 (97.9%)
0.25 (Weak)		SECM
	Selection Criteria	
	Rissanen. MDL	0.962 (96.2%)
	G. Prior. MDL	0.941 (94.1%)
	Norm. MDL	0.908 (90.8%)
	Exact. MDL	0.935 (93.5%)
	AIC	0.861 (86.1%)
	BIC	0.963 (96.3%)

5. Varing Levels of asymmetry of n=500

Difference		Model Fitted
0.50 (Strong)		SECM
	Selection Criteria	
	Rissanen. MDL	0.995 (99.5%)
	G. Prior. MDL	0.967 (96.7%)
	Norm. MDL	0.948 (94.8%)
	Exact. MDL	0.964 (96.4%)
	AIC	0.836 (83.6%)
	BIC	0.987 (98.7%)
0.25 (Weak)		SECM
	Selection Criteria	
	Rissanen. MDL	0.993 (99.3%)
	G. Prior. MDL	0.964 (96.4%)
	Norm. MDL	0.942 (94.2%)
	Exact. MDL	0.958 (95.8%)
	AIC	0.836 (83.6%)
	BIC	0.987 (98.7%)

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui



6. Stable and unstable conditions of asymmetry

Difference	Model Fitted	
Stable		SECM
	Selection Criteria	
	Rissanen. MDL	0.995 (99.5%)
	G. Prior. MDL	0.977 (97.7%)
	Norm. MDL	0.964 (96.4%)
	Exact. MDL	0.973 (97.3%)
	AIC	0.861 (86.1%)
	BIC	0.979 (97.9%)
Unstable		SECM
	Selection Criteria	
	Rissanen. MDL	0.261 (26.1%)
	G. Prior. MDL	0.491 (49.1%)
	Norm. MDL	0.271 (27.1%)
	Exact. MDL	0.447 (44.7%)
	AIC	0.452 (45.2%)
	BIC	0.307 (30.7%)

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

I 3. Results for SECM DGP (other tables and charts)

2. Varying Sample Size of n=50, 150, 500

Sample Size	Model Fitted	
50		CECM
	Selection Criteria	
	Rissanen. MDL	0.639 (63.9%)
	G. Prior. MDL	0.491 (49.1%)
	Norm. MDL	0.698 (69.8%)
	Exact. MDL	0.551 (55.1%)
	AIC	0.626 (62.6%)
	BIC	0.340 (34.0%)
150		CECM
	Selection Criteria	
	Rissanen. MDL	0.901 (90.1%)
	G. Prior. MDL	0.923 (92.3%)
	Norm. MDL	0.950 (95.0%)
	Exact. MDL	0.933 (93.3%)
	AIC	0.969 (96.9%)
	BIC	0.870 (87.0%)
500		CECM
	Selection Criteria	
	Rissanen. MDL	1.000 (100%)
	G. Prior. MDL	1.000 (100%)
	Norm. MDL	1.000 (100%)
	Exact. MDL	1.000 (100%)
	AIC	1.000 (100%)
	BIC	1.000 (100%)

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

2. When Stochastic Variance (noise) in the data increasing as against the standardized conditions

Compared at n=150

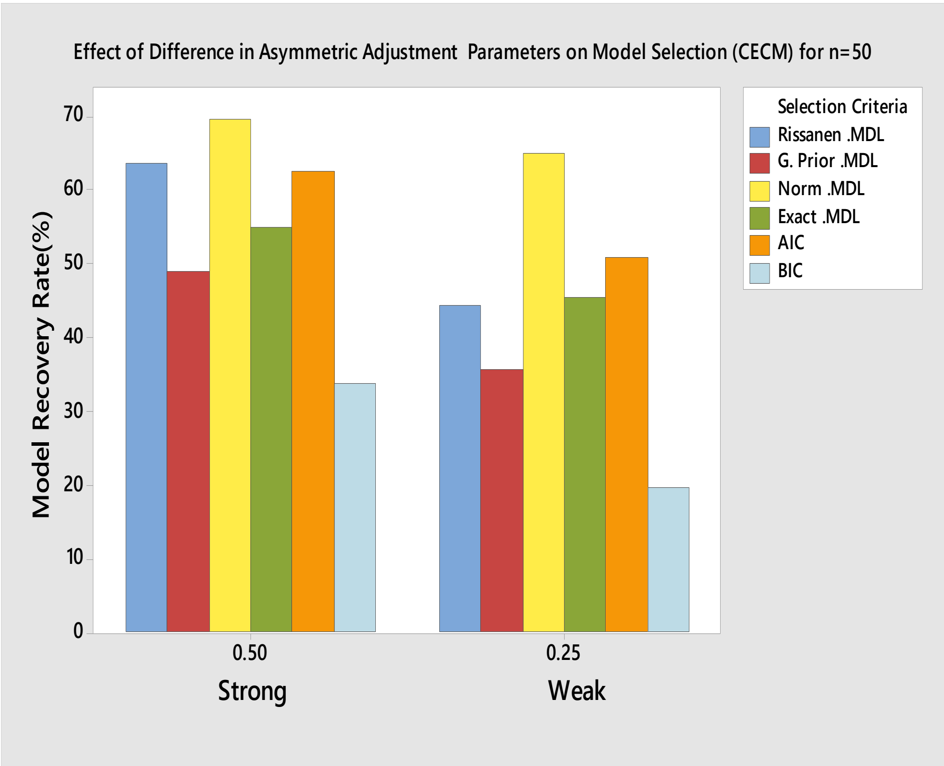
Stochastic Variance(noise)	Model Fitted	
1		CECM
	Selection Criteria	
	Rissanen. MDL	0.901 (90.1%)
	G. Prior. MDL	0.923 (92.3%)
	Norm. MDL	0.950 (95.0%)
	Exact. MDL	0.933 (93.3%)
	AIC	0.969 (96.9%)
	BIC	0.870 (87.0%)
2		CECM
	Selection Criteria	
	Rissanen. MDL	0.627 (62.7%)
	G. Prior. MDL	0.441 (44.1%)
	Norm. MDL	0.638 (63.8%)
	Exact. MDL	0.515 (51.5%)
	AIC	0.530 (53.0%)
	BIC	0.160 (16.0%)
3		CECM
	Selection Criteria	
	Rissanen. MDL	0.333 (33.3%)
	G. Prior. MDL	0.233 (23.3%)
	Norm. MDL	0.602 (60.2%)
	Exact. MDL	0.329 (32.9%)
	AIC	0.231 (23.1%)
	BIC	0.020 (2.0%)

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

7. Varying Levels of asymmetry of n=50

Difference	Model Fitted	
0.50 (Strong)		CECM
	Selection Criteria	
	Rissanen. MDL	0.639 (63.9%)
	G. Prior. MDL	0.491 (49.1%)
	Norm. MDL	0.698 (69.8%)
	Exact. MDL	0.551 (84.2%)
	AIC	0.626 (79.5%)
	BIC	0.340 (34.0%)
0.25 (Weak)		CECM
	Selection Criteria	
	Rissanen. MDL	0.446 (44.6%)
	G. Prior. MDL	0.358 (35.8%)
	Norm. MDL	0.652 (65.2%)
	Exact. MDL	0.455 (45.5%)
	AIC	0.511 (51.1%)
	BIC	0.198 (19.8%)



Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

8. Varying Levels of asymmetry of n=150

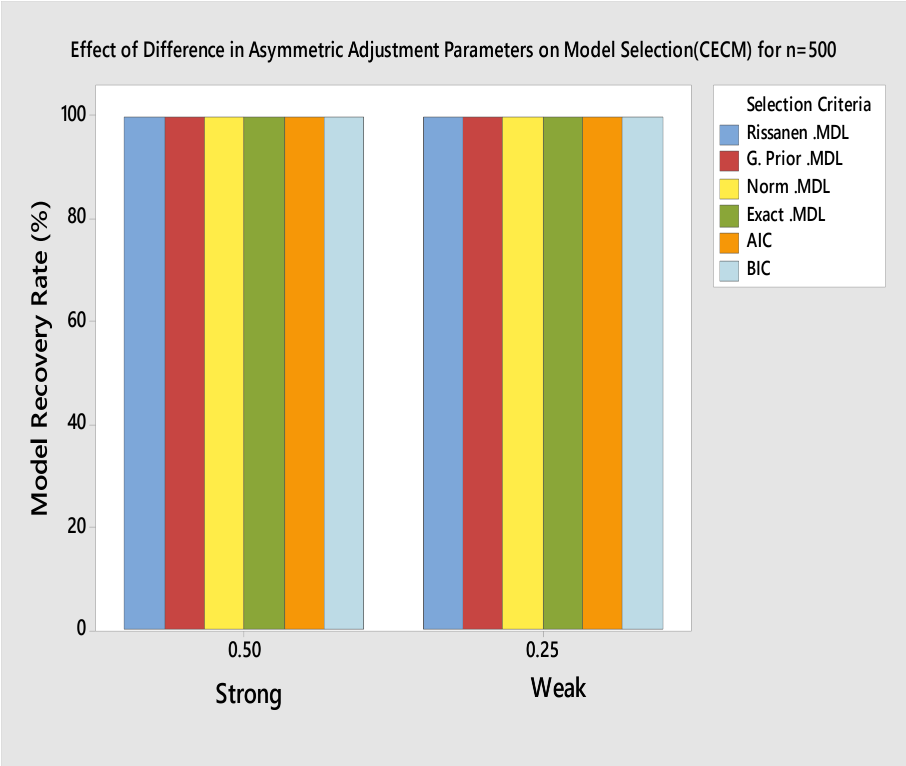
Difference	Model Fitted	
0.50 (Strong)		CECM
	Selection Criteria	
	Rissanen. MDL	0.901 (90.1%)
	G. Prior. MDL	0.923 (92.3%)
	Norm. MDL	0.950 (95.0%)
	Exact. MDL	0.933 (93.3%)
	AIC	0.969 (96.9%)
	BIC	0.870 (87.0%)
0.25 (Weak)		CECM
	Selection Criteria	
	Rissanen. MDL	0.883 (88.3%)
	G. Prior. MDL	0.905 (90.5%)
	Norm. MDL	0.944 (94.4%)
	Exact. MDL	0.923 (92.3%)
	AIC	0.964 (96.4%)
	BIC	0.809 (80.9%)

9. Varing Levels of asymmetry of n=500

Difference	Model Fitted	
0.50 (Strong)		CECM
	Selection Criteria	
	Rissanen. MDL	1.000 (100%)
	G. Prior. MDL	1.000 (100%)
	Norm. MDL	1.000 (100%)
	Exact. MDL	1.000 (100%)
	AIC	1.000 (100%)
	BIC	1.000 (100%)
0.25 (Weak)		CECM
	Selection Criteria	
	Rissanen. MDL	1.000 (100%)
	G. Prior. MDL	1.000 (100%)
	Norm. MDL	1.000 (100%)
	Exact. MDL	1.000 (100%)
	AIC	1.000 (100%)
	BIC	1.000 (100%)

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui



10. Stable and unstable conditions of asymmetry

Difference	Model Fitted	
Stable		CECM
	Selection Criteria	
	Rissanen. MDL	1.000 (100%)
	G. Prior. MDL	1.000 (100%)
	Norm. MDL	1.000 (100%)
	Exact. MDL	1.000 (100%)
	AIC	1.000 (100%)
	BIC	1.000 (100%)
Unstable		CECM
	Selection Criteria	
	Rissanen. MDL	0.186 (18.6%)
	G. Prior. MDL	0.188 (18.8%)
	Norm. MDL	0.669 (66.9%)
	Exact. MDL	0.313 (31.3%)
	AIC	0.152 (15.2%)
	BIC	0.038 (3.8%)

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

APPENDIX J

DERIVATION OF R-FUNCTIONS AND PACKAGES

Appendix J1: Brief Introduction to R

In order to understand what R is doing for us, there are a few fundamentals about the language and how it operates. Readers should download and install R and read the help file inbuilt or any R basic book to catch up with the pace. The following might be helpful:

Prompts

When R is waiting for us to tell it what to do, it begins the line with

>

and this is called the prompt.

If we give it an incomplete command and it cannot finish the task requested it provides:

+ or other versions (R-studio) will show a stop sign.

To get out of R we use the command:

> q ()

The R software can be downloaded at:

https://www.rstudio.com/products/rstudio/download/

Note: download R first, install it, then download R studio and install that as well. The order is important and the instructions are in the link. If it doesn't work, please Google "how to install R".

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Appendix J2: Derivation of Rissanen MDL (rDML) from AIC

set.seed(189)

Note that the R codes come with its alphanumeric characters or its own language.

....

x1 <- rnorm(1000)

We obtain the following values:

0.53897541, 0.21970589, -0.74246202, -0.25227473, -0.05193457, 1.01742024, ..., etc

x2 <- rnorm(1000)

We obtain the following values:

0.3542655, -1.6688001, -0.2104529, 0.9828985, -0.5264001, -0.5318512, ..., etc

We obtain the following values:

x3 <- rnorm(1000)

0.1662904, -1.2096978, -1.3239037, 0.8006888, 1.8387752, -1.2672122, ..., etc

x4 <- rnorm(1000)

We obtain the following values:

-0.96014053, 0.75370303, -0.03409247, -0.54589023, -1.15223634,

-0.76674984, ..., etc

y <- 2*x1 + 4*x2 + 17*rnorm(1000)

....

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Appendix J3: Other packages associated with “HMISC”

All codes developed for the computations of the various criteria (AIC, BIC, MDLs) can be found at Appendix H4. The package ‘Hmisc’ was installed to help with R-codes used in calculating ‘lags’ or differencing a time series (our price models). In order to load package ‘Hmisc’ it is necessary to also load the following packages: lattice, survival, Formula, ggplot2, from the R-domain before package ‘Hmisc’ was attached. The following objects were masked from ‘package: base’: format.pval, round.POSIXt, trunc.POSIXt, units. Note that,

1: package ‘survival’ was built under R version 3.2.5

2: package ‘Formula’ was built under R version 3.2.3

3: package ‘ggplot2’ was built under R version 3.2.5

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

APPENDIX K

ADDITIONAL TABLES

Supplementary Tables for Table 3 and Table 4

Table K1: Criteria Averages & Rankings when DGP is SECM

	rMDL	Ranking	gMDL	Ranking	nMDL	Ranking	eMDL	Ranking	AIC	Ranking	BIC	Ranking
CECM	33.41405	2	18.13018	2	12.59736	2	21.06456	2	2830.934	2	2855.448	2
SECM	26.64921	1	15.89862	1	10.79782	1	16.89012	1	2829.915	1	2849.526	1
HOUCK'S	199.73897	3	138.44222	3	133.95012	3	262.38379	3	3080.173	3	3094.881	3

Sample size (1000) 1000 Monte Carlo Simulations

Table K2: Criteria Averages & Rankings when DGP is CECM

	rMDL	Ranking	gMDL	Ranking	nMDL	Ranking	eMDL	Ranking	AIC	Ranking	BIC	Ranking
CECM	30.20129	1	18.35132	1	12.81850	1	21.50684	1	2830.934	1	2855.448	1
SECM	78.47769	2	55.12114	2	50.02034	2	95.33516	2	2908.553	2	2928.164	2
HOUCK'S	192.48686	3	137.68654	3	133.19444	3	260.87244	3	3078.316	3	3093.024	3

Sample size (1000) 1000 Monte Carlo Simulations

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

APPENDIX L

GRAPHICAL REPRESENTATION OF MODEL SELECTION CRITERIA UNDER STUDY CONDITIONS

Appendix L1: Graphical representation of model selection criteria for SECM analysis

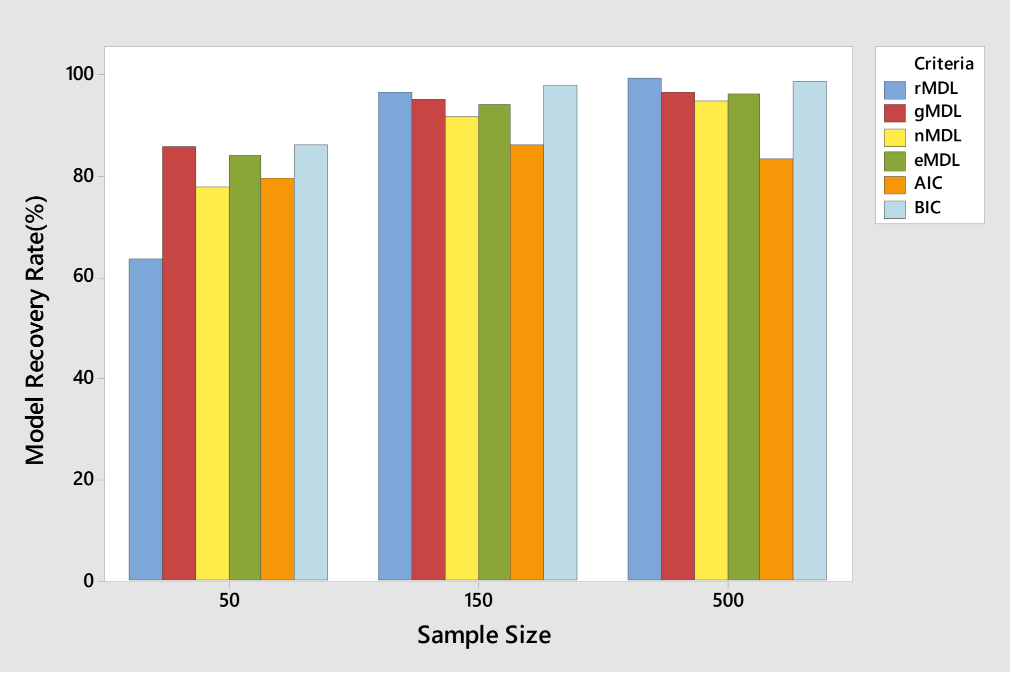


Figure 2: Effect of Sample Size on Model Selection-SECM

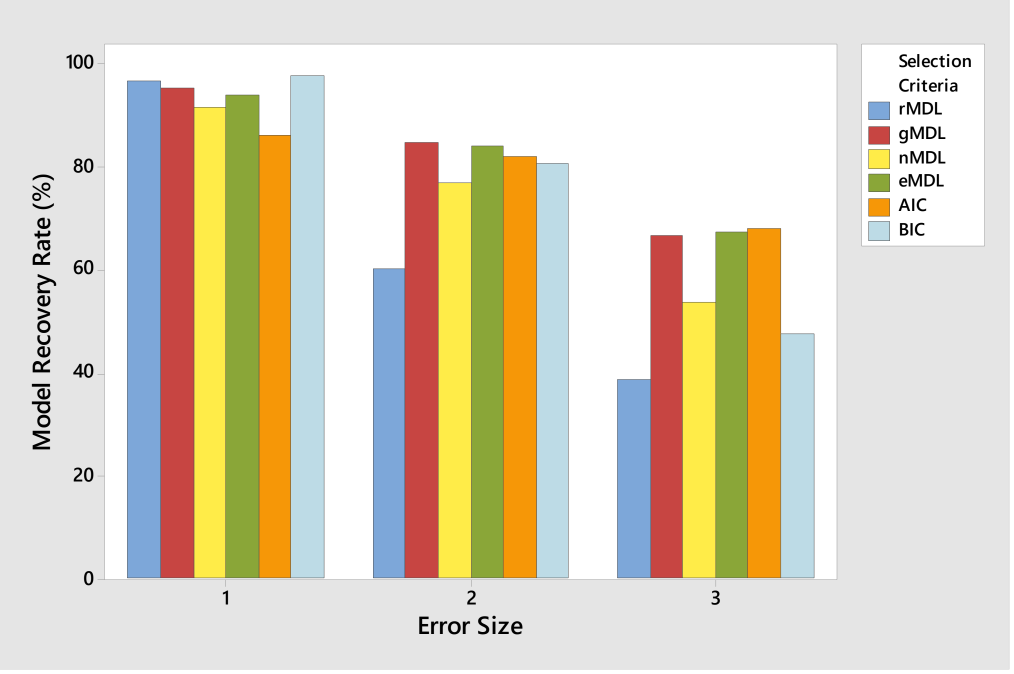


Figure 3: Effect of Stochastic Variance on Model Selection-SECM

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

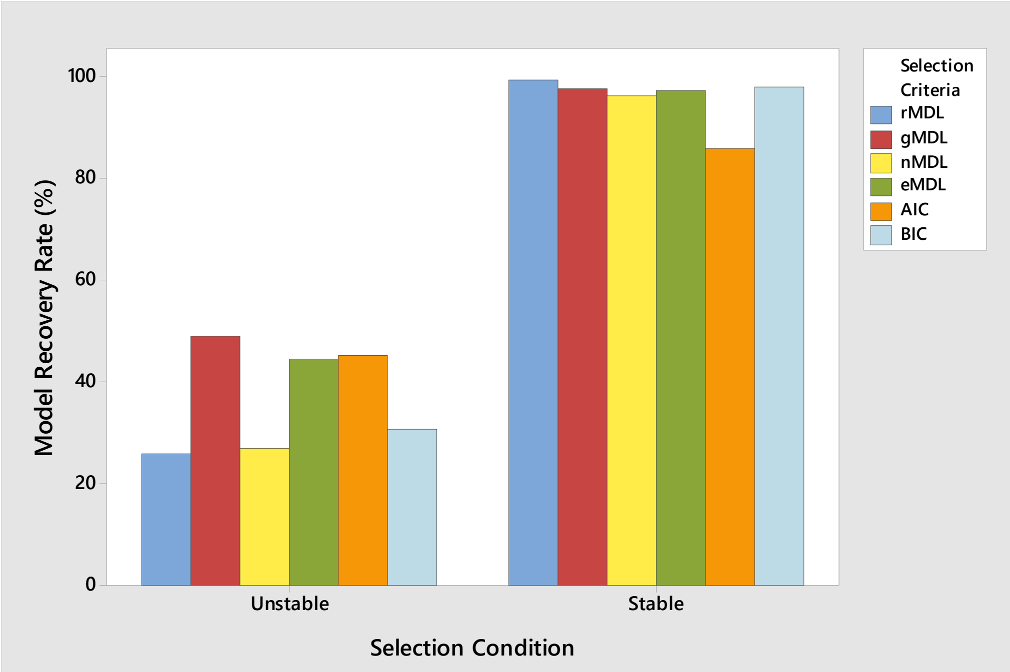


Figure 4: Effect of Stochastic Variance and Sample Size on Model Selection-SECM

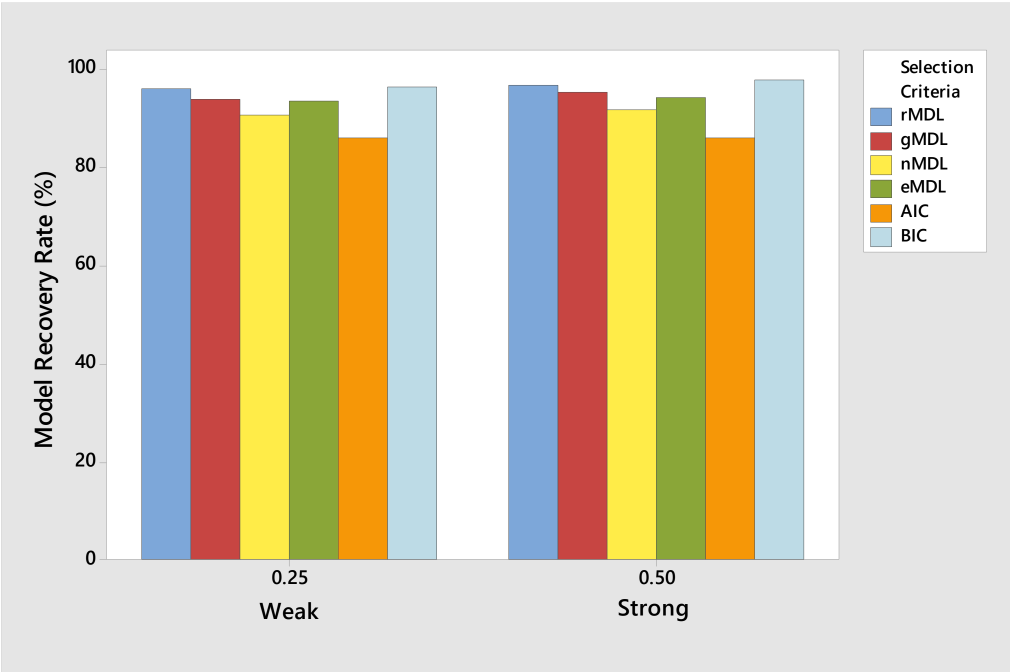


Figure 5: Effect of Asymmetric Adjustment Parameters on Model Selection (SECM for n = 150)

Digitized by Sam Jonah Library

© University of Cape Coast https://erl.ucc.edu.gh/jspui

Appendix L2: Graphical representation of model selection criteria for CECM analysis

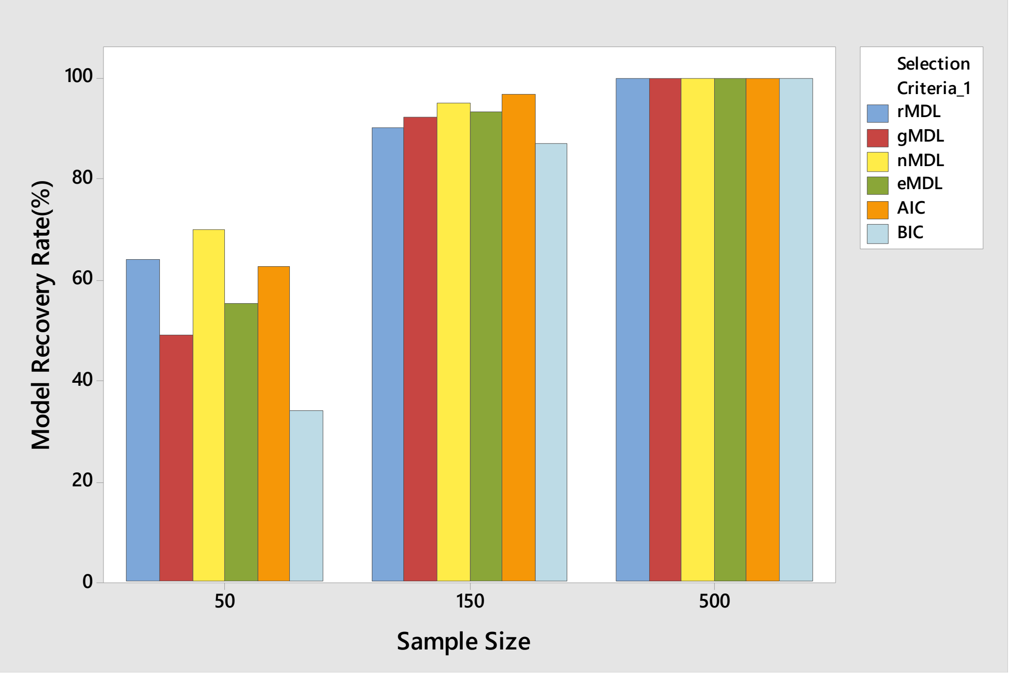


Figure 6: Effect of Sample Size on Model Selection-CECM

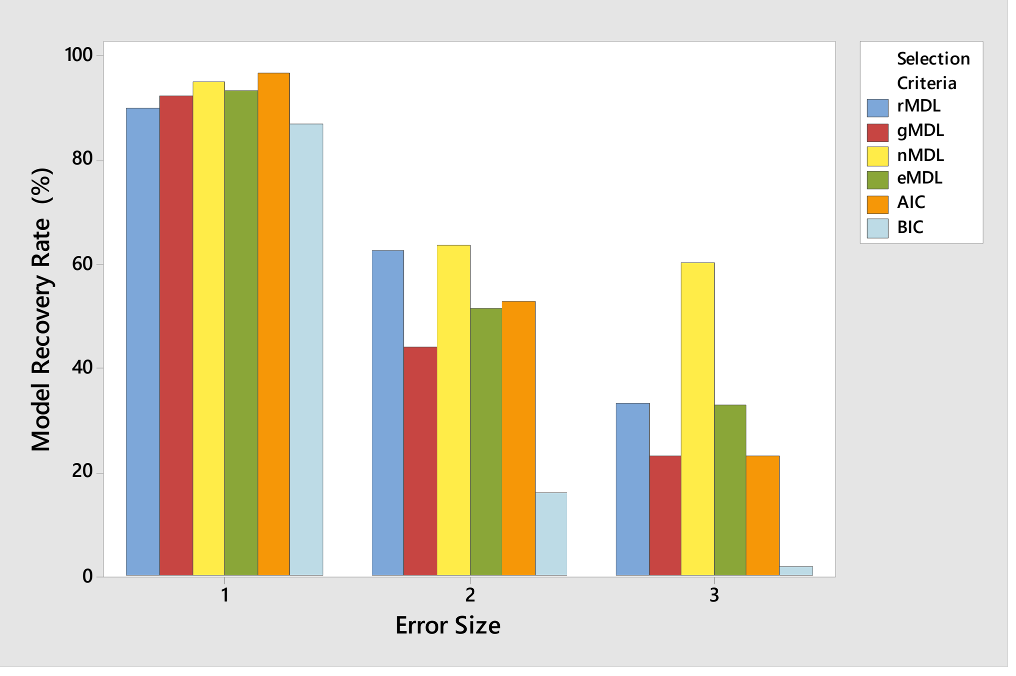


Figure 7: Effect of Stochastic Variance on Model Selection-CECM

Digitized by Sam Jonah Library

© University of Cape Coast   https://erl.ucc.edu.gh/jspui

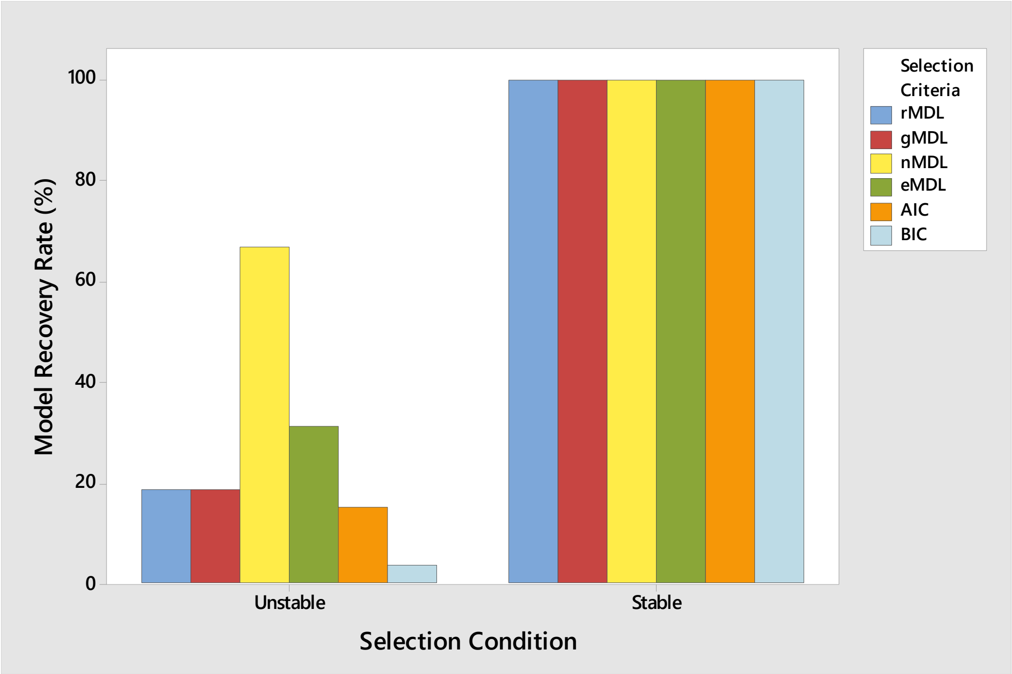


Figure 8: Effect of Stochastic Variance and Sample Size on Model Selection-CECM

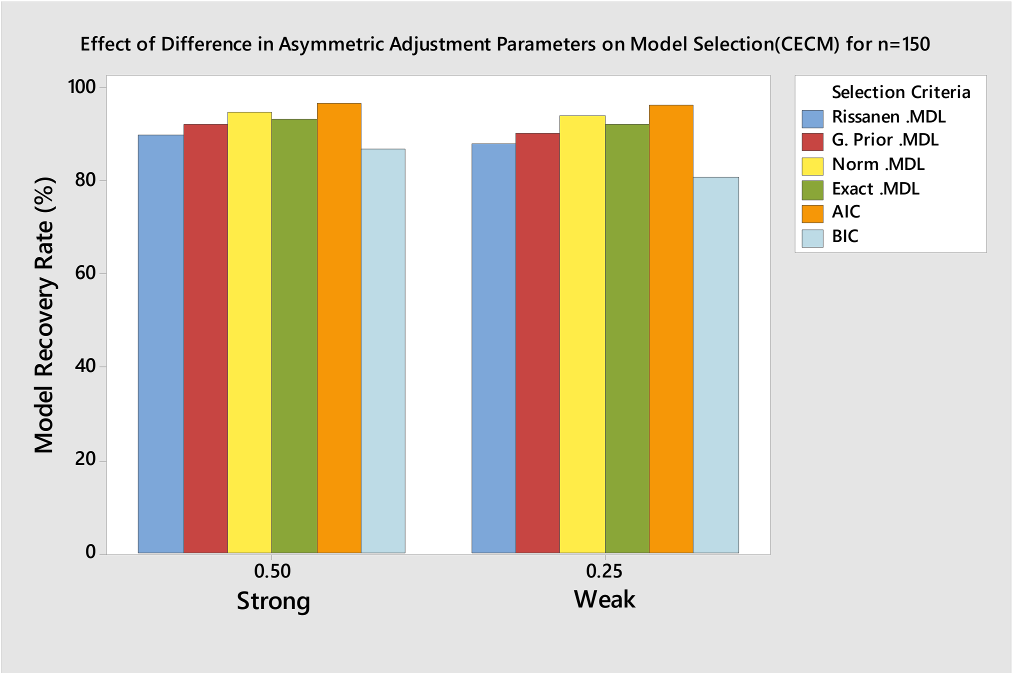


Figure 9: Effect of Asymmetric Adjustment Parameters on Model Selection (CECM for n = 150)

Digitized by Sam Jonah Library