An Empirical Study of Implied Volatility in Australian Index Option Markets

by

Qianqian Yang

Bachelor of Science (Mathematics, XMU)

A thesis submitted as partial requirement for the degree of

Master of Business (Research)

School of Economics and Finance

Faculty of Business

Queensland University of Technology

November 2006
Acknowledgements

I would like to express my gratitude and appreciation to all the scholarly staffs who deserve a mention for their assistance during the course of this thesis.

Firstly, I would like to give my appreciation to my supervisor Dr. Steven Li for proposing such an interesting and intellectually stimulating research topic, and for all his support, advice and encouragement throughout the year as I would not have been able to complete this thesis without his guidance.

Secondly, to Mr. Mark Christenson, I would like to express my gratitude for all his kindly help and precious time.

Thirdly, to Charles Wong, Jason Hay, Rany Thach, Kohei Marumo and other colleagues, thank you for all your patience and time in helping me with my questions.

Finally, a special thanks to my family for all their love and encouragement throughout the time when I study abroad. I would like to take this opportunity to say, “I love you, Daddy!”
Statement of Original Authorship

The work contained in this thesis has not been previously submitted for a degree for diploma at any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Signed: Qianqian Yang

Date: 7th Nov, 2006
Abstract

With the rapid development of option markets throughout the world, option pricing has become an important field in financial engineering. Among a variety of option pricing models, volatility of underlying asset is associated with risk and uncertainty, and hence is treated as one of the key factors affecting the price of an option. In particular, in the framework of the Black-Scholes option pricing model, volatility of the underlying stock is the only unobservable variable, and has attracted a large amount of attention of both academics and practitioners. This thesis is concerned with the implied volatility in the Australian index option market. Two interesting problems are examined.

First, the relation between implied volatility and subsequently realized volatility is investigated by using the S&P/ASX 200 (XJO) index options over a five-year period from April 2001 to March 2006. Unlike the S&P 100 index options in the US market, the XJO index options are traded infrequently, in low volumes, and with a long maturity cycle. This implies that the errors-in-variable problem for the measurement of implied volatility is more likely to exist. After accounting for this problem by the instrumental variable method, it is found that both call and put options implied volatilities are nearly unbiased and superior to historical volatility in forecasting future realized volatility.

Second, the volatility structure implied by the XJO index options is examined during the period from April 2001 to June 2005. The volatility structure with
respect to moneyness and time to maturity are investigated for both call and put option price series. It is found that the volatility smile largely exists, with call (put) option implied volatilities decreasing monotonically as the call (put) goes deeper out of the money (in the money). This result is consistent with the well-documented evidence of volatility smile on other index options since the stock market crash of 1987.

In summary, this thesis presents some important findings on the volatility inferred from the XJO index options traded on the ASX.
Table of Contents

AN EMPIRICAL STUDY OF IMPLIED VOLATILITY IN AUSTRALIAN INDEX OPTION MARKETS ............................................................................................................................................................................ I

ACKNOWLEDGEMENTS .................................................................................................................................................................................................................... II

STATEMENT OF ORIGINAL AUTHORSHIP .................................................................................................................................................................................. III

ABSTRACT ........................................................................................................................................................................................................................................ IV

TABLE OF CONTENTS ...................................................................................................................................................................................................... VI

LIST OF TABLES AND FIGURES ......................................................................................................................................................................................... VIII

CHAPTER 1 INTRODUCTION ............................................................................................................................................................................................ 1

1.1 BACKGROUND ............................................................................................................................................................................................................. 1

1.2 MOTIVATION OF THIS STUDY ............................................................................................................................................................................ 5

1.3 STRUCTURE OF THE THESIS ......................................................................................................................................................................... 6

CHAPTER 2 THE RELATION BETWEEN IMPLIED AND REALIZED VOLATILITY ............................................................................................................ 7

2.1 INTRODUCTION ........................................................................................................................................................................................................... 7

2.1.1 Background and rationale ........................................................................................................................................................................ 7

2.1.2 Statement of the research problem ....................................................................................................................................................... 10

2.1.3 Research methodology ........................................................................................................................................................................... 11

2.1.4 Structure ...................................................................................................................................................................................................... 11

2.2 LITERATURE REVIEW ..................................................................................................................................................................................... 13

2.2.1 Early studies based on cross-sectional data ............................................................................................................................................. 13

2.2.2 Later studies based on overlapping time-series data .................................................................................................................................. 18

2.2.3 Recent studies based on non-overlapping time series data .......................................................................................................................... 24

2.3 CONTRACT SPECIFICATIONS AND DATA ................................................................................................................................................... 30

2.3.1 Contracts ........................................................................................................................................................................................................ 30

2.3.2 Data .............................................................................................................................................................................................................. 33

2.4 METHODOLOGY ...................................................................................................................................................................................................... 38

2.4.1 Measurement of time to maturity .......................................................................................................................................................... 38

2.4.2 Measurement of IV .................................................................................................................................................................................. 39

2.4.3 Measurement of RV and HV ................................................................................................................................................................. 47

2.4.4 Regression tests ................................................................................................................................................................................................ 49

2.5 EMPIRICAL RESULTS ................................................................................................................................................................................................ 59

2.5.1 Descriptive statistics ............................................................................................................................................................................. 59

2.5.2 Conventional method – OLS estimates .............................................................................................................................................. 62

2.5.3 Alternative method – IV estimates ..................................................................................................................................................... 66

2.6 CONCLUSION ............................................................................................................................................................................................................. 71

CHAPTER 3 THE VOLATILITY STRUCTURE IMPLIED BY S&P/ASX 200 INDEX OPTIONS ........................................................................................................ 73

3.1 INTRODUCTION ........................................................................................................................................................................................................... 73

3.1.1 Background and rationale ........................................................................................................................................................................ 73

3.1.2 Statement of the research problem ....................................................................................................................................................... 75

3.1.3 Research methodology ........................................................................................................................................................................... 75

3.1.4 Structure ...................................................................................................................................................................................................... 75

3.2 LITERATURE REVIEW ..................................................................................................................................................................................... 77

3.2.1 Evidence on foreign currency options ............................................................................................................................................. 77

3.2.2 Evidence on stocks and stock index options ............................................................................................................................................ 79

3.2.3 Evidence on futures options ................................................................................................................................................................. 81

3.3 DATA AND METHODOLOGY ........................................................................................................................................................................ 84

3.3.1 Data .............................................................................................................................................................................................................. 84

3.3.2 Methodology .................................................................................................................................................................................................. 85

3.4 EMPIRICAL RESULTS ................................................................................................................................................................................................ 87

3.4.1 Volatility smile and sneer ................................................................................................................................................................. 87

An empirical study of implied volatility in ASX index option market ................................................................................................................................. Q. Yang
List of Tables and Figures

Table 2.1: Contract Specifications ................................................................. 32
Table 2.2: Methods for computing weighted implied volatility ......................... 44
Table 2.3: Descriptive statistics ..................................................................... 60
Figure 2.1: Level series of various volatilities .................................................. 60
Table 2.4: Information content of implied volatility: OLS estimates .................. 63
Table 2.5: Pairwise correlation Matrix ............................................................. 64
Table 2.6: Information content of implied call volatility: 2SLS estimates .......... 68
Figure 3.1: Graphs of call option implied volatility against moneyness (S/X-1) ... 88
Table 3.1: Data used to create graphs in Figure 3.1 .......................................... 89
Figure 3.2: Graphs of put option implied volatility against moneyness (X/S-1) ... 90
Table 3.2: Data used to create graphs in Figure 3.2 .......................................... 91
Figure 3.3: 3-D view of volatility structure for call options .............................. 94
Table 3.3: Group averages of call option implied volatilities ............................ 95
Figure 3.4: 3-D view of volatility structure for put options .............................. 96
Table 3.4: Group averages of put option implied volatilities ............................ 97
Chapter 1 Introduction

1.1 Background

One of the exciting developments in finance over the last 30 years has been the growth of options markets. An option gives the holder of the option the right to buy or sell an asset for a certain price. The holder does not have to exercise this right. For this reason, many hedgers and speculators find it more attractive to trade an option on an asset than to trade the asset itself. European and American options on stocks, stock indices, foreign currencies and futures contracts are currently actively traded in both the exchange-traded and over-the-counter markets. Regardless of where an option is traded, the central issue relating to option trading lies in the valuation of the option.

Since Black and Scholes published their seminal article on option pricing in 1973, there has been an explosion of theoretical and empirical work on option pricing. Among most option pricing models, volatility, associated with risk and uncertainty, has become one of the key inputs, and has deservedly attracted an amount of attention of both academics and practitioners. For example, models for pricing options under stochastic volatility have been developed in Hull and White (1987), Johnson and Shanno (1987), Scott (1987), and Wiggins (1987). In the field of empirical option pricing, many studies have concentrated on the volatility implied in the option’s market price based on an option pricing model.
In the framework of the Black-Scholes model, an equilibrium option price is a function of five variables: the current stock price, the time to maturity of the option, the exercise price, the risk free interest rate, and the volatility of the return of the underlying stock. Among these variables, volatility is the only unobservable one. Historical stock price data may be used to estimate the volatility parameter, which then can be plugged into the option pricing formula to derive option values. As an alternative, once an option price is observed from the market, one can solve the Black-Scholes formula backward to obtain the volatility implied in this option price. This volatility is called \textit{implied volatility}.

Since implied volatility is viewed as the market’s assessment of the underlying asset’s volatility over the option’s life, an interesting topic appears. That is to test the relationship between implied volatility and realized volatility. The null hypothesis of the test is that implied volatility is an \textit{unbiased} and \textit{informationally efficient} prediction of future realized volatility over the remaining life of the option. In addition, the relation between implied and realized volatility is a joint test of market efficiency and applicability of the option pricing model. Therefore, this topic is of great importance in the field of empirical study of option pricing model and hence has attracted a large amount of attention of researchers.

The first study to examine the predictive power of implied volatility is Latane and Rendleman (1976). In the light of this paper, Chiras and Manaster (1978) and Beckers (1981) use a broader sample of individual CBOE stock options to conduct static cross-sectional regressions. These earlier papers essentially
document that implied volatility is superior to historical volatility in forecasting future volatility. With cross-sectional data, one can only grasp the static relationship between implied and future realized volatility across a number of stocks during a single time period.

For this reason and with the availability of sufficient time series data, later studies have focused on the information content of implied volatility in dynamic settings. Such studies as Day and Lewis (1992) and Canina and Figlewski (1993) for S&P 100 index options, and Lamoureux and Lastrapes (1993) for CBOE stock options, examine whether the volatility implied in an option price predicts the realized volatility over the remaining life of the option. To this point, the time series literature seems to believe that volatility implied by option prices is biased and inefficient predictor of future volatility, and does not outperform volatility obtained from historical information.

In contrast to the previous studies, Jorion (1995), who studies currency options, and Fleming (1998), who studies S&P 100 index options, conclude that implied volatility is informationally efficient, although biased. Nevertheless, all of these later studies are based on overlapping datasets and suffer from the serial correlation problem. That is, historical volatility may contain part of information in the future realized volatility, which leads to overstating the explanatory power of historical volatility.

To address this problem, Christensen and Prabhala (1998) introduce a new sampling procedure which produces non-overlapping volatility series. That is,
exactly one implied-volatility is responding to one realized-volatility for each time period under consideration. Apart from the conventional OLS method, Christensen and Prabhala (1998) use the instrumental variable (IV) method to correct for the errors-in-variable (EIV) problem in the measurement of implied volatility, and conclude that implied volatility is an unbiased and informationally efficient forecast of future realized volatility.

Following the sampling procedure proposed by Christensen and Prabhala (1998), more recent studies, such as Hansen (2001) for Danish KFX index options, Christensen and Hansen (2002) for S&P 100 index options, Shu and Zhang (2003) for S&P 500 index options, and Szakmary et al (2003) for 35 futures options, view implied volatility as a better forecast of future realized volatility than historical volatility.

Apart from the relation between implied and realized volatility, another interesting problem in empirical option pricing is to investigate the volatility structure implied by option prices. Although the Black-Scholes (1973) model assumes the volatility of underlying asset to be constant, it is well known that, at any moment of time the implied volatility obtained by the Black-Scholes model varies across time to maturity as well as strike prices. The pattern of implied volatility for different time to maturity is known as the term structure of implied volatility, and the pattern across strike prices is known as the volatility smile or the volatility sneer. The term of volatility structure is used generally to refer to the pattern across both strike price and time to maturity.
Many early studies, such as Shastri and Wethyavivorn (1987), Fung and Hsieh (1991), Heynen (1993), Taylor and Xu (1994), and Sheikh (1991), have documented a U-shape smile pattern for implied volatility in many options markets prior to the stock market crash of 1987.

In contrast to the above theoretical and empirical results showing a symmetric pattern for implied volatility against strike price, Dumas, Fleming and Whaley (1998) illustrate that the volatility structure for S&P 500 index options has changed from the symmetric smile pattern to more of a sneer since the stock market crash of 1987. A J-shaped volatility pattern appears. This result has been extended to the SFE SPI 200 futures options by Brown (1999), with implied call (put) volatilities decreasing monotonically as the call (put) option goes deeper out-of-the-money (in-the-money).

1.2 Motivation of this study

Most existing studies on the two topics described above have concentrated on the most actively traded options in the U.S. markets, while options traded on the Australian markets have seldom been examined, despite an explosive growth in trading volume and availability of abundance of price data. The primary motivation of this study is to contribute to the literature by investigating (i) the relation between implied and realized volatility and (ii) the implied volatility structure in the context of S&P/ASX 200 (XJO) index options on the floor of Australian Stock Exchange (ASX).
1.3 Structure of the thesis

Chapter 1 briefly discusses the background, the motivation and the structure of this thesis.

Chapter 2 examines the relation between implied volatility and subsequently realized volatility over a five-year period from April 2001 through March 2006. Considering the errors-in-variable (EIV) problem in the measurement of implied volatility, both Ordinary Least Squares (OLS) method and Instrumental Variables (IV) method are employed, and the results from both methods are compared and analysed. Special attention is paid to the issues of how the EIV problem biases the OLS estimates and of why the consistent estimates can be obtained from the IV method by a 2SLS procedure.

Chapter 3 examines the volatility structure implied by the XJO index options over the period from April 2001 to June 2005. Two dimensional graphs are plotted to investigate the effect of volatility smile or sneer, i.e. how implied volatilities vary across different strike prices for options with the same time to maturity on a particular day in the data set. To examine how volatility smile or sneer is affected by time to maturity, three dimensional graphs of implied volatility against moneyness and maturity of the option are plotted. Then analyses of these figures and a comparison with the shapes in the previous studies are provided.

Chapter 4 provides a summary of this thesis, an overview of its contributions, and an outline of potential extensions for future research in this field.
Chapter 2 The Relation between Implied and Realized Volatility

2.1 Introduction

2.1.1 Background and rationale
An implied volatility is the volatility implied by an option price observed in the market based on an option pricing model. Assuming that the market is efficient and the option pricing model is valid, then implied volatility should be an unbiased and informationally efficient prediction of future realized volatility over the remaining life of the option. That is, implied volatility should subsume all the information content contained in all other variables used to explain future realized volatility. In addition, the relation between implied and realized volatility is a joint test of market efficiency and applicability of the option pricing model. Hence, the relation between implied and realized volatility has been an important research topic and many studies have been devoted to this topic.

The first study is done by Latane and Rendleman (1976). They use closing prices of options and stocks for 24 companies whose options are traded on the Chicago Board Options Exchange (CBOE) and conclude that implied volatility outperforms historical volatility in forecasting future realized volatility. After that, Chiras and Manaster (1978) and Beckers (1981) also obtain the same
conclusion based on a broader sample of the CBOE stock options. It should be noted that these studies concentrate on static cross-sectional relation rather than time-series forecasts.

Due to the availability of sufficient time series data, later studies have focused on testing the relation of implied volatility and realized volatility in a dynamic setting. Day and Lewis (1992) examine options on the S&P 100 index between 1983 and 1989, and report that implied volatility does not contain more information content of future realized volatility than historical volatility. However, this study ignores the term structure of volatility in the measurement of realized volatility, which is not matched with the remaining life of options.

Canina and Figlewski (1993) conduct the regression of realized volatility over the remaining life of the option on the implied volatility of the S&P 100 index options from 1983 to 1986. Surprisingly, they find that “implied volatility has virtually no correlation with future return volatility”. Lamoureux and Lastrapes (1993) examine options on individual stocks from 1982 to 1984, and find that information contained in historical volatility about future realized volatility is more than that contained in implied volatility. This result is also consistent with the result of Day and Lewis (1992).

Responding to the mixed conclusions in the previous studies of individual stock options and index options, Jorion (1995) points out that there are two possible explanations: one is that the test procedure is faulty; the other is that the option
markets are inefficient. In contrast with individual stock options and index options, he uses options on foreign currency futures traded on Chicago Mercantile Exchange (CME), and concludes that implied volatility is an efficient but biased forecast of future realized volatility.

Fleming (1998) examines the S&P 100 index options from 1985 to 1992, and indicates that the implied volatility is an upward biased forecast, but also that it contains more information regarding to future realized volatility than historical volatility.

All of these studies above are based on overlapping datasets and suffer from the serial correlation problem. That is, historical volatility may contain part of information in the future realized volatility, which leads to overstating the explanatory power of historical volatility.

To address this problem, Christensen and Prabhala (1998) introduce a new sampling procedure which produces non-overlapping volatility series. That is, exactly one implied-volatility is responding to one realized-volatility for each time period under consideration. With this sampling procedure and a longer volatility series from 1983 to 1995, they find that implied volatility of S&P 100 index (OEX) options outperforms historical volatility in predicting future realized volatility. These conclusions are further enhanced by Christensen and Hansen (2002) where implied volatility is constructed as a trade weighed average of implied volatilities from both OEX calls and puts.
Following the same sampling procedure as employed in Christensen and Prabhala (1998), a few studies in other options markets have been carried out and view implied volatility as a better predictor of future volatility of underlying assets. For instance, Hansen (2001) analyses the information content of options on the Danish KFX share index. This option market is very illiquid with infrequent trade and low volume. However, he finds that when the errors-in-variable problem is controlled by the instrumental variable technique, call implied volatility still contains more information about future volatility than historical volatility in such an illiquid option market. More recently, Shu and Zhang (2003) examine the options on S&P 500 index, and also report that implied volatility implied in the option prices outperforms the subsequently historical index return volatility. Szakmary et al (2003) examine 35 futures options markets across eight separate exchanges and find that for a large majority of the commodities studied, implied volatility is a better predictor of future realized volatility than historical volatility.

2.1.2 Statement of the research problem

The relationship between implied and realized volatility is an important topic in empirical studies and as noted above, many studies have been devoted to this topic. However, most existing studies have focused on the options in the US market and tend to conclude that implied volatility outperforms historical volatility as a predictor of the subsequently realized volatility over the remaining life of the option. To the authors’ best knowledge, no such investigation has been carried out for the S&P/ASX 200 index options on the ASX. Chapter 2 of the
thesis aims to contribute to the literature by investigating the relation between implied and realized volatility in an Australian setting.

2.1.3 Research methodology

Firstly, to avoid the overlapping of data, this study follows the sampling procedure presented by Christensen and Prabhala (1998). This approach results in that exactly one implied-volatility is responding to one realized-volatility for each time period under consideration.

Secondly, two types of estimation methods are conducted to assess the relation between realized and implied volatility. One is the conventional analysis, namely Ordinary Least Squares (OLS) method. But the OLS estimates might be biased and inconsistent. This is because the S&P/ASX 200 index options are traded infrequently, in low volumes, and with long maturity cycle, relative to the stock index options such as the S&P 100 index options in the US market. Hence, the errors-in-variable (EIV) problem for the measurement of implied volatility is more likely to exist. The presence of the EIV problem is then confirmed by a Hausman (1978) test. Therefore, to solve the EIV problem, another estimation method is implemented, which is called Instrumental Variables (IV) method. The analyses are based on log-volatility series throughout this chapter.

2.1.4 Structure

The rest of Chapter 2 is organized as follows. Section 2.2 provides a literature review of the previous studies on the relation between implied and realized
volatility, including the studies based on static cross-sectional data, and studies based on overlapping and non-overlapping time-series data.

Section 2.3 describes the contracts and data used in this chapter. Introductions of the Australia Stock Exchange (ASX), the S&P/ASX 200 index, and the contract specifications of the XJO index options and the SFE SPI 200 Futures are provided. Then, the data used in this chapter are described, with a special attention given to the sampling criteria and procedure.

Section 2.4 presents the methodologies used in this chapter. Firstly, the measurements of implied volatility, realized volatility and historical volatility will be discussed in turn. Particular attention is given to the measurement errors of implied volatility. After that, two estimation methods, namely the Ordinary Least Squared (OLS) method and the Instrument Variables (IV) method, will be described. Additionally, the Hausman (1978) test will also be presented in this section, which is used to prove the existence of the errors-in-variable (EIV) problem.

Section 2.5 provides the analysis of empirical results. Two estimation methods will be used and the results will be presented and analysed.

Finally, a summary and the conclusions of Chapter 2 are presented in Section 2.6.
2.2 Literature review

This section aims to present a thorough and deep review of the literature investigating the issues of whether implied volatility predicts future realized volatility and whether it does so efficiently. The literature can be classified into three categories: early studies based on cross-sectional data, later studies based on overlapping time-series data, and recent studies based on non-overlapping time-series data.

2.2.1 Early studies based on cross-sectional data

2.2.1.1 Latane and Rendleman (1976)

It is important to begin reviewing the literature with the first insight into the relation between implied and realized volatility. Using closing prices of options and stocks for 24 companies whose options were traded on the Chicago Board Options Exchange (CBOE), Latane and Rendleman (1976) established four measures to describe the standard deviation of underlying stock returns. The first one was the weighted average standard deviation (WISD). Since each stock had several traded options, different implied standard deviations (ISDs) were derived by numerically solving the basic Black and Scholes (1973) model for each stock. To get a single estimate of standard deviation, different weights were given to

\[ WISD_{it} = \left[ \sum_{j=1}^{N} ISD_{ijt}^2 \cdot d_{ijt} \right]^{\frac{1}{2}} \cdot \left[ \sum_{j=1}^{N} d_{ijt} \right]^{-1} \]

where \( WISD_{it} \) = Weighted average implied standard deviation for company \( i \) in period \( t, i=1 \) to 24, \( t=1 \) to 39,

\( ISD_{ijt} \) = Implied standard deviation for option \( j \) of company \( i \) in period \( t, N \) denotes the number of options analysed for company \( I \) and is always greater than or equal to 2,
options with different strike prices and different times to maturity. Then this
WISD was computed for each of the 24 companies over a sample period of 38
weeks between October 5, 1973 and June 28, 1974. The other three measures of
standard deviations of underlying stock returns were actual return volatilities but
over different periods. One measure was calculated over a four-year period
ending September 30, 1973 for each company, which could be seen as the
historical volatility. Another measure was computed over the same time period
with WISD for each company, which could be recognized as the concurrent
realized volatility. The last measure was computed over a two-year period ending
March 31, 1975 for each company, which was partially extended to the future
and hence could be viewed as the subsequent realized volatility.

To investigate the relation between WISD and actual return volatility, Latane and
Rendleman (1976) conducted the cross-sectional correlations among these four
measures and found that the highest degree of correlation is between WISD and
the subsequent realized volatility. Hence, Latane and Rendleman (1976)
concluded that WISD outperforms historical volatility in forecasting future
volatility of underlying stock returns.

Nevertheless, there are a number of shortcomings in Latane and Rendleman
(1976). Firstly, this study does not consider the dividend payments of the
individual stocks of these 24 companies during the sample period. In fact, thr
dividends on some stocks may be substantial and can have a significant effect on
the valuation of the options if such payments are made during the life of the

\[ d_{ijt} = \text{Partial derivation of the price of option } j \text{ of company } i \text{ in period } t \text{ with respect to its implied standard deviation using the Black-Scholes model.} \]
options. Secondly, this study does not consider the possibility of early exercise, since stock options traded on CBOE are American style. Thirdly, due to the different closing times in stock market and option market, the closing prices of stocks and options used in this study may not be synchronous. Fourthly, this study only calculates and compares the cross-sectional correlations between those four volatility measures, which can not reflect the forecasting power in dynamic settings. Finally, as pointed by Chiras (1977), ‘weighted average’ in Latane and Rendleman (1976) is not truly a weighted average since the sum of the weights is less than one. Therefore, the WISD is downward biased to zero. Moreover, the bias increases with an increase in the sample size even if the ISDs calculated for each option are the same.

### 2.2.1.2 Chiras and Manaster (1978)

Based on a broader sample of CBOE individual stock options during 23 months beginning June 1973 and ending April 1975, Chiras and Manaster (1978) compared the cross-sectional information content of their weighted implied standard deviation (WISD) with that of historical volatility, and found that the

\[ WISD = \sum_{j=1}^{N} ISD_j \frac{\partial W_j}{\partial \nu_j} \frac{\nu_j}{W_j} \sum_{j=1}^{N} \frac{\partial W_j}{\partial \nu_j} \frac{\nu_j}{W_j}, \]

where

- \( N \) = the number of options recorded on a particular stock for the observation date,
- \( WISD \) = the weighted implied standard deviation for a particular stock on the observation date,
- \( ISD_j \) = the implied standard deviation of option \( j \) for the stock,
- \( \frac{\partial W_j}{\partial \nu_j} \frac{\nu_j}{W_j} \) = the price elasticity of option \( j \) with respect to its implied standard deviation (\( \nu \)).

---

2 The contract specifications of equity option are listed in [http://www.cboe.com](http://www.cboe.com).
3 The weighing system in Chiras and Manaster (1978) is given by
WISD was superior to historical volatility in forecasting future realized volatility of stock returns. In particular, the predictive ability for the WISD was 32 percent on average and this value dramatically increased to 39 percent after March 1974, as compared to 21 percent in the prior period. Furthermore, historical volatility in isolation explained approximately 26 percent of the future volatility of stock returns, but after February 1974 it did not add any statistically significant information beyond that was already contained in the WISD. However, the WISD was a substantially biased forecast of cross-sectional stock volatility, with monthly slope coefficients ranging from 0.29 to 0.83.

Compared to Latane and Rendleman (1976), there are two improvements in Chiras and Manaster (1978). Firstly, they took into account the dividend payments during the option’s life and hence used the dividend adjusted Black-Scholes model, proposed by Merton (1973), to calculate the ISDs for each stock. Secondly, instead of running the correlation, they conducted several OLS regression tests, by which the information content and forecasting power of the WISD and historical volatility were illustrated more straightforwardly.

Although some improvements have been made, there are still many shortcomings remaining in Chiras and Manaster (1978). Firstly, although the Merton (1973) model allows inclusion of dividend data, it only can be used to price European style options. As noted in Section 2.2.1.1, individual stock options traded on the CBOE are American style. Therefore, using European style options model to value American style options is not appropriate since the possibility of early exercise should be considered in the pricing model. Secondly, the non-
synchronous problem in stock prices and option prices is also left in this study. Thirdly, the forecasting interval of implied volatility for any given option price is the remaining life of the option, thus, the future realized volatility should be calculated carefully to match the option’s remaining life. However, this study does not consider the various maturities for each stock option and hence does not match future realized volatility with different times to maturity. Finally, instead of assigning heavier weight to near-the-money options, Chiras and Manaster (1978) focused on the percentage change in the price of an option with respect to the percentage change in its ISD, and hence resulted in the heaviest weight falling on the deepest out-of-the-money call and put options.

2.2.1.3 Beckers (1981)

Using daily closing price data on 62-115 CBOE stock options over April 28, 1975 to July 22, 1977, Beckers (1981) investigated the predictive ability of implied volatility. He took into account the dividend problem and the problem of optimal weighting schemes of the volatilities. Three alternative measures for implied volatility were compared. The first one was the modified Latane-Rendleman approach, which distinguished between options on the same stock with different maturities since they have a different forecasting horizon\(^4\). The second one was his own method\(^5\), which put more weight on the at-the-money

\[f(ISD) = \sum_{i=1}^{I} w_i \left[ C_i - BS_i(ISD) \right]^2 \left/ \sum_{i=1}^{I} w_i \right.,\]

where

\(^4\) Latane and Rendleman (1976) ignored the various maturities of options on the same stock. The forecasting horizon of implied volatility could be different depending on the time to maturity of the option.

\(^5\) To obtain an estimate of implied volatility, Beckers (1981) used an alternative weighting scheme, which minimizes the following function:
options than Latane and Rendleman (1976) and Whaley (1982). The last approach was simply the volatility obtained from the at-the-money options.

After comparing the predictive power of the different measures, Beckers (1981) concluded that, in general, all implied volatility measures were superior to the quarterly historical volatility in forecasting the cross-sectional stock volatility. Furthermore, at-the-money implied volatility had at least as much information contents as other implied volatility measures. Hence, this result questioned the usefulness of a weighing scheme in computing implied volatilities. In addition, he also noted that implied volatility was biased and informationally inefficient, since historical volatility provided additional information for volatility forecasting.

2.2.2 Later studies based on overlapping time-series data

2.2.2.1 Day and Lewis (1992)

Previous studies of the information content of implied volatility from the market prices of call options have relied on the cross-sectional regression tests, whereas Day and Lewis (1992) started to look at the time series behaviour of implied volatility. The data they used were call options on the S&P 100 index for 319 weeks over the period from November 11, 1983 to December 28, 1989 (including the 1987 stock market crash). They compared implied volatility with

\[ C_i = \text{market price of option } i, \]
\[ BS_i = \text{Black-Scholes option price as a function of the ISD}, \]
\[ I = \text{total number of options on a given stock with the same maturity}, \]
\[ w_i = \text{weight for the } i\text{th option } = \frac{\delta BS_i(ISD)}{\delta ISD}. \]
the weekly volatility estimates from the GARCH (Generalized Autoregressive Conditional Heteroscedasticity) and the Exponential GARCH models, and concluded that implied volatility was informative and unbiased predictor of future index return volatility. However, they also concluded that the GARCH and the EGARCH forecasts contained incremental information content which was not captured by implied volatility. This implied that implied volatility was not an informationally efficient forecast of future volatility.

There are two main advantages in the study by Day and Lewis (1992). Firstly, they accounted for the dividend problem in the Black-Scholes formula by reducing the current stock price by the present value of the dividends to be paid over the life of the call option. Secondly, since they placed the greatest weight on the at-the-money options and eliminated the deep-in- and deep-out-of-money options from their sample, their estimation procedure tends to minimize the measurement errors caused by either non-synchronous trading or the size of the bid-ask spread.

However, their results are subject to the criticism due to the use of overlapping data and the maturity mismatch between realized and implied volatility. In particular, they calculated implied volatility from the options with the shortest expiration traded at the beginning of the week, while estimated the volatility from the GARCH and the EGARCH models based on Wednesday to Wednesday and Friday to Friday return intervals.
2.2.2.2 **Canina and Figlewski (1993)**

Canina and Figlewski (1993) examined the information content of implied volatility from the closing prices of S&P 100 index call options over March 15, 1983 to March 28, 1987. They performed a regression of realized volatility over the remaining life of the option on the corresponding implied volatility. Rather startlingly, they found that implied volatilities from options across various maturities and moneynesses have ‘no statistically significant correlation with realized volatility at all’ and appear to be even worse than the 60-day historical volatility to forecast future S&P 100 index volatility. They attributed their results either to the options market’s inefficiency or to the applicability of the Black-Scholes formula.

However, as pointed out by Christensen and Prabhala (1998), the S&P 100 index option market is the most active option market in the U.S. and hence it is less likely that the option market processes volatility information inefficiently. Furthermore, although some assumptions under the Black-Scholes option pricing formula, such as no transaction costs, continuous trading, unlimited arbitrage and so on, conflict with the real market, those frictions could not explain the apparent failure of the Black-Scholes pricing model for the S&P 100 index options. Hence, Christensen and Prabhala (1998) suggested that the results reported by Canina and Figlewski (1993) seem to be the outcome of their extremely overlapping sampling procedure, which leads to highly autocorrelated errors and inaccurate and potentially inconsistent regression estimates.
2.2.2.3 Lamoureux and Lastrapes (1993)

Under the framework of Hull and White’s (1987) stochastic volatility option pricing model, Lamoureux and Lastrapes (1993) investigated implied volatility from at-the-money CBOE call options on 10 non-dividend paying stocks for the period between April 19, 1982 and March 31, 1984, and compared the 1-day and option-lifetime volatility forecasts with those from GARCH and historical volatility estimates. Using both in-sample and out-of-sample tests, they concluded that implied volatility was biased but informative, and that historical volatility contributed additional information in forecasting future stock-return volatility. These results were consistent with those in Day and Lewis (1992). However, both studies were based on overlapping samples and characterized by the maturity mismatch problem between realized and implied volatility.

2.2.2.4 Jorion (1995)

Responding to the conclusions in previous studies that implied volatility was biased and inefficient forecast of future volatility, Jorion (1995) pointed out that there are two possible explanations: one is that the test procedure is faulty; the other is that the option markets are inefficient. In contrast with individual stock options and stock index options, Jorion (1995) examined the information content and predictive power of implied volatility from foreign currency futures options on Chicago Mercantile Exchange (CME). The data he used were daily closing prices of options on the German deutsche mark (DM) futures, the Japanese yen (JY) futures, and the Swiss franc (SF) futures, over January 1985 to February 1992. He concluded that implied volatility was close to an unbiased forecast of the volatility over the next day, but was a more biased forecast of the volatility.
over the remaining life of the option. In both cases, implied volatility was informationally efficient, i.e. 20-day historical volatility and GARCH-based volatility assessments contributed no additional information beyond that was already contained in implied volatility. He also provided simulations to fully account for small-sample biases usually ignored in regressions of predictive ability, and showed that measurement errors can substantially distort inferences.

2.2.2.5 Fleming (1998)

Fleming (1998) examined the performance of implied volatility as a forecast of future stock price volatility, using daily transaction data on the S&P 100 index options over October 1985 to April 1992, excluding the October 1987 stock market crash period. The implied volatility used in this study was estimated from the modified binomial model of Fleming and Whaley (1994). To avoid spurious regression estimates caused by high serial correlation, Fleming (1998) utilised the GMM approach of Hansen (1982) to evaluate the unbiasedness and efficiency of implied volatility of S&P 100 index. He found that the implied volatility was a biased but substantially informative forecast of future index volatility, and that implied volatility was informationally efficient relative to other estimates of volatility such as 28-day historical volatility and GARCH(1,1) forecast. In contrast to the previous evidence for S&P 100 index options, both the biasness and informationally efficiency results were consistent with the evidence of Jorion (1995) for options on foreign currency futures.
2.2.2.6  Gemmill (1986, 1993)

Gemmill (1986, 1993) are the first two studies on information content of implied volatility using U.K. data. Gemmill (1986) examined the implied volatility from equity options, and reported that in-the-money and at-the-money options produce the most accurate implied volatility measures but a historical volatility also performs well. Implied volatilities did not contain all the information available in historical volatility, indicating implied volatility was not an informationally efficient forecast of future volatility. Gemmill (1993) compared implied volatility from at-the-money FTSE100 options’ prices with another volatility measure from an autoregressive random variance model, and concluded that implied volatility contains information about future volatility beyond that in the time-series forecast.

2.2.2.7  Gwilym and Buckle (1999)

Gwilym and Buckle (1999) is another study investigating the performance of implied volatilities as a forecast of future volatility in the U.K. option market. In particular, using daily prices of at-the-money FTSE100 index options on the London International Financial Futures and Options Exchange (LIFFE) over the period from 21 June, 1993 to 19 May, 1995, Gwilym and Buckle (1999) found that implied volatility, albeit upwardly biased, contains more information of future realized volatility than historical volatility. They also found that the forecast accuracy is better for longer forecast horizons for all measures of implied volatilities and historical volatilities. They explained this result by the fact that as the option expiry date approaches, the realized volatility series becomes itself more volatile.
2.2.3 Recent studies based on non-overlapping time series data

2.2.3.1 Christensen and Prabhala (1998)

Using daily closing prices of at-the-money S&P 100 index call options over a period of 139 months from November 1983 to May 1995, Christensen and Prabhala (1998) examined the ability of implied volatility to forecast future realized volatility over the remaining life of the option. Responding to the mixed conclusions in the previous studies based on overlapping samples, they proposed an ad hoc sampling procedure to construct non-overlapping volatility series. That is, exactly one implied volatility is corresponding to one realized volatility for each time period under consideration. They believed that this non-overlapping sample yields more reliable regression estimates relative to less precise and potentially inconsistent estimates obtained from overlapping samples used in previous work.

Based on their non-overlapping sample, Christensen and Prabhala (1998) accounted for the errors-in-variable problem afflicting the measurement of implied volatility by using the Instrumental Variables method. They concluded that implied volatility is an unbiased and efficient forecast of future volatility, i.e. implied volatility subsumes all the information in the option market to forecast future volatility whereas historical volatility does not contribute incremental information beyond that already exists in implied volatility. They also documented that there was a regime shift around the October 1987 stock market crash period, which helps to explain the biasness of implied volatility in previous work.
2.2.3.2 Hansen (2001)

Following the sampling procedure proposed by Christensen and Prabhala (1998), Hansen (2001) investigated the information content of implied volatility in Danish option market, which is illiquid relative to the U.S. and U.K. index options markets. Using daily closing prices of at-the-money Danish KFX index options over 52 months from September 1995 to December 1999, she found that implied volatility seems to be a biased and inefficient forecast of subsequent realized volatility based on the OLS regression results. However, after controlling for the errors-in-variables problem via instrumental variables techniques, she concluded that implied call volatility is an unbiased and informationally efficient forecast of subsequent realized volatility, while implied put volatility is a biased forecast and even does not outperform historical volatility to predict future index volatility.

Hansen (2001) also documented that the measurement errors, except those caused by misspecification of the Black-Scholes model, could be reduced by combining implied call and put volatility into a common implied volatility measure. Using instrumental variables method to control for the remaining errors, she extended the unbiasedness and informational efficiency results of the combined implied volatility measures to the Danish index option market despite its poor liquidity.
2.2.3.3 Christensen and Hansen (2002)

Christensen and Hansen (2002) examined the unbiasedness and informational efficiency of both call and put implied volatility from the S&P 100 index (OEX) options, over a more recent period from April 1993 to February 1997. The implied call and put volatilities in their study were collected from DataStream, where implied volatility is constructed as a traded weighted average of the individual Black (1976) implied volatilities of all near term call or put OEX options, traded over the last five days. According to the OLS regression results, Christensen and Hansen (2002) concluded that both implied call and put volatilities are unbiased and efficient forecasts of future volatility, however, a slightly higher adjusted-$R^2$ associated with implied call volatility indicates that implied call volatility is a better volatility forecast than implied put volatility. To handle the possible measurement errors problem, they utilised a full information maximum likelihood (FIML) method to estimate a three-equation model as a simultaneous system, and confirmed the conclusions from the previous OLS regressions that both implied call and put volatilities are unbiased and efficient forecasts of future volatility.

---

6 The weighting system in Datastream’s implied volatility is given by

$$\sigma_t = \frac{\sum_d \sum_i \sigma_{idt} N_{dt}}{\sum_d \sum_i N_{dt}}$$

where

$s$ = the number of series with one month to expiration

$N_{dt}$ = the number of trades $d$ days ago in series $t$

$\sigma_{idt}$ = the implied call or put volatility $d$ day ago on series $t$ calculated using Black's (1976) option pricing model
2.2.3.4 Shu and Zhang (2003)

Following the non-overlapping sampling procedure proposed by Christensen and Prabhala (1998), Shu and Zhang (2003) examined the relationship between realized and implied volatility by using daily closing prices of at-the-money S&P 500 index options traded on the CBOE over January 1, 1995 to December 9, 1999. Moreover, they examined how the measurement errors in realized volatility and implied volatility affect the stability of this relation between realized and implied volatility. For doing this, four different measures of realized volatility are computed, which are the standard deviation of daily return, the Parkinson (1980) extreme value volatility estimator, the Yang and Zhang (2000) range estimator, and the Andersen’s (2000) square root of intraday return squares estimator. Then implied volatility derived from the Black-Scholes formula is compared with that from the Heston (1993) stochastic volatility option pricing model.

After regressing realized volatility on implied volatility in isolation and with combination of historical volatility, Shu and Zhang (2003) found that implied volatility is a biased but informationally efficient forecast of subsequent realized volatility, i.e., implied volatility outperforms historical volatility in forecasting future volatility since historical volatility does not contribute additional information. They also concluded that this relation is stable under various measures of realized volatility and implied volatility. In particular, the measurement of realized volatility by intraday 5-minute return significantly improves the predictive ability of implied volatility. Implied volatility computed...
from the Black-Scholes model has higher explanatory power than that from the Heston (1993) stochastic volatility model.

2.2.3.5  Szakmary et al (2003)

Using data from 35 futures options markets\(^7\) on eight separate exchanges\(^8\), Szakmary et al (2003) examined the power of implied volatility in forecasting subsequently realized volatility. They found that for this broad array of futures options, implied volatility performs well in a relative sense. For a large majority of the commodities studied, the implied volatility, albeit not completely unbiased, outperforms historical volatility as a predictor of the subsequently realized volatility in the underlying futures prices over the remaining life of the option. In particular, in most markets examined, regardless of whether it is modelled as a simple moving average or in a GARCH framework, historical volatility does not contribute incremental information beyond what is already impounded in implied volatility. Hence, implied volatilities in most futures options markets are informationally efficient.


\(^7\) These options are written on a broad array of underlying futures contracts, including equity-index, interest rate, currency, energy, metals, agriculture and livestock futures.

\(^8\) They are Chicago Board of Trade (CBOT); Chicago Mercantile Exchange (CME); Coffee, Sugar, Cocoa Exchange (CSCE); Commodity Exchange (COMEX); London International Financial Futures Exchange (LIFFE); Marché à Terme International De France (MATIF); New York Mercantile (NYM); and New York Cotton Exchange (NYCE).
and the underlying futures contracts are traded on the same exchange, and hence their closing prices are less likely to be afflicted by the non-synchronous trading problem, and transactions costs are relatively low.
2.3 Contract specifications and Data

This section describes the contracts and data utilised in Chapter 2. In the first subsection, introductions of the Australia Stock Exchange (ASX), the S&P/ASX 200 index, and the contract specifications of the XJO options and the SFE SPI 200 Futures are provided. A brief explanation on the reason why these contracts were chosen is given. In the second subsection, the data used in this chapter are described, with a special attention given to the sampling criteria and procedure.

2.3.1 Contracts

2.3.1.1 Australian Stock Exchange (ASX)

There are two exchanges in Australia, namely, the Australian Stock Exchange (ASX) and the Sydney Futures Exchange (SFE). The ASX is the 12th largest share market in the world, and the second largest in the Asia Pacific region on the basis of market capitalisation. Apart from individual stocks, the ASX is also an exchange for trading index options, which are currently written on the S&P/ASX 200 Index (XJO)\(^9\), the S&P/ASX 50 Index (XFL) and the S&P/ASX 200 Property Trusts Index (XPJ). The S&P/ASX 200 index options were chosen for this thesis because they are the most popular and liquid index options on the ASX. The SFE is the tenth largest financial futures and options exchange in the world, and the second largest in the Asia Pacific region by volume turnover. The SPI 200 Futures contracts are traded on the SFE. These contracts were chosen for the consideration of the dividend adjustment in the option pricing model.

\(^9\) XJO, XFL and XPJ are the security code for index options on the ASX. (see www.asx.com.au)
2.3.1.2 **S&P/ASX 200 Index**

The S&P/ASX 200 index is the underlying index of both the S&P/ASX 200 index options contracts and the SFE SPI 200 index futures contracts. This index is composed of the S&P/ASX 100 plus an additional 100 stocks listed on the Australian stock market. It is recognized as the investable benchmark for the Australian stocks market. It represents approximately 90% of total market capitalisation of Australian equities. This index allows investment managers to benchmark against a portfolio with ample size and liquidity.

2.3.1.3 **S&P/ASX 200 Index Options**

The S&P/ASX 200 index options are traded on the ASX and written against the S&P/ASX 200 index. The listing date of these contracts was March 31, 2001. These options contracts are European in exercise style and with quarterly expiry cycle: March, June, September and December. The types of these contracts are call and put options. The exercise prices are set at intervals of 25 index points with new exercise prices automatically created as the underlying index oscillates. The expiration day is the third Thursday\(^{10}\) of the expiry month or the following business day when an expiry Thursday happens to be a public holiday. Trading will cease at 12 noon on expiry Thursday. This means trading will continue after the settlement price has been determined. The S&P/ASX 200 index options are cash settled. The settlement amount is based on the opening prices of the stocks in the underlying index on the morning of the last trading date\(^{11}\). The options are

---

\(^{10}\) It was the third Friday before September 2004.

\(^{11}\) As the stocks in the relevant index open, the first traded price of each stock is recorded. Once all stocks in the index have opened, an index calculation (the Opening Price Index Calculation (OPIC)) is made using these opening prices.
quoted in index point and each index point has a multiplier of AUD $10. These contracts can be traded during the normal trading hours between 6.00am and 5.00pm, and night trading hours between 5.30pm to 8.00pm\textsuperscript{12}.

Table 2.1: Contract Specifications

<table>
<thead>
<tr>
<th>Name</th>
<th>S&amp;P/ASX 200 Index Options</th>
<th>SFE SPI 200 Futures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underlying Index</td>
<td>S&amp;P/ASX 200 Index</td>
<td>Same</td>
</tr>
<tr>
<td>Exchange</td>
<td>Australian Stock Exchange ASX</td>
<td>Sydney Futures Exchange (SFE)</td>
</tr>
<tr>
<td>Multiplier</td>
<td>AUD$10 per index point</td>
<td>AUD$25 per index point</td>
</tr>
<tr>
<td>Listing date</td>
<td>31-03-2001</td>
<td>02-05-2000</td>
</tr>
<tr>
<td>Minimum price movement</td>
<td>One index point (A$10)</td>
<td>One index point (A$25)</td>
</tr>
<tr>
<td>Margin requirement</td>
<td>/</td>
<td>A$1750</td>
</tr>
<tr>
<td>Exercise intervals</td>
<td>25 index points</td>
<td>/</td>
</tr>
<tr>
<td>Exercise style</td>
<td>European</td>
<td>/</td>
</tr>
<tr>
<td>Settlement</td>
<td>Cash settled</td>
<td>Same</td>
</tr>
<tr>
<td>Contract months</td>
<td>March, June, September, December</td>
<td>March, June, September, December out to six quarter months</td>
</tr>
<tr>
<td>Expiry day</td>
<td>The third Thursday of the settlement month</td>
<td>Same</td>
</tr>
<tr>
<td>Last trading day</td>
<td>Trading in expiring contracts ceases at 12 noon on expiry Thursday</td>
<td>Same</td>
</tr>
<tr>
<td>Trading hours</td>
<td>6.00am to 5.00pm and 5.30pm to 8.00pm</td>
<td>9.50am to 4.30pm and 5.10pm to 7.00am</td>
</tr>
<tr>
<td>Settlement day</td>
<td>The first business day following the last trading day</td>
<td>Same</td>
</tr>
</tbody>
</table>

Note: All the information in the table are obtained from ASX website: www.asx.com.au

2.3.1.4 SFE SPI 200 Index Futures

The SFE SPI 200 index futures contracts are traded on the SFE written on the S&P/ASX 200 index. These contracts were listed on May 2, 2000. They are marked-to-market at the end of each day and valued at AUD $25 per index point. Trading in these contracts ceases at 12.00pm on the third Thursday of the

\textsuperscript{12} Unless otherwise indicated, all times are Sydney local times.
settlement month and the contracts have March, June, September and December month cycles out to six quarter months. The contracts are cash settled on the first business day after expiry. These contracts can be traded during the normal trading hours between 9.50am – 4.30pm and overnight from 5.10pm – 7.00am (During US daylight saving time)\(^{13}\). Table 2.1 provides the contract specifications for both options and futures on S&P/ASX 200 index.

### 2.3.2 Data

#### 2.3.2.1 Data description

Chapter 2 of this thesis is concerned with the S&P/ASX 200 index options (XJO) traded on the Australian Stock Exchange (ASX) during a five-year period from April 2, 2001 to March 16, 2006.

Proceeding April 2, 2001, there have been a number of changes in the index options market on the ASX. On November 15, 1985, the ASX first listed options on the All Ordinaries Index (XAO), which is the main benchmark for securities trading on the ASX. On April 3, 2000, the ASX, in association with S&P, introduced the S&P/ASX 200 index as the new benchmark for the Australian stock market. Consequently, the underlying index for the ASX index options also switched from the All Ordinaries Index to the S&P/ASX 200 index on April 3, 2000. However, during the period from April 3, 2000 and March 31, 2001, a continuation of the former All Ordinaries Index was calculated and disseminated.

\(^{13}\)US daylight saving begins first Sunday in April and ends last Sunday in October. During US non daylight saving time, the trading hours would be 5.10pm-8.00am and 9.50am-4.30pm. All times are Sydney local times.
by the ASX to allow for the maturity of futures contracts based on the superseded All Ordinaries Index. During this period the ASX listed index options on the All Ordinaries Index. Thus, from March 31, 2001, the S&P/ASX 200 index has been formally used as the underlying asset of index options on the ASX. Additionally, for the reason of thin trading, the All Ordinaries Index options were delisted twice\textsuperscript{14} and finally relaunched until November 8, 1999.

Given these historical changes and research purpose of this study, the data of S&P/ASX 200 index options over a five-year period from April, 2001 to March, 2006 is used. The period before April, 2001 is omitted to avoid possible excess market movements due to the change in the underlying index and the infrequent trading in an emerging index option market. The daily index options data are provided by the ASX, consisting of trading date, expiration date, close price, strike price and trading volume for each trading option.

The corresponding daily close index levels are obtained from Bloomberg. Australian 90-day Bank Accepted Bill (BAB) rate is used as a proxy for the risk-free interest rate $r$, whose maturity most closely matches the option’s life. The interest rate data are obtained from the Reserve Bank of Australia (RBA). For the purpose of estimating dividends during the life of the options, daily closing prices of the SFE SPI 200 futures are also collected from Bloomberg.

\textsuperscript{14} The first time it was delisted on May 26, 1988 and relaunched on December 8, 1993. The second time it was delisted on September 29, 1996 and relaunched until November 8, 1999.
2.3.2.2  **Sampling procedure**

To obtain a relatively accurate measurement of implied volatility, the options must be chosen carefully. The following sampling criteria are applied:

(1) Options must be collected on the business day immediately following the expiry date and this option must expire on the next expiry date

(2) Options must be in the range of $S_t / X_t \in (0.95, 1.05)$, where $S_t$ is the index level and $X_t$ is the exercise price of the option

(3) Options must be traded actively, i.e., a relatively high trading volume

Criterion (1) is used to avoid the overlapping of data\(^{15}\). The S&P/ASX 200 index options expire on the third Thursday of the settlement month: March, June, September and December, thereby four observations can be obtained each year for call and put options, respectively. Criterion (2) is used as the definition of at-the-money option in this thesis. Since options of different exercise prices have been known to produce different implied volatilities, a decision has to be made as to which of these implied volatilities should be used, or which weighing scheme should be adopted. The most common strategy is to choose the implied volatility derived from the at-the-money option based on the argument that at-the-money options are the most liquid and hence at-the-money implied volatility is less likely to have measurement errors. However, some options in the range of $S_t / X_t \in (0.95, 1.05)$ are thinly traded compared with others, thus Criterion (3) is

---

\(^{15}\) Christensen and Prabhala (1998), Hansen (2001), Christensen and Hansen (2002), and Shu and Zhang (2003) also use non-overlapping time series.
used to collect options with highest trading volume in the range of $S_t / X_t \in (0.95, 1.05)$.

Following the criteria described above, the sampling procedure is as follows. Let $t$ be the business day that immediately follows the first expiry date. On day $t$, the closing prices ($C_t$ and $P_t$) and strike prices ($X_{t,c}$ and $X_{t,p}$) are recorded for a call option and a put option, each of which expires on the next expiry date $t+1$ and has highest trading volume in the range of at-the-money options. The corresponding underlying index level ($S_t$) is also recorded. To estimate the dividend paid during the life of an option, the closing price of the SFE SPI 200 futures ($F_t$) is also recorded on day $t$. The next call option is sampled on the business day that immediately follows the second expiry date. This sampling date is labelled $t+1$. Similarly the values of $C_{t+1}, P_{t+1}, X_{t+1,c}, X_{t+1,p}, S_{t+1}, F_{t+1}$ are recorded. The whole sequence of call and put options contracts is constructed in this manner.

Due to the fact that the trades in the ASX options markets are not very active, a few complications appear in the collection of the data. In some instances, no options with three months to maturity are traded on the business day that immediately follows the expiry date, or the options traded on that day deviate too much from being at-the-money. In these cases, the options contracts which are close to at-the-money will be collected on the nearest following business day. In some other cases, it may be hard to find the suitable call and put options on the

---

16 The details of dividend estimation will be discussed in Section 2.4.2.
same day. Thus, to satisfy all the sampling requirements for both call and put options, the sampling time has to be postponed to the next closest business day. However, there is a trade-off between how close to at-the-money and how far from the last expiry date. It is worth to mention that these changes in the sampling date are acceptable, since the realized volatility will be calculated for the remaining life of the option. Hence, the maturity mismatch problem affecting many previous studies will not happen in this study.

Applying the above sampling criteria and procedure, 20 observations are obtained for each series of call and put options.
2.4 Methodology

This section presents the methodologies used in Chapter 2. Firstly, the measurements of time to maturity and three volatility series (i.e. implied-, realized-, and historical- volatility) will be discussed in turn. Particular attention is paid to the measurement errors of implied volatility. After that, two estimation methods, namely the Ordinary Least Squared (OLS) method and the Instrument Variables (IV) method, will be described. Additionally, a Hausman (1978) test will also be presented in this section, which is used to prove the existence of the errors-in-variable (EIV) problem.

2.4.1 Measurement of time to maturity

In practice, the time for paying interest is based on calendar days, while the time for the life of an option is based on trading days. French (1984) suggests that, when calculating the option’s price, time for paying interest and time for option’s life should be measured separately. That is,

\[
\text{Time for option's life} = \frac{\text{trading days until maturity}}{\text{trading days per year}},
\]

\[
\text{Time for interest paying} = \frac{\text{calendar days until maturity}}{\text{calendar days per year}}.
\]

Hull (2003) suggests that, however, in practice, these two measurements do not result in a big difference except for options with very short life. The option’s life in this study is about three months, thus the differentiation of these two
measurements is ignored and the following definition of time to maturity is used throughout the thesis.

Time to maturity, $T_t$, is measured by the number of the trading days between the day of trade and the day immediately prior to expiry day divided by the number of trading days per year which is taken as 252\(^{17}\). Note that the expiry day is not taken into account when calculating the trading days, because all prices recorded are daily closing prices, while all expiring contracts cease at 12:00 noon on the expiry day and the cash settlement price is calculated on the expiry morning.

### 2.4.2 Measurement of IV

Implied volatility is the volatility implied by an option price observed in the market based on an option pricing model. Thus this study is very dependent on option pricing model used. In this subsection, the original Black-Scholes (BS) model and its limitations are briefly discussed. Then a dividend-adjusted BS model is presented in details and used as the option pricing model in this thesis. Furthermore, it is well known that at any time implied volatility derived from the Black-Scholes model varies across different levels of moneyness, thus a decision of which implied volatility to use has to be made. At last, the measurement errors in implied volatility should also be considered before taking any further action.

\(^{17}\) As suggested by Hull (2003).
2.4.2.1 The option pricing models

2.4.2.1.1 Black-Scholes (1973) model

The assumptions underlying the original Black-Scholes European option pricing model are as follows:

1. The stock price follows a lognormal distribution with a constant volatility.
2. The short selling of securities with full use of proceeds is permitted.
3. There are no transactions costs or taxes. All securities are perfectly divisible.
4. There are no dividends during the life of the derivative.
5. There are no riskless arbitrage opportunities.
6. Securities trading is continuous.
7. The risk-free rate of interest is constant and the same for all maturities.

Based on the assumptions above, the Black-Scholes option pricing formulae for European call and put options can be written as

\[ C = S N(d_1) - X e^{-rT} N(d_2), \]  

(2.1)

and

\[ P = X e^{-rT} N(-d_2) - S N(-d_1), \]  

(2.2)

respectively, where

\[ d_1 = \frac{\ln(S/X) + rT + \sigma^2 T/2}{\sigma \sqrt{T}} \]

\[ d_2 = d_1 - \sigma \sqrt{T} \]
where

\[ C = \text{the call option price} \]
\[ P = \text{the put option price} \]
\[ S = \text{the current stock price} \]
\[ X = \text{the exercise price of the option} \]
\[ T = \text{the time remaining until expiration of the option} \]
\[ r = \text{the continuously compounded annualized riskfree rate of interest for the period of } T \]
\[ \sigma = \text{the annualized volatility of the stock's return} \]
\[ N(\cdot) = \text{the cumulative normal density function of } (\cdot) \]

One of the assumptions employed by Black and Scholes (1973) is that the stock pays no dividend (Assumption 4). However, dividends on some stocks may be substantial and can have a significant effect on the valuation of options whose stocks make such payments during the life of the options. Therefore, a dividend adjustment must be allowed for in the option pricing formulae.

### 2.4.2.1.2 Merton (1973) model

Given that the ASX index options are European dividend paying options, it is natural to use Merton (1973) model, which generalizes the Black and Scholes (1973) model by relaxing the assumption of no dividend. Merton (1973) allows for a constant continuous dividend yield on the stock and/or stock index.

Let \( \sigma_{t,i} \) denote the volatility implied by an index option price and \( q_t \) denote the annualized continuously compounded dividend yield during the remaining life of
An empirical study of implied volatility in ASX index option market

~ Chapter 2 ~  42

the option, then the Merton (1973) formulae for call and put options can be expressed as:

\[
\begin{align*}
C_t &= S_t e^{-rT} N(d_1) - X_t e^{-rT} N(d_2) \\
P_t &= X_t e^{-rT} N(-d_2) - S_t e^{-rT} N(-d_1)
\end{align*}
\]

respectively, where

\[
\begin{align*}
d_1 &= \left( \ln(S_t / X_t) + (r_t - q_t)T_t + \sigma^2_{i,t} T_t / 2 \right) / \sigma_{i,t} \sqrt{T_t} \\
d_2 &= d_1 - \sigma_{i,t} \sqrt{T_t}
\end{align*}
\]

Hence, an estimate of the dividend yield \( q_t \) is needed. For the purpose of this study, the SFE index futures contracts are utilised, which have exactly the same expiry cycle, expiry date and underlying asset with the XJO index options. The cost of carry model (see, e.g. Hull, 2003) gives

\[
F_t = S_t e^{(r_t-q_t)T_t}
\]

Combining (2.5) with (2.3) and (2.4) yields

\[
\begin{align*}
C_t &= F_t e^{-rT} N(d_1) - X_t e^{-rT} N(d_2) \\
P_t &= X_t e^{-rT} N(-d_2) - F_t e^{-rT} N(-d_1)
\end{align*}
\]

and

\[
\begin{align*}
d_1 &= \left( \ln(F_t / X_t) + \sigma^2_{i,t} T_t / 2 \right) / \sigma_{i,t} \sqrt{T_t} \\
d_2 &= d_1 - \sigma_{i,t} \sqrt{T_t}
\end{align*}
\]
By solving (2.6) and (2.7) numerically\(^{18}\), the implied volatility series for call and put options can be constructed.

### 2.4.2.2 ATM option volatility or weighted implied volatility?

Under the framework of the Black-Scholes option pricing model, a single option price is sufficient to estimate the implicit parameter — volatility of underlying assets’ return over the remaining life of the option. However, according to the fact that options with the same maturity but different strikes yield different implied volatilities, a decision has to be made as to which of these implied volatilities should be used, or which weighing scheme should be adopted, to produce a single volatility assessment.

It has been widely accepted in the literature that implied volatility computed from an ATM\(^{19}\) option is superior to volatilities obtained from OTM or ITM options. This is because ATM options are often most actively traded and hence volatility derived from these options should be least prone to measurement errors. ATM implied volatility is also theoretically the most sound. For example, Feinstein (1989) pointed out that for the stochastic volatility process described in Hull and White (1987), implied volatilities from ATM and near expiration options provide the closest approximation to the average volatility over the option’s life, provided that volatility risk premium is either zero or a constant.

---

\(^{18}\) An iteration procedure is used to calculate implied volatilities. For more details, please see Appendix 1. Explicit formulae can also be employed, see e.g. Li (2005).

\(^{19}\) ATM, OTM and ITM are short for at-the-money, out-of-the-money, and in-the-money, respectively.
This means that if volatility is stochastic, implied volatilities from ATM options are less likely to be biased compared to those from OTM or ITM options.

### Table 2.2: Methods for computing weighted implied volatility

<table>
<thead>
<tr>
<th>Model</th>
<th>Formula</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schmalensee and Trippi (1978)</td>
<td>$\hat{\sigma} = \frac{1}{N} \sum \sigma_i$, where $\sigma_i$ is the implied volatility from the $i$th option price $O_i$</td>
<td>Equal weights. Typically implemented on a restricted set of options (e.g., excluding deep out-of-the-money options).</td>
</tr>
<tr>
<td>Latane and Rendleman (1976)</td>
<td>$\hat{\sigma}^2 = \frac{\sum w_i^2 \sigma_i^2}{\left(\sum w_i^2\right)^2}$, $w_i = \frac{\partial O_i}{\partial \sigma} \sigma_i$</td>
<td>Weights don't sum to one, creating biased volatility estimates.</td>
</tr>
<tr>
<td>Modified Latane and Rendleman</td>
<td>$\hat{\sigma} = \frac{\sum w_i \sigma_i}{\sum w_i}$, $w_i = \frac{\partial O_i}{\partial \sigma} \sigma_i$</td>
<td>Heaviest weight on near-the-money options. In-and out-of-the-money options weighted symmetrically.</td>
</tr>
<tr>
<td>Whaley (1982)</td>
<td>$\hat{\sigma} = \arg \min \sum [O_i - \hat{O}_i(\sigma)]^2$</td>
<td>Even heavier weight on near-the-money options than the Modified Latane and Rendleman. Typically implemented on transactions data, which affects the relative weights.</td>
</tr>
<tr>
<td>Beckers (1981)</td>
<td>$\hat{\sigma} = \arg \min \sum \frac{w_i [O_i - \hat{O}_i(\sigma)]^2}{\sum w_i}$</td>
<td>Even heavier weight on near-the-money options than Whaley (1982).</td>
</tr>
<tr>
<td>Chiras and Manaster (1978)</td>
<td>$\hat{\sigma} = \frac{\sum w_i \sigma_i}{\sum w_i}$, $w_i = \frac{\partial O_i}{\partial \sigma} \sigma_i$</td>
<td>Elasticity-weighted, with heaviest weight on low-priced, deep out-of-the-money options.</td>
</tr>
<tr>
<td>At-the-money</td>
<td>$\hat{\sigma} = \sigma_{ATM}$</td>
<td>Increasingly standard. A readily replicable benchmark based on actively traded options.</td>
</tr>
</tbody>
</table>

On the other hand, many studies have focused on computing a composite implied volatility by placing different weights on ATM, OTM and ITM options. Bates (1996) provides a good summary for these weighing schemes, as shown in Table 2.2. Most assign the heaviest weight to ATM implied volatility. Other weighing
scheme, such as equally weighted model by Schmalensee and Trippi (1978), and elasticity weighted model by Chiras and Manaster (1978), are less popular due to not emphasizing ATM implied volatility.

Furthermore, Beckers (1981) shows that using only at-the-money options is preferable to various other weighing schemes, and hence questions the usefulness of weighting systems. Jorion (1995) also concentrates on at-the-money options, and suggests that using the arithmetic average of the implied volatilities from at-the-money call and put options alleviates some of the measurement problems. In addition, since the plot of implied volatility against strikes can take many shapes, it is not likely to remove all pricing errors consistently.

Therefore, this chapter will follow the conclusions from previous studies and use implied volatilities from ATM options instead of weighted implied volatility.

2.4.2.3 Measurement errors

Before assessing the empirical results, it is worth to be aware of that there are several sources of measurement errors which may afflict the estimation of implied volatility.

2.4.2.3.1 Non-synchronous prices

First, option prices, closing index levels and futures prices may be non-synchronous. This is either because the closing times for the three markets are different, or because the XJO index options are not traded frequently. The
closing option prices may correspond to a trade taking place before the market is closed. However, the index levels and futures contracts do not suffer from such illiquid problems. For examples, when some good news enters into the market between the trade of option and the time when closing index level and futures price are recorded, then the recorded index level or futures price will be higher than the index level or futures price simultaneously corresponding to the option price, indicating that implied call (put) volatility underestimates (overestimates) the true implied volatility. A similar situation can happen with bad news which may lead to deviations in the opposite direction. In reality, good news and bad news come randomly, and hence the two effects can offset each other and the computed implied volatility will not deviate consistently from the true volatility. However, this non-synchronous measurement does cause an errors-in-variable (EIV) problem, which leads to the correlation between the explanatory variable and the error term in the subsequent regressions.

2.4.2.3.2 Distribution assumption

The Black-Scholes (BS) option pricing model assumes that the index level follows a log-normal distribution with constant volatility during the life of the option. In the real-world market, this assumption could be violated, for instance, due to the jumps in the index level. Hence, the BS model can be misspecified and implied volatility computed from the BS model can be consequently misspecified. Shu and Zhang (2003) compare implied volatility computed from the BS model with the one implied from the Heston (1993) stochastic volatility model, and conclude that implied volatility computed from the BS model still has higher explanatory power than that computed from the Heston model. Thus, up to now,
the BS model may still be the best model for estimating volatility implied in the option prices.

2.4.2.3.3 Dividend adjustment

It should be noted that the dividend adjustments are required for applying the original BS option pricing model. To this end, the relationship between futures prices and spot prices, namely the cost of carry model, is utilised. However, in practice, the cost of carry model may not hold and thus it may contribute to the EIV problem.

2.4.2.3.4 Effect of long maturity cycle

The XJO options have a three month maturity cycle, which implies that the index volatility has to be assumed to be constant over the three month period. However, in practice, as evidenced in many empirical studies, volatility is not constant and follows its own stochastic process. Hence, the longer maturity period may exacerbate the EIV problem more than that for the S&P 100 index options in the U.S. market whose maturity cycle is only 1 month.

In sum, it is more likely that the EIV problem exists within the XJO options series and thus it must be accounted for when assessing the relation between implied volatility and realized volatility.

2.4.3 Measurement of RV and HV
2.4.3.1 Realized volatility

The realized volatility can be measured by the sample standard deviation of the daily index returns over the remaining life of an option\textsuperscript{20}.

Let \( n \) be the number of trading days before expiration, \( S_i \) be the index level on the day of the remaining life of the option, and \( R_i \) denote log-return on the \( i \)th day. Then

\[
R_i = \ln(S_i / S_{i-1})
\]

for \( i = 2, 3 \ldots n \). Thus, the annually realized volatility can be expressed as:

\[
\sigma_{r,t} = \sqrt{\frac{252}{n-2} \sum_{t=2}^{n} (R_{i,t} - \bar{R}_t)^2}, \tag{2.8}
\]

where \( \bar{R}_t \) denotes the mean of daily index log-returns during period \( t \).

2.4.3.2 Historical volatility

In the previous studies, historical volatility at time \( t \) is often defined as realized volatility at time \( t-1 \). However, in this study, the time to maturity ranges from 53 to 63 days. If the previous measurement is used, the information contained in the gap between two consecutive contracts will be ignored. It is believed that the more recent data contains more relevant information about the future. Thus, in this study, a different definition of historical volatility is presented.

\textsuperscript{20} This measure is only a proxy of true but unknown realized volatility. Several other measure methods of realized volatility are outlined in Appendix 2.
For a chosen option contract with $T$ days to maturity at time $t$, the corresponding historical volatility is calculated by using the daily index log-returns of the period going back $T$ days from time $t$. That is,

$$\sigma_{h,t-1} = \sqrt{\frac{252}{T-2} \sum_{i=2}^{T} (R_{i,t-1} - \bar{R}_{t-1})^2}, \quad (2.9)$$

where $\bar{R}_{t-1}$ denote the mean of daily index log-returns during the period $t-1$.

### 2.4.4 Regression tests

This section is the core part of this chapter. First, the OLS regressions used to examine the relation between implied and realized volatility are constructed. Then the effects of EIV problems are mathematically derived. A formal test for the presence of EIV problems, Hausman (1978) test, is also given. At last, the solution for the EIV problems, i.e. the IV estimation method, are described.

#### 2.4.4.1 OLS method

##### 2.4.4.1.1 OLS regression specifications

To explore the information content of implied volatility by OLS estimates, the following regressions are conducted for both call and put series:

$$rv_i = \alpha_0 + \alpha_i iv_i + \varepsilon_i, \quad (2.10)$$

$$rv_i = \alpha_0 + \alpha_i iv_i + \alpha_h hv_{t-1} + \varepsilon_i, \quad (2.11)$$

For reasons of comparison, the regression of realized volatility on historical volatility is also performed as below:
where \( rv_t, iv_t, \) and \( hv_{t-1} \), respectively, denote the natural logarithm of the realized volatility, implied volatility and historical volatility.

There are a few testable hypotheses of main interest (see e.g., Christensen and Prabhala, 1998; Szakmary et al., 2003). Firstly, if implied volatility contains some information about future realized volatility, then the coefficient of the implied volatility \( \alpha_i \) in both (2.10) and (2.11) should be nonzero. Secondly, if implied volatility is an unbiased forecast of future realized volatility, then the coefficient of implied volatility \( \alpha_i \) should be equal to one and the coefficient of intercept \( \alpha_0 \) should be equal to zero in both (2.10) and (2.11). Thirdly, if implied volatility includes more information than historical volatility, then implied volatility should have greater explanatory power than historical volatility, i.e., a higher \( \text{Adjusted }- R^2 \) would be expected from regression (2.10) than from (2.12). Finally, if implied volatility is an informationally efficient predictor of future realized volatility, i.e., implied volatility efficiently includes all information to predict future volatility, then the coefficient of historical volatility \( \alpha_h \) in (2.11) should be zero, and the error term \( \varepsilon_t \) should be white noise and thereby uncorrelated with any explanatory variable in the market’s information set.

In addition, to explore the relative information content of implied call and put volatility, the following regressions can be conducted:

\[
rv_t = \alpha_0 + \alpha_h hv_{t-1} + \epsilon_t,
\]  

(2.12)
\[ rV_t = \alpha_0 + \alpha_c iv_{t,c} + \alpha_p iv_{t,p} + \alpha_h hv_{t-1} + \epsilon_t \]  

(2.14)

where \( iv_{t,c} \) and \( iv_{t,p} \) denote the natural logarithm of implied call and put volatility, respectively.

Several hypotheses can be tested within regression (2.13) and (2.14) (see e.g. Hansen, 2001; Christensen and Hansen, 2002). Firstly, the unbiased hypothesis becomes that \( \alpha_c + \alpha_p = 1 \) and \( \alpha_0 = 0 \) in regression (2.13). Secondly, the issue of which of the two implied volatilities is more informative about the future realized volatility can be investigated by comparing \( \alpha_c \) with \( \alpha_p \) in regression (2.13). In fact, this regression can help to find out an optimal weighting of these two implied volatilities to forecast future volatility. Finally, the efficiency hypothesis is the same as before. The coefficient of historical volatility is expected to be zero in regression (2.14) if implied call and put volatility contain all the information to predict future volatility.

The empirical results for OLS regressions (2.10)-(2.14) will be presented in Section 2.5.2.

### 2.4.4.1.2 The effects of EIV problem

As discussed in Section 2.4.2.3, due to the shortcomings of Black-Scholes formula and the poor liquidity of S&P/ASX 200 index option market, a number of errors exist in the measurement of implied volatility. If these errors result in the implied volatility series being correlated with the regression error term, then
the EIV problem arises. This subsection illustrates how the EIV problem causes
the OLS estimates to appear both biased and inconsistent, based on standard least
squares asymptotics (see e.g. Greene, 2000, Chapter 9).

Let \( iv_t^* \) denotes the true implied volatility. The relation between realized volatility
and true implied volatility is given by

\[
rv_t = \alpha_0 + \alpha_i iv_t^* + \epsilon_t^*,
\]

(2.15)

where \( iv_t^* \) is not correlated with the error term \( \epsilon_t^* \). Because the true implied
volatility \( iv_t^* \) is not observed, the regression actually estimated is the one based
on imprecisely measured implied volatility \( iv_t \), as shown in (2.10).

The relation between true implied volatility \( iv_t^* \) and measured implied
volatility \( iv_t \) can be written as

\[
iv_t = iv_t^* + u_t,
\]

(2.16)

where \( u_t \) denotes the measurement error. Because the error is purely random, it is
reasonable to suppose that \( u_t \) has mean zero and variance \( \sigma_u^2 \) and is uncorrelated
with \( iv_t^* \) and the regression error \( \epsilon_t^* \).

Substituting (2.16) into (2.15) yields the actual estimated regression (2.10)

\[
rv_t = \alpha_0 + \alpha_i (iv_t - u_t) + \epsilon_t^* = \alpha_0 + \alpha_i iv_t + (\epsilon_t^* - \alpha_i u_t) = \alpha_0 + \alpha_i iv_t + \epsilon_t,
\]

where \( \epsilon_t = \epsilon_t^* - \alpha_i u_t \). Since \( iv_t \) equals \( iv_t^* + u_t \), the regressor in (2.10) is correlated
with the disturbance:

\[
\text{cov}[iv_t, \epsilon_t] = \text{cov}[iv_t^* + u_t, \epsilon_t^* - \alpha_i u_t] = -\alpha_i \sigma_u^2.
\]

(2.17)
This result violates one of the key assumptions of the classical linear regression model (CLRM), thus the OLS slope estimate for (2.10)

\[
\hat{\alpha}_i = \frac{(1 / n) \sum_{i=1}^{n} (iv_i - i\bar{v})(rv_i - r\bar{v})}{(1 / n) \sum_{i=1}^{n} (iv_i - i\bar{v})^2}
\]

(2.18)

is inconsistent. That is, the estimate does not converge to the true population value as the sample size increases infinitely. Since \(iv_i^*, \epsilon_i^*\) and \(u_i\) are mutually independent, the probability limit of \(\hat{\alpha}_i\) is given by\(^{21}\)

\[
p \lim \hat{\alpha}_i = \frac{\alpha_i}{1 + \sigma_u^2 / \sigma_{iv}^2},
\]

(2.19)

where \(\sigma_{iv}^2\) and \(\sigma_u^2\) denote the variances of true implied volatility and the measurement error, respectively. Hence, as long as \(\sigma_u^2\) is larger than zero, the OLS estimate \(\hat{\alpha}_i\) for the univariate regression (2.10) is inconsistent, and has a persistent bias toward zero. This effect of EIV problem is called attenuation. Furthermore, the greater the variance of measurement error in implied volatility, the worse the bias.

In the multiple regression case, with mismeasured implied volatility \(iv_i\) and historical volatility \(hv_{i-1}\) as the regressors, the matters are getting worse. The relation among realized volatility \(rv_i\), true implied volatility \(iv_i^*\) and historical volatility \(hv_{i-1}\) can be expressed as

\[
rv_i = \alpha_0 + \alpha_i iv_i^* + \alpha_{hv} hv_{i-1} + \epsilon_i^*,
\]

(2.20)

\(^{21}\) The derivation of the probability limit of \(\hat{\alpha}_i\) is presented in Appendix 3.
Substituting (2.16) into (2.20) yields the actual estimated regression (2.11):

\[
rv_t = \alpha_0 + \alpha_i iv_t + \alpha_h hv_{t-1} + (\epsilon_t^* - \alpha_i u_t) = \alpha_0 + \alpha_i iv_t + \alpha_h hv_{t-1} + \epsilon_t,
\]

where \( \epsilon_t = \epsilon_t^* - \alpha_i u_t \). The relation between the measured implied volatility \( iv_t \) and the regression disturbance \( \epsilon_t \) remains the same as that in (2.17), and hence the OLS estimates for regression (2.11) are inconsistent.

To simplify, let \( RV_t = rv_t - r\bar{v} \), \( IV_t = iv_t - i\bar{v} \) and \( HV_{t-1} = hv_{t-1} - h\bar{v} \). The OLS estimates in (2.11) can be written as

\[
\hat{\alpha}_i = \frac{\sum_{t=1}^{n} IV_t RV_t \sum_{t=1}^{n} HV_{t-1}^2 - \sum_{t=1}^{n} HV_{t-1} RV_t \sum_{t=1}^{n} IV_t HV_{t-1}}{\sum_{t=1}^{n} IV_t^2 \sum_{t=1}^{n} HV_{t-1}^2 - \left( \sum_{t=1}^{n} IV_t HV_{t-1} \right)^2},
\]

(2.21)

and

\[
\hat{\alpha}_h = \frac{\sum_{t=1}^{n} HV_{t-1} RV_t \sum_{t=1}^{n} IV_t^2 - \sum_{t=1}^{n} IV_t RV_t \sum_{t=1}^{n} IV_t HV_{t-1}}{\sum_{t=1}^{n} IV_t^2 \sum_{t=1}^{n} HV_{t-1}^2 - \left( \sum_{t=1}^{n} IV_t HV_{t-1} \right)^2}.
\]

(2.22)

As the sample size increases infinitely, the probability limits of the OLS estimate \( \hat{\alpha}_i \) and \( \hat{\alpha}_h \) are given by\(^{22}\)

\[
p\lim \hat{\alpha}_i = \alpha_i \frac{1 - \rho_i^2}{1 - \rho_i^2 + \sigma_i^2 / \sigma_{iv_i}^2},
\]

(2.23)

\[
p\lim \hat{\alpha}_h = \alpha_h + p\lim \hat{\alpha}_i \frac{\sigma_i^2}{\sigma_{iv_i}^2 \sigma_{hv_i}^2} \frac{\rho_i}{1 - \rho_i^2},
\]

(2.24)

\(^{22}\) The derivations of the probability limit of \( \hat{\alpha}_i \) and \( \hat{\alpha}_h \) are presented in Appendix 4.
respectively, where $\sigma_{iv}^2$, $\sigma_{hv}^2$, and $\sigma_u^2$ denote the variances of true implied volatility, historical volatility, and the measurement error, respectively; and $\rho$ denotes the correlation between true implied volatility and historical volatility.

Hence, in the context of multiple regression, the measurement error in implied volatility leads both $\hat{\alpha}_i$ and $\hat{\alpha}_h$ to appear biased and inconsistent. In particular, the attenuation effect, i.e., $\hat{\alpha}_i$ is biased downward towards zero, is more severe in multiple regression (2.11) than in univariate regression (2.10), since the probability limit of $\hat{\alpha}_i$ in (2.11) is less than that in (2.10). In addition, the slope coefficient for historical volatility is biased upward, since $\hat{\alpha}_i$ and $\rho$ are positive (as shall shown in empirical results). Even when historical volatility has no explanatory power, i.e., $\alpha_h = 0$, the probability limit of $\hat{\alpha}_h$ is still positive. Further, the upward bias will be greater when $\hat{\alpha}_i$ and $\rho$ are large. That is, the smaller the bias in the slope coefficient of implied volatility and the stronger the relation between implied and historical volatility, the more severe the bias in the slope coefficient of historical volatility.

### 2.4.4.2 Instrument variable (IV) method

#### 2.4.4.2.1 Hausman (1978) test

This subsection presents a formal test to verify the presence of the EIV problem in this study. As noted previously, if the EIV problem exists, then the OLS estimates can be not only biased but also inconsistent. That is, the estimates do not converge to the true population value as the sample size increases infinitely.
Thus, the OLS estimates of regression (2.10) and (2.11) may yield misleading results. In this case, the remedy is the alternative method, IV method. However, the IV estimates are less efficient than OLS estimates if the EIV problem does not exist. Thus the Hausman (1978) test is needed to test for the presence of EIV problem.

The basic idea of Hausman (1978) test is to construct a \( \chi^2 \) test statistic based on the difference between OLS estimator and IV estimator. But as suggested by Davidson and MacKinnon (1989, 1993), one never need to construct the difference between two estimators to compute the statistic. Hausman (1978) test can be illustrated by a simple version in which an auxiliary regression is utilised. To carry out the Hausman test, two OLS regressions are run. The first regression is to regress the suspect variable \( iv_t \) on all exogenous variables and instrument variables. In this case, the first regression is given as

\[
iv_t = \beta_0 + \beta_1 iv_{t-1} + \beta_2 hv_{t-1} + e_t .
\]  

(2.25)

The second regression is to re-estimate regressions (2.10) by including the residuals from the first regression as an additional regressor. That is,

\[
riv_t = \alpha_0 + \alpha_1 iv_t + \alpha_2 e_t + \epsilon_t
\]

(2.26)

If the OLS estimates are consistent, then the coefficient on the first stage residuals \( \alpha_e \) should not be significantly different from zero. The empirical results will be presented in Section 2.5.3.1.
2.4.4.2.2 IV regression specifications

This subsection explains why coefficient estimators from IV method are consistent, and presents how to perform the IV method by a two-stage-least-squares procedure (2SLS). As suggested by Greene (2000), the idea behind IV method is to find out a set of variables, termed instruments, which are highly correlated with the suspect explanatory variables, but uncorrelated with the error term. These instruments are used to capture part of the information in implied volatility, which is uncorrelated with the error term. Hence, the coefficient estimates from IV method are consistent\(^{23}\).

To perform the IV method successfully, the number of instruments must satisfy the order condition for identification, i.e., there must be at least as many instruments as there are coefficients in the regressions. The natural candidates of instruments are lagged implied volatility \(iv_{t-1}\) and historical volatility \(hv_{t-1}\), since \(iv_{t-1}\) and \(hv_{t-1}\) are highly correlated with implied volatility at time \(t\), as shown in Table 2.5, but are quite plausibly uncorrelated with \(\varepsilon_t\), the error term associated with implied volatility sampled three months later.

The IV method can be achieved by the two stage least squares procedure (2SLS). In the first stage, implied volatility \(iv_t\) is regressed on the instrumental variables \(iv_{t-1}\) and \(hv_{t-1}\) by OLS method:

\(^{23}\) The proof of consistent estimate from 2SLS is presented in Appendix 5.
\[ iv_t = \beta_0 + \beta_1 iv_{t-1} + \beta_2 hv_{t-1} + e_t. \]  \hspace{1cm} (2.27)

In the second stage, the regressions (2.10) and (2.11) are re-examined by replacing implied volatility \( iv_t \) with the fitted implied volatility \( \hat{iv}_t \) from the first stage regression (2.27). That is,

\[ rv_t = \alpha_0 + \alpha_1 \hat{iv}_t + e_t, \]  \hspace{1cm} (2.28)

and

\[ rv_t = \alpha_0 + \alpha_1 \hat{iv}_t + \alpha_2 hv_{t-1} + e_t. \]  \hspace{1cm} (2.29)

The empirical results of IV estimates will be presented in Section 2.5.3.2.
2.5 Empirical results

This section presents the empirical results on the relation between implied and realized volatility for the XJO index options. Firstly, the descriptive statistics for various volatility series, which are constructed in the Section 2.4, are provided. Then regression results from two estimation methods are reported. One is the conventional analysis, namely Ordinary Least Squares (OLS) method. To solve the EIV problems, the other estimation method is implemented, which is called Instrumental Variables (IV) method. The following analyses are based on log-volatility series.

2.5.1 Descriptive statistics

Table 2.3 presents descriptive statistics for the level and natural logarithm series for implied-, realized- and historical- volatility, while Figure 2.1 gives an intuitive impression for the level series of these volatilities.

Table 2.3 clearly indicates that the average (log) implied volatility is larger than the average (log) realized and average (log) historical volatility for both call and put options series. This indicates that the Black-Scholes model tends to overprice both at-the-money call and put options on average. A similar finding has been demonstrated for options on the S&P 100 index (Christensen and Prabhala, 1998), for options on the S&P 500 index (Lin et al., 1998) and for options on the Danish KFX index (Hansen, 2001).
Table 2.3: Descriptive statistics

This table reports descriptive statistics for quarterly level series and natural logarithm series of implied call and put volatility, realized volatility, and historical volatility for the S&P/ASX 200 stock index. Here, implied volatility is computed each quarter using the Black-Scholes-Merton model; realized volatility is the annualized ex-post daily return volatility (sample standard deviation) of the index over the remaining life of the option. Historical volatility is calculated as the annualized sample standard deviation of the daily index return over a period before the sampling time, but of the same time length as that for calculating realized volatility. Statistics are based on 20 non-overlapping quarterly observations for each series, covering a five-year period from April 2, 2001 to March 16, 2006.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Implied call volatility</th>
<th>Implied put volatility</th>
<th>Realized volatility</th>
<th>Historical volatility</th>
<th>Log implied call volatility</th>
<th>Log implied put volatility</th>
<th>Log realized volatility</th>
<th>Log historical volatility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.1166</td>
<td>0.1330</td>
<td>0.0986</td>
<td>0.0992</td>
<td>-2.1763</td>
<td>-2.0415</td>
<td>-2.3614</td>
<td>-2.3543</td>
</tr>
<tr>
<td>Std.Dev</td>
<td>0.0288</td>
<td>0.0304</td>
<td>0.0313</td>
<td>0.0313</td>
<td>0.2371</td>
<td>0.2217</td>
<td>0.3043</td>
<td>0.2994</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.7968</td>
<td>0.7358</td>
<td>0.7607</td>
<td>0.8858</td>
<td>2.5664</td>
<td>2.7260</td>
<td>2.0570</td>
<td>2.2424</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.9936</td>
<td>3.1325</td>
<td>2.6507</td>
<td>3.1169</td>
<td>0.4726</td>
<td>0.2096</td>
<td>1.0916</td>
<td>0.9386</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>2.1164</td>
<td>1.8194</td>
<td>2.0304</td>
<td>2.6269</td>
<td>0.2096</td>
<td>1.0916</td>
<td>0.9386</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2.1: Level series of various volatilities

This figure plots the level series of various volatilities defined in Section 2.4. Each series contains 20 non-overlapping observations, covering a five-year period from April 2001 to March 2006. *ive, ivp, rel, and his* denote the level series of implied call volatility, implied put volatility, realized volatility and historical volatility, respectively.
The average historical (log) volatility is sightly higher than the average (log) realized volatility. This may be due to the excess volatility on the expiration days, since the expiration day is excluded in the measurement of realized volatility while it is included in the measurement of historical volatility.

Comparing the call series and put series, it is found that the average (log) implied volatility for put option series is sightly higher than that for call option series. This may be due to the fact, as suggested by Harvey and Whaley (1991, 1992), that buying put options is a convenient and relative cheap method for hedging. Therefore, buying pressure on index put options is larger than that on index call options. Consequently, implied put volatility is higher than implied call volatility on average.

Turning to the standard deviations, it is found that realized volatility is always more volatile than implied call and put volatility in both level and log series,
which accords with the notion that implied volatility is a smoothed expectation of future realized volatility.

Now move to the distributions of the level series in contrast to the log series for all three volatility measures. Table 2.3 reveals that the skewness of log-volatility series is closer to zero than that of the corresponding level volatility series, while the kurtosis of the log-volatility series departs more from three than that of level volatility series. Overall, according to the Jarque-Bera test, it appears that the log-volatility series conform better to the normal distribution than the level volatility series. For this reason, the following analyses are based on the log-volatility series. Throughout the paper, $rv_t$, $iv_t$ ($iv_{t,c}$, $iv_{t,p}$) and $hv_{t-1}$ denote the natural logarithm of the realized volatility, implied (call and put) volatility, and historical volatility, respectively.

### 2.5.2 Conventional method – OLS estimates

Table 2.4 presents the results of OLS estimates of regressions (2.10)-(2.14). Before looking at the significance of individual coefficient estimates, F-test is used to check the overall significance of each regression. According to the F-statistics, it is found that regression (2.10) is significant at the 5% level for put series, but not significant at this level for call series. This result appears to indicate that implied put volatility contains some information about future realized volatility, while implied call volatility has no significant relation with future realized volatility and thus can not be used as a market forecast of future index volatility.
Table 2.4: Information content of implied volatility: OLS estimates

This table presents the OLS estimates of regressions (2.10)-(2.14) for both call and put series:

\[ rv_t = \alpha_0 + \alpha_i iv_t + \varepsilon_t, \]
\[ rv_t = \alpha_0 + \alpha_i iv_t + \alpha_h hv_{t-1} + \varepsilon_t, \]
\[ rv_t = \alpha_0 + \alpha_i iv_{t,c} + \varepsilon_t, \]
\[ rv_t = \alpha_0 + \alpha_i iv_{t,c} + \alpha_p iv_{t,p} + \varepsilon_t, \]
\[ rv_t = \alpha_0 + \alpha_i iv_{t,c} + \alpha_p iv_{t,p} + \alpha_h hv_{t-1} + \varepsilon_t, \]

where \( rv_t \) denotes the natural logarithm of the daily return realized volatility of the index over the remaining life of option; \( iv_t (= iv_{t,c} \) for the call series and \( iv_{t,p} \) for the put series) denotes the natural logarithm of Black-Scholes-Merton implied volatility for close to be at-the-money options on the underlying index, measured at time \( t \); \( hv_{t-1} \) denotes the natural logarithm of historical index return volatility during the period before time \( t \) and with the same time length as that of the future realized volatility. The data consist of 20 non-overlapping quarterly observations for each series, covering a five-year period from April 2001 to March 2006. Note that DW denotes Durbin-Watson statistic and the numbers in parentheses are \( t \)-statistics. * denotes significant at the 5% level.

<table>
<thead>
<tr>
<th>Dependent variable: Log realized volatility ( rv_t )</th>
<th>Independent variables: ( iv_{t,c} ), ( iv_{t,p} ), ( hv_{t-1} )</th>
<th>Adj.R(^2)</th>
<th>DW</th>
<th>F-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.39*</td>
<td>0.45</td>
<td>7%</td>
<td>1.71</td>
</tr>
<tr>
<td>(-2.24)</td>
<td>(1.58)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.01</td>
<td>0.66*</td>
<td>19%</td>
<td>2.07</td>
<td>5.45*</td>
</tr>
<tr>
<td>(-1.74)</td>
<td>(2.33)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.23*</td>
<td>0.48*</td>
<td>18%</td>
<td>2.33</td>
<td>5.14*</td>
</tr>
<tr>
<td>(-2.46)</td>
<td>(2.27)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.20</td>
<td>0.042</td>
<td>13%</td>
<td>2.32</td>
<td>2.43</td>
</tr>
<tr>
<td>(-1.95)</td>
<td>(0.11)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.95</td>
<td>0.40</td>
<td>17%</td>
<td>2.31</td>
<td>2.96</td>
</tr>
<tr>
<td>(-1.59)</td>
<td>(0.91)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.06</td>
<td>-0.10</td>
<td>14%</td>
<td>2.06</td>
<td>2.61</td>
</tr>
<tr>
<td>(-1.68)</td>
<td>(-0.24)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.03</td>
<td>-0.21</td>
<td>13%</td>
<td>2.34</td>
<td>1.95</td>
</tr>
<tr>
<td>(-1.62)</td>
<td>(-0.45)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If this result were true, the joint hypothesis of option market efficiency and applicability of BS model would be rejected. However, as noted previously, this result might be distorted because of the EIV problem. For reasons of comparison,
regression (2.12) is conducted and significant at the 5% level, indicating that historical volatility can be used to forecast future volatility. Regression (2.11), (2.13) and (2.14) are not significant at the 5% level. This may be due to the EIV problem or the high correlation between the independent variables in the regressions, namely the multicollinearity problem. The correlation matrix in Table 2.5 provides some evidence on the presence of the multicollinearity problem. The correlation coefficient is 0.70 between implied call and historical volatility, and 0.76 between implied put and historical volatility.

Table 2.5: Pairwise correlation Matrix

This table presents a pairwise correlation matrix for the dependent and independent variables in the regressions in this study, where \( r_{\text{vt}} \) denotes the natural logarithm of the daily return realized volatility of the index over the remaining life of option; \( iv_{t,c} \) and \( iv_{t,p} \) denote the natural logarithm of Black-Scholes-Merton implied volatility for at-the-money call and put options on the underlying index, measured at time \( t \); \( iv_{t-1,c} \) and \( iv_{t-1,p} \) denote the first lagged value of implied call and put volatility, respectively; and \( h\text{vt}_{t-1} \) denotes the natural logarithm of historical index return volatility during the period before time \( t \) and with the same time length as that of the future realized volatility. The data consist of 20 non-overlapping quarterly observations for each series except for \( iv_{t-1,c} \) and \( iv_{t-1,p} \) (19 observations for each), covering a five-year period from April 2001 to March 2006.

<table>
<thead>
<tr>
<th></th>
<th>( r_{\text{vt}} )</th>
<th>( iv_{t,c} )</th>
<th>( iv_{t,p} )</th>
<th>( h\text{vt}_{t-1} )</th>
<th>( iv_{t-1,c} )</th>
<th>( iv_{t-1,p} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_{\text{vt}} )</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( iv_{t,c} )</td>
<td>0.35</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( iv_{t,p} )</td>
<td>0.48</td>
<td>0.79</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( h\text{vt}_{t-1} )</td>
<td>0.47</td>
<td>0.70</td>
<td>0.76</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( iv_{t-1,c} )</td>
<td>0.27</td>
<td>0.31</td>
<td>0.27</td>
<td>0.33</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>( iv_{t-1,p} )</td>
<td>0.30</td>
<td>0.37</td>
<td>0.27</td>
<td>0.44</td>
<td>0.79</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Now turn to the significance of individual coefficient estimates. In regression (2.10) and (2.12), both coefficients of the historical volatility and implied put volatility are significant at 5% level. The coefficient of historical volatility is around 0.48 and that of implied put volatility is 0.66, indicating that both of them are biased predictors of the future realized volatility. The Adjusted-$R^2$ of regression (2.10) suggests that implied put volatility explains about 19% of the variation of future volatility, while that for historical volatility in regression (2.12) is about 18%. It appears that implied put volatility dose not significantly overwhelm historical volatility and only has slightly higher explanatory power than historical volatility in forecasting future volatility.

In addition, none of the coefficients of implied call volatility in (2.10) and (2.11) is significant, and the $t$-value of the coefficient of historical volatility in regression (2.11) is even higher than that for implied call volatility. There are two possible explanations of the fact that historical volatility contains more information about subsequent realized volatility than implied call volatility. One is that historical volatility is indeed more informative than implied call volatility for ASX index options. Since ASX index option market is an illiquid market, volatility implied in the option prices is stale and might not be an unbiased and efficient forecast of future realized volatility. If this explanation were true, then it would reject the joint hypothesis that BS formula is valid and ASX index option market is efficient. The other explanation is that the OLS results are seriously affected by the measurement errors in the implied volatility, namely, EIV problem. As discussed in Section 2.4.4.1.2, the presence of EIV problem can lead to a few consequences: the estimated coefficient of implied volatility $\hat{\alpha}$ is
downward biased, even to zero, and the estimated coefficient of historical volatility $\hat{\alpha}_h$ is upward biased. This suggests that the OLS estimates of regressions (2.10) and (2.11) appear to be biased and inconsistent, and hence leads to the incorrect conclusions that XJO implied volatility is biased and inefficient.

Up to this point, the results reveal that XJO implied volatility is neither unbiased nor efficient. Implied put volatility has slightly more predictive power than historical volatility, whether assessed by the Adjusted-$R^2$ of each regression or by the magnitude of the regression slope coefficients. In this respect, XJO index options do not appear to be dramatically different from OEX options (Christensen and Hansen, 2002), and KFX options (Hansen, 2001), except that implied call volatility does not have significant relation with future index volatility.

Given these OLS regression results, it is needed to investigate whether EIV problem is significant for the implied volatility series, and if so, to take the EIV problem into account in the analysis.

### 2.5.3 Alternative method – IV estimates

#### 2.5.3.1 Presence of EIV problem

After performing Hausman (1978) test by a simple version proposed by Davidson and MacKinnon (1989, 1993), it is found that, for call option series, the coefficient estimate of $\alpha_e$ in the second regression (2.26) is -0.86 and
significant at 15%. For put option series, the coefficient estimate of $\alpha_e$ is $-0.40$ but not statistically significant. Hence, the presence of EIV problem is confirmed for call option series but not for put option series. Thus the IV method will be applied to call option series in the following analysis.

2.5.3.2 \textit{IV estimates}

Table 2.6 reports the results of 2SLS estimates by using lagged implied call volatility and historical volatility as the instruments for implied call volatility.

Panel A presents the OLS estimates of the first stage regression (2.27). It appears that constant term, lagged implied call volatility $iv_{t-1}$ and historical volatility $hv_{t-1}$ altogether explains about 44 percent of the total variation of implied call volatility $iv_t$. The high level of adjusted $R^2$ makes the second stage IV regression more significant. The coefficient of historical volatility $hv_{t-1}$ is highly significant at the 1% level, whilst the constant term and coefficient of lagged implied volatility $iv_{t-1}$ is not significant in the regression (2.27). Furthermore, these results reveal that implied volatility at time $t$ can be predicted at large by historical volatility, which are known to the market in advance of time $t$. Thus, specification (2.27) provides a way to forecast future implied volatility, and hence future option prices, using past information available in the market.
Table 2.6: Information content of implied call volatility: 2SLS estimates

This table presents the results of 2SLS estimates by using historical volatility $h_{v,t-1}$ and lagged implied call volatility $iv_{t-1}$ as the instruments for implied call volatility $iv_t$.

Panel A reports the OLS estimates of the first stage regression (2.27) for call option series,

\[ iv_t = \beta_0 + \beta_1 iv_{t-1} + \beta_2 h_{v,t-1} + \epsilon_t \]

where $iv_t$ denotes the natural logarithm of Black-Scholes-Merton implied volatility for at-the-money options on the underlying index, measured at time $t$, especially, $iv_{t,c}$ denotes implied call volatility; $iv_{t-1,c}$ denotes the first lagged value of implied call volatility; and $h_{v,t-1}$ denotes the natural logarithm of historical index return volatility during the period before time $t$ and with the same time length as that of the future realized volatility.

Panel B presents the IV estimates of the second stage regression (2.28) and (2.29),

\[ rv_t = \alpha_0 + \alpha_1 iv_t + \epsilon_t \]

\[ rv_t = \alpha_0 + \alpha_1 iv_t + \alpha_2 h_{v,t-1} + \epsilon_t \]

where $\hat{iv}_t$ is the fitted values from the first stage regression (2.27). Note that DW denotes Durbin-Watson statistic and the numbers in parentheses are $t$-statistics. ** and * denote significant at 1% and 5%, respectively.

<table>
<thead>
<tr>
<th>Panel A: First stage regression estimates</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent</td>
<td>Intercept</td>
<td>$iv_{t,c}$</td>
<td>$h_{v,t-1}$</td>
<td>Adj.$R^2$</td>
<td>DW</td>
<td>F-Statistic</td>
</tr>
<tr>
<td>$iv_{t,c}$</td>
<td>-0.73</td>
<td>0.09</td>
<td>0.53**</td>
<td>44%</td>
<td>2.29</td>
<td>8.09**</td>
</tr>
<tr>
<td></td>
<td>(-1.66)</td>
<td>(0.50)</td>
<td>(3.61)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Second stage regression estimates</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent</td>
<td>Intercept</td>
<td>fitted-$iv_{t,c}$</td>
<td>$h_{v,t-1}$</td>
<td>Adj.$R^2$</td>
<td>DW</td>
<td>F-Statistic</td>
</tr>
<tr>
<td>$rv_t$</td>
<td>-0.42</td>
<td>0.89*</td>
<td></td>
<td>19%</td>
<td>2.00</td>
<td>5.33*</td>
</tr>
<tr>
<td></td>
<td>(-0.49)</td>
<td>(2.31)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$rv_t$</td>
<td>0.35</td>
<td>1.79</td>
<td>-0.51</td>
<td>15%</td>
<td>1.93</td>
<td>2.56</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.56)</td>
<td>(-0.29)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel B in Table 2.6 presents the second stage IV estimates of the regression (2.28) and (2.29). In the case of univariate regression (2.28), which removes historical volatility and only includes fitted implied call volatility as explanatory variable, it is found that the coefficients, t-values and adjusted $R^2$ all increase significantly, comparing to the OLS estimates results in the first line of Table 2.4.
This suggests that the measurement of implied call volatility is indeed affected by EIV problems. In particular, the coefficient of implied call volatility, $\alpha_i$, in IV estimates increases dramatically from 0.45 to 0.89, and is not significantly different from 1 (with $t$-statistics of -0.28), indicating that implied call volatility is an unbiased forecast of future realized volatility. The $t$-value increases from 1.58 to 2.31, indicating that implied call volatility now is significant at 5% to explain future realized volatility. The adjusted $R^2$ in the IV estimates is also improved, increasing dramatically from 7% to 19%.

In addition, comparing to the OLS estimate results for historical volatility in Table 2.4, it is found that, after accounting for the EIV problem, implied call volatility $iv_t$ does appear to contain more information about future realized volatility $rv_t$ than historical volatility $hv_{t-1}$, whether judged by the adjusted $R^2$ of each regression or by the magnitude of the regression slope coefficients. In sum, both implied call and put volatility are superior to historical volatility in forecasting future index return volatility, and implied call volatility appears to be less biased than implied put volatility and historical volatility.

In the case of multivariate regression (2.29), which includes both fitted implied call volatility and historical volatility as regressors, it appears that both regressors are not statistically significant. This is probably due to the weak performance of lagged implied call volatility $iv_{t-1}$ as instrument in the first stage regression (2.27), which leads to that fitted implied call volatility is highly correlated with historical volatility, and in turn leads to the multicollinearity problem in the
second stage regression (2.29). For this reason, the issue of whether implied volatility is an informationally efficient forecast of future index volatility could not be decided, even after solving for the EIV problem by using IV method.
2.6 Conclusion

This chapter investigates the relation between implied volatility and subsequently realized volatility in the Australian S&P/ASX 200 index options (XJO) market during a period of five years from April 2001 to March 2006. This study is based on non-overlapping quarterly data, which is constructed by a sampling procedure proposed by Christensen and Prabhala (1998). Due to the quarterly maturity cycle of the XJO options, only 20 observations can be obtained each for the call option series and put option series.

Beginning with the conventional analysis (OLS method), it is found that implied call volatility has no relation with future volatility, whilst implied put volatility and historical volatility can be used to forecast future volatility. Furthermore, implied put volatility is less biased and has slightly higher predictive power than historical volatility. However, these results can be misleading because of the presence of measurement errors in implied volatility, namely the errors-in-variable (EIV) problem.

After discussing the possible factors which may be responsible for the EIV problem, the Hausman (1978) test is performed and formally confirms the existence of EIV problem in the measurement of implied call volatility. To account for this problem, an alternative method (IV method) is utilised. It is found that implied call volatility is indeed an unbiased estimator of future volatility and has slightly higher predictive power than historical volatility.
In sum, both implied call and put volatility are better than historical volatility in forecasting future realized volatility, whether assessed by the $Adjusted-R^2$ of each regression or by the magnitude of the regression slope coefficients. Moreover, implied call volatility is nearly an unbiased forecast of future volatility. However, the issue of whether implied volatility is informationally efficient could not be decided, either because of the multicollinearity problem between volatility series, or because of the weak performance of lagged implied call volatility as the instrument for implied call volatility.
Chapter 3  The Volatility Structure

Implied by S&P/ASX 200 Index Options

3.1  Introduction

3.1.1  Background and rationale
The Black-Scholes (1973) option pricing model is widely used to value options written on a large variety of underlying assets. Despite this widespread acceptance among practitioners and academics, the discrepancies between market and Black-Scholes prices are obvious and systematic. If the market were to price options according to the Black-Scholes model, all options on one stock would have the same implied volatility. However, it is well known that, at any moment of time implied volatility obtained by the Black-Scholes model varies across time to maturity as well as strike prices. The pattern of implied volatility for different time to maturity is known as the term structure of implied volatility, and the pattern across strike prices is known as the volatility smile or the volatility sneer. A term of volatility structure is used generally to refer to the pattern across both strike prices and time to maturity.

Many early studies have documented a U-shape smile pattern for implied volatility in many options markets prior to the 1987 stock market crash. For example, Shastri and Wethyavivorn (1987) find that implied volatilities from foreign currency options traded on the Philadelphia Stock Exchange (PHLX) in
1983 and 1984 are a U-shaped function of the exchange rate divided by the strike price. Sheikh (1991) has argued that a U-shaped pattern occurred for the S&P100 options during various subperiods between 1983 and 1985. Fung and Hsieh (1991) discuss informally some empirical smile effect for foreign currency options traded on the Chicago Mercantile Exchange (CME). Heynen (1993) has shown that U-shaped functions can describe the pattern of implied volatility from nine months’ transaction prices of European Options Exchange (EOE) stock-index options, which are European style options on an index of 25 active stocks on the Amsterdam Stock Exchange. Taylor and Xu (1994) fit U-shaped functions to implied volatilities from foreign currency options traded on the PHLX over a longer period from November 1984 to January 1992. They also find that the magnitude of the smile effect is a decreasing function of time to maturity.

In contrast to the above theoretical and empirical results showing a symmetric pattern for implied volatility against strike price, Dumas, Fleming and Whaley (1998) illustrate that the volatility structure for S&P 500 index options has changed from the symmetric smile pattern to more of a sneer since the stock market crash of 1987. That is, implied call (put) volatilities decrease monotonically as the call (put) option goes deeper out-of-the-money (in-the-money).

Brown (1999) extends Dumas, Fleming and Whaley’s (1998) results to the SPI 200 futures options on the Sydney Futures Exchange (SFE), covering the period from June 1993 to June 1994. In-the-money call (put) options are generally
trading at higher (lower) implied volatilities than out-of-the-money call (put) options.

### 3.1.2 Statement of the research problem

Volatility structure, the pattern of implied volatility across both strike prices and time to maturity, has attracted a great deal of interest of both practitioners and academics. However, most existing studies have focused on stock index options and foreign currency options in the U.S. market. To the authors’ best knowledge, no such investigation has been carried out for the S&P/ASX 200 (XJO) index options on the Australian Stock Exchange (ASX). Therefore, Chapter 3 of this thesis aims to contribute to the literature by investigating the volatility structure of the XJO options and providing possible explanations consistent with its shape.

### 3.1.3 Research methodology

The data used in this chapter are daily closing prices of the XJO options from April 2001 to June 2005. Since the XJO options are European style and dividend-paying options, the Merton (1973) model is employed to derive implied volatilities from the observed market prices. Then two dimensional graphs are plotted to examine the effect of volatility smile or sneer, i.e. how implied volatilities vary across different strike prices for options with same time to maturity on a particular day in the data set. Furthermore, to examine how volatility smile or sneer is affected by time to maturity, three dimensional graphs of implied volatility against moneyness and maturity of the option are plotted.

### 3.1.4 Structure
The rest of Chapter 3 is organized as follows. Section 3.2 provides a literature review of the previous studies on the volatility structure implied by options written on different underlying assets. Section 3.3 presents the data and methodology used in Chapter 3. Section 3.4 provides the empirical results and analysis. Section 3.5 presents a summary and the conclusions for Chapter 3.
3.2 Literature review

This section presents a review of the literature investigating the volatility structure implied by the option prices. According to the underlying assets of the options, the literature can be classified into three categories: evidence on foreign currency options, evidence on stocks and stock index options, and evidence on futures options.

3.2.1 Evidence on foreign currency options

3.2.1.1 Shastri and Wethyavivorn (1987)

Shastri and Wethyavivorn (1987) examine the volatility pattern implied by the options written on Japanese Yen, Swiss Franc, and West German Mark exchange rates on the floor of the Philadelphia Stock Exchange (PHLX), covering the period from March 1, 1983 to August 31, 1984. 28,527 observations for the three currencies are classified into one of 36 categories by time to expiration, and by spot exchange rate to exercise price ratio. Implied volatility is then calculated for each option in the dataset by using pure diffusion model, i.e. the Black-Scholes-Merton model. They find that mean implied volatilities are U-shaped functions of the spot exchange rate to exercise price ratio (S/X), except for options with time to maturity greater than 220 days. They also find that there is no unique implied volatility pattern with respect to option maturity.
3.2.1.2  *Taylor and Xu (1994)*

Taylor and Xu (1994) provide both theoretical and empirical evidence of volatility smile for the foreign currency options. Firstly, they show that the existence of stochastic volatility is a sufficient reason for smiles to exist. In particular, they show that an approximation to the theoretical implied volatility is a quadratic function of $\ln(F/X)$ where $F$ is the forward price and $X$ is the strike price, and that this approximate function has a minimum at $X=F$. This theoretical result requires that asset price and volatility differentials are uncorrelated and that volatility risk is not priced. The curvature of the quadratic function depends on the time to maturity of the option and several volatility parameters including the present level, any long-run median level and the variance of future average volatility. The magnitude of the smile effect is a decreasing function of time to maturity.

Turning to empirical tests, using foreign currency option data from the Philadelphia Stock Exchange (PHLX) over the period from November 1984 to January 1992, Taylor and Xu (1994) regress a function of theoretical and observed implied volatilities on moneyness, and find little evidence of asymmetry in these implied volatilities. Hence, the empirical evidence for foreign currency options supports the theoretical predictions, although the empirical smile is about twice the size predicted by the theory. These results are broadly consistent with those of Shastri and Wethyavivorn (1987) for two earlier years although they simply examine the mean implied volatilities across different maturities and different levels of moneyness.
3.2.2 Evidence on stocks and stock index options

3.2.2.1 Rubinstein (1985)

Of the early studies documenting volatility smiles, the most systematic and complete is that of Rubinstein (1985). Using all reported trades and bid-ask quotes on the 30 most liquid CBOE option classes from August 1976 to August 1978, Rubinstein (1985) conduct a non-parametric test for the null hypothesis that implied volatilities computed from the Black-Scholes formulae exhibit no systematic differences across strike prices or across time to maturity for otherwise identical options.

The most robust result in his study is that for out-of-the-money call options implied volatility is systematically higher for options with shorter time to maturity. His other results were statistically significant but changed across subperiods. He divides the sample into two subperiods: Period I from August 23, 1976 to October 21, 1977, and Period II from October 24, 1977 to August 31, 1978. For at-the-money calls, Rubinstein (1985) finds that in Period I, implied volatilities for options with short time to maturity are higher than for those with longer time to maturity, while the opposite result is true in Period II. Furthermore, in Period I, implied volatilities are higher for options with lower striking prices, but again, the result is reversed in Period II. Thus, Rubinstein (1985) concludes that systematic deviations from the Black-Scholes model appear to exist, but the pattern of deviations varies over time.
3.2.2.2  Sheik (1991)

Following the non-parametric approach of Rubinstein (1985) and using 14 months transaction data selected from the period July 1983 to December 1985, Sheik (1991) examines the implied volatility pattern for the S&P 100 call index options, and finds that the characteristics of the observed biases are apparently different from the biases reported upon option prices on an individual stock. He suggests that the observed biases correspond to biases that arise if option market prices incorporate a stochastically changing volatility of the underlying index. Furthermore, in two of the three subperiods in his sample, the implied volatilities with different strike prices but same time to maturity exhibit a U-shaped pattern, which is often characterized as a “smile pattern”.

3.2.2.3  Heynen (1993)

Heynen (1993) examines the implied volatility patterns for European Options Exchange (EOE) stock-index call options, which are European style options on an index of 25 active stocks on the Amsterdam Stock Exchange. Using Rubinstein’s (1985) non-parametric approach and transactions data from January 23 to October 31, 1989, it is shown that implied volatility patterns are significantly U-shaped. Heynen (1993) reviews the predictions of various stochastic volatility models, finds the observed smile pattern to be inconsistent with them, and suggests an alternative explanation for the volatility smile, based on market imperfections.
3.2.2.4  **Dumas, Fleming and Whaley (1998)**

In contrast to the above theoretical and empirical results showing a symmetric pattern for implied volatility across strike prices, Dumas, Fleming and Whaley (1998) document that the patterns of the Black-Scholes implied volatilities for the S&P 500 index options has changed since the stock market crash of 1987. The smile pattern prior to the 1987 crash has changed to be more of a sneer, with call (put) option implied volatilities decreasing monotonically as the call (put) goes deeper out of the money (in the money).

Dumas, Fleming and Whaley (1998) explain the emergence of the volatility sneer as a consequence of an increase in investors’ probability assessment of downward moves in the index level. This explanation accords with one of those explanations made by Bates (2000) for the S&P 500 futures options implied volatilities over 1988-1993. Since out-of-the-money put options provide explicit portfolio insurance against substantial downward movements in the market, the heavy demand for out-of-the-money put options has driven up prices and in turn implied volatilities, when the market places a relatively greater probability on a downward price movement than an upward movement, resulting in a negatively skewed implied terminal asset distribution.

3.2.3  **Evidence on futures options**

3.2.3.1  **Fung and Hsieh (1991)**

Fung and Hsieh (1991) conduct an empirical analysis of implied volatilities across three futures options markets—stocks, bonds and currencies, over the period from March 1, 1983 through July 31, 1989. For stocks, they use S&P 500
futures contracts on the Chicago mercantile Exchange (CME), T-bond futures contracts on the Chicago Board of Trade (CBT) for bonds, and Deutschemark futures contracts on the International Monetary Market (IMM) for foreign currencies. In all three markets, smile or sneer patterns of implied volatility against strike prices are observed. For S&P 500 futures options, they find that call (put) option implied volatilities decrease monotonically as the call (put) goes deeper out of the money (in the money). For T-bond futures options, similar volatility pattern of monotonically decreasing appears for in-the-money call and out-of-the-money put options, but implied volatilities from in-the-money put and out-of-the-money call options display no unique patterns across levels of moneyness. A smile pattern is more pronounced for the Deutschemark futures options.

3.2.3.2 Brown (1999)

Brown (1999) examines the volatility structure implied by options on the SPI 200 index futures contract traded on the Sydney Futures Exchange (SFE), over the period June 1993 through June 1994. Since the SPI futures options are subject to futures style marging, implied volatilities are derived from the Asay (1982) model, which modifies the Black (1976) futures option pricing model to account for the daily marging of the option contract. 2,488 call options and 2,029 put options traded at the same time with their underlying futures are chosen and assigned into 99 groups according to moneyness and days to maturity. Then two dimensional graphs of implied volatility against moneyness and three dimensional graphs of implied volatility against both moneyness and maturity of

---

24 Another exchange which also margins its futures options is the LIFFE.
the option are plotted to illustrate the volatility structure of the SPI futures options.

Brown (1999) provides evidence of volatility skew for both call and put implied volatilities. In particular, in-the-money call options have a tendency to trade at higher implied volatilities than out-of-the-money options, while the reverse is true for put options. This volatility skew pattern for SPI futures options is consistent with the pattern hypothesized by Dumas, Flemming and Whaley (1998) for S&P 500 options since the stock market crash of 1987. Rather than focusing on the pricing biases of the Asay (1982) model, Brown (1999) explains this shape of volatility structure by viewing implied volatilities as prices reflective of the willingness of market participants to take on and lay off the risk of volatility changing and other risks not priced by the model. Therefore, the volatility structure of SPI futures options is largely driven by the demand by institutional traders for option strategies that protect equity portfolios, and by the requirement of no arbitrage opportunity in the market.
3.3 Data and methodology

3.3.1 Data

This chapter uses daily closing option prices between April 2001 and June 2005 to examine the volatility structure implied by the XJO index options on the ASX.

Since the XJO options are European dividend-paying options, the upper and lower boundary conditions of option prices are

\[ Fe^{-rT} \geq C \geq \max\left( (F - X) e^{-rT}, 0 \right) \]  \hspace{1cm} (10.1)

for call options, and

\[ X e^{-rT} \geq P \geq \max\left( (X - F) e^{-rT}, 0 \right) \]  \hspace{1cm} (10.2)

for put options. If an option price is above the upper bound or below the lower bound, there are profitable opportunities for arbitrageurs. However, in a competitive market, any available arbitrage opportunities disappear very quickly. For the purpose of analyses, it is reasonable to assume that there are no arbitrage opportunities. Hence, option records that violate one of these conditions are excluded.

\(^{25}\) For non-dividend paying European options, the boundary conditions are

\[ S_0 \geq C \geq \max(S_0 - X e^{-rT}, 0) \]

for call options, and

\[ X e^{-rT} \geq P \geq \max((X e^{-rT} - S_0), 0) \]

for put options. For dividend paying options, the boundary conditions are constructed simply by replacing \( S_0 \) with \( S_0 e^{-qT} \), where \( q \) denotes the annualized continuously compounded dividend yield during the remaining life of the option. In this thesis, to estimate \( q \), the relation between spot futures prices and spot index level, \( F = S_0 e^{(r-q)T} \), is utilised. Hence, the boundary conditions for XJO options are obtained according to \( S_0 e^{-qT} = F e^{-rT} \).
Consequently, 12,486 call option contracts and 14,634 put option contracts are remained to investigate the volatility structure implied by the XJO index options.

### 3.3.2 Methodology

Since the XJO options are European style and dividend paying options, the Merton (1973) model is employed to derive implied volatilities from the observed market prices. The Merton (1973) model generalizes the Black and Scholes (1973) model by relaxing the assumption of no dividend paid during the life of options and allows for a constant continuous dividend yield on the stock or stock index. To obtain the dividend yield estimation, the SFE SPI 200 futures contracts are utilised.\(^{26}\)

Then two dimensional graphs are plotted to examine the effect of volatility smile or sneer, i.e. how implied volatilities vary across different strike prices for options with same time to maturity on a particular day in the data set. The empirical results will be presented in Section 3.4.1.

Furthermore, to examine how volatility smile or sneer is affected by time to maturity, three dimensional graphs of implied volatility against moneyness and maturity of the option are plotted. This is done by assigning all the options records under investigation into groups according to moneyness and time to maturity, and then taking an average for implied volatilities in each group.

---

\(^{26}\) For details on the measurement of implied volatility, please refer to Section 2.4.2.
In particular, the 12,486 call option contracts in the data set are assigned into 12 moneyness (S/X-1) intervals and 11 days to maturity intervals, i.e. 132 groups. Within each group the average implied volatility is calculated. It is necessary to mention that there is an implicit assumption when taking average in each group, as suggested by Brown (1999), that the trends in implied volatility will influence implied volatilities from options with different strike prices and different maturities equally. Similarly, the 14,634 put option contracts are placed into 12 moneyness (X/S-1) intervals and 9 days to maturity intervals\(^{27}\), i.e. 108 groups. The empirical results will be presented in Section 3.4.2.

\(^{27}\) Since put options with time to maturity larger than 160 trading days are not traded as frequently as other put options with shorter life, these put options are assigned into one expiration interval.
3.4 Empirical results

3.4.1 Volatility smile and sneer

The inconsistency between the market prices of XJO options and the prices predicted by the dividend-adjusted BS model is illustrated in Figures 3.1 and 3.2 for call and put options, respectively. All the graphs in both figures are yielded from particular days in the data set, and represent the variety of shapes presented in the data set. Either a linear or a polynomial of best fit\(^{28}\) has been superimposed on the Implied volatility-Moneyness plots. Data used to create the graphs in Figures 3.1 and 3.2 are provided in Tables 3.1 and 3.2 for call and put options, respectively.

The call option implied volatilities illustrated in Figure 3.1 display a sneer pattern. This sneer or skew pattern conforms to the general shape hypothesized by Dumas, Flemming and Whaley (1998) for S&P 500 options since the 1987 stock market crash, and those shapes in Brown (1999) for call options on the SFE SPI 200 index futures over the period June 2003 through June 2004. That is, call option implied volatilities decrease monotonically as the call goes deeper out of the money. Dumas, Flemming and Whaley (1998) explain the emergence of the volatility sneer after the crash by an increase in investors’ probability assessment of downward moves in the index level.

\(^{28}\) For all graphs presented in Figure 3.1 and 3.2, the polynomial of best fit with order 2 is superimposed when the second derivative of volatility with respect to strike prices is positive; otherwise a linear of best fit is applied.
Figure 3.1: Graphs of call option implied volatility against moneyness (S/X-1)
Figure 3.1 – Continued

Table 3.1: Data used to create graphs in Figure 3.1

<table>
<thead>
<tr>
<th>Date</th>
<th>No. of observations</th>
<th>Days to maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td>25/10/2001</td>
<td>12</td>
<td>44</td>
</tr>
<tr>
<td>13/06/2002</td>
<td>7</td>
<td>70</td>
</tr>
<tr>
<td>19/08/2003</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>18/11/2004</td>
<td>9</td>
<td>80</td>
</tr>
<tr>
<td>12/01/2005</td>
<td>12</td>
<td>44</td>
</tr>
</tbody>
</table>
Figure 3.2: Graphs of put option implied volatility against moneyness (X/S-1)
Figure 3.2-Continued

![Graph of implied volatility vs moneyness for 2003-3-14, 2004-1-6, and 2005-6-1](image)

Table 3.2: Data used to create graphs in Figure 3.2

<table>
<thead>
<tr>
<th>Date</th>
<th>No. of observations</th>
<th>Days to maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td>7/09/2001</td>
<td>8</td>
<td>78</td>
</tr>
<tr>
<td>20/08/2002</td>
<td>11</td>
<td>22</td>
</tr>
<tr>
<td>18/12/2002</td>
<td>13</td>
<td>62</td>
</tr>
<tr>
<td>14/03/2003</td>
<td>13</td>
<td>65</td>
</tr>
<tr>
<td>6/01/2004</td>
<td>19</td>
<td>51</td>
</tr>
<tr>
<td>1/06/2005</td>
<td>13</td>
<td>74</td>
</tr>
</tbody>
</table>
Figure 3.2 illustrates the typical patterns in the put option implied volatilities. Strikingly, both sneer and smile patterns are displayed. The sneer pattern conforms to the general shape hypothesized by Dumas, Fleming and Whaley (1998) and Brown (1999), with put option implied volatilities decrease monotonically as the put goes deeper in the money. Some smile pattern graphs, e.g. the second graph in Figure 3.2, illustrate that the put option implied volatilities are symmetric around zero moneyness, with in-the-money and out-of-the-money options trading at higher implied volatilities than at-the-money options. While other graphs with smile pattern, e.g. the third graph in Figure 3.2, illustrate that the put option implied volatilities is symmetric around a positive moneyness, with at-the-money options having higher implied volatilities than some less deep in-the-money options.

### 3.4.2 3-D view of volatility structure

To investigate the volatility structure further, three dimensional graphs of implied volatilities against moneyness and time to maturity of the option are plotted in Figures 3.3 and 3.4 for call and put options, respectively. This is done by assigning all the options records under investigation into groups according to moneyness and expiration, and then taking an average for implied volatilities in each group. Tables 3.3 and 3.4 report the average implied volatility within each group for call and put options, respectively.

Figure 3.3 shows some evidence of volatility sneer for the XJO call options. The volatility sneer tends to be downward sloping. This means that in-the-money call options on average trade at higher implied volatilities than out-of-the-money call
options, except for the short maturity options. For the options with short life, a volatility smile pattern is more pronounced, with deep-in-the-money and deep-out-of-money call options having higher implied volatilities than at-the-money call options. The term structure of implied volatility is nearly flat for at-the-money call options, but more fluctuant for deep-in-the-money and deep-out-of-money call options.

Figure 3.4 illustrates both smile and sneer patterns for the XJO put options. In particular, the volatility smile pattern is more pronounced for long maturity put options, e.g. $T > 160$ days, and short maturity put options, e.g. $T < 60$ days. The volatility pattern for options’ maturities between 60 and 160 days is more of a sneer, with out-of-the-money put options on average trading at higher implied volatilities than in-the-money put options. The volatility term structure of put option is similar to that of call options. That is, the term structure is close to be flat for at-the-money put options, and tends to be fluctuant for deep-in-the-money and deep-out-of-money put options.

Overall, these results indicate that the dividend-adjusted Black-Scholes model used in this study tends to overprice in-the-money call options and out-of-the-money put options and underprice in-the-money put options and out-of-the-money call options. At-the-money options are often most actively traded, and hence they are less likely to be mis-priced.
Figure 3.3: 3-D view of volatility structure for call options

This figure illustrates a 3D-view of the volatility structure implied by the XJO call options over the period from April 2001 to June 2005. The 12486 call option contracts in the data set are assigned into one of 132 groups by moneyness and time to maturity. Implied volatility is obtained by numerically solving the dividend-adjusted Black-Scholes formula for each call option contract, and then an average implied volatility is calculated for each group. Days to maturity is the number of trading days left to expiration for a call option. Moneyness is defined as S/X-1 for the XJO call options.
Table 3.3: Group averages of call option implied volatilities

| Days to maturity | Moneyness (S/X-1) | <=-0.10 | (-0.10, -0.08] | (-0.08, -0.06] | (-0.06, -0.04] | (-0.04, -0.02] | (-0.02, 0.00] | (0.00, 0.02] | (0.02, 0.04] | (0.04, 0.06] | (0.06, 0.08] | (0.08, 0.10] | >0.10 |
|-----------------|-------------------|--------|-----------------|-----------------|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-------|
| (0,20]          |                   | 0.402  | 0.350           | 0.161           | 0.151           | 0.127          | 0.112          | 0.116          | 0.168          | 0.240          | 0.281          | 0.372          | 0.577 |
| (20,40]         |                   | 0.180  | 0.165           | 0.134           | 0.118           | 0.110          | 0.110          | 0.118          | 0.124          | 0.144          | 0.158          | 0.203          | 0.257 |
| (40,60]         |                   | 0.153  | 0.138           | 0.130           | 0.116           | 0.112          | 0.113          | 0.120          | 0.131          | 0.141          | 0.150          | 0.163          | 0.215 |
| (60,80]         |                   | 0.164  | 0.132           | 0.121           | 0.112           | 0.108          | 0.111          | 0.115          | 0.125          | 0.135          | 0.137          | 0.144          | 0.189 |
| (80,100]        |                   | 0.129  | 0.121           | 0.110           | 0.110           | 0.111          | 0.113          | 0.119          | 0.127          | 0.141          | 0.147          | 0.158          | 0.194 |
| (100,120]       |                   | 0.138  | 0.127           | 0.124           | 0.120           | 0.117          | 0.121          | 0.127          | 0.150          | 0.138          | 0.153          | 0.157          | 0.245 |
| (120,140]       |                   | 0.146  | 0.129           | 0.119           | 0.118           | 0.119          | 0.115          | 0.124          | 0.124          | 0.135          | 0.131          | 0.165          | 0.164 |
| (140,160]       |                   | 0.117  | 0.113           | 0.113           | 0.118           | 0.112          | 0.113          | 0.122          | 0.123          | 0.129          | 0.141          | 0.130          | 0.162 |
| (160,180]       |                   | 0.113  | 0.119           | 0.111           | 0.116           | 0.126          | 0.124          | 0.130          | 0.126          | 0.138          | 0.141          | 0.194          | 0.186 |
| (180,200]       |                   | 0.123  | 0.123           | 0.117           | 0.120           | 0.120          | 0.118          | 0.117          | 0.136          | 0.128          | 0.127          | 0.151          | 0.146 |
| (200,282]       |                   | 0.109  | 0.112           | 0.112           | 0.110           | 0.112          | 0.118          | 0.132          | 0.116          | 0.133          | 0.139          | 0.129          | 0.142 |
Figure 3.4: 3-D view of volatility structure for put options

This figure illustrates a 3-D view of the volatility structure implied by the XJO put options over the period from April 2001 to June 2005. The 14634 put option contracts in the data set are assigned into one of 132 groups by moneyness and time to maturity. Implied volatility is obtained by numerically solving the dividend-adjusted Black-Scholes formula for each put option contract, and then an average implied volatility is calculated for each group. Days to maturity is the number of trading days left to expiration for a put option. Moneyness is defined as X/S-1 for the XJO put options.
Table 3.4: Group averages of put option implied volatilities

| Days to maturity | Moneyness (X/S-1) | <=-0.10 | (-0.10, -0.08] | (-0.08, -0.06] | (-0.06, -0.04] | (-0.04, -0.02] | (-0.02, 0.00] | (0.00, 0.02] | (0.02, 0.04] | (0.04, 0.06] | (0.06, 0.08] | (0.08, 0.10] | >0.10 |
|------------------|------------------|--------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|------|
| (0,20)           |                  | 0.337  | 0.225          | 0.209          | 0.176          | 0.149          | 0.127          | 0.129          | 0.204          | 0.303          | 0.438          | 0.486          | 0.764         |
| (20,40)          |                  | 0.210  | 0.178          | 0.164          | 0.150          | 0.132          | 0.123          | 0.115          | 0.121          | 0.138          | 0.164          | 0.201          | 0.344         |
| (40,60)          |                  | 0.194  | 0.178          | 0.162          | 0.150          | 0.135          | 0.126          | 0.126          | 0.126          | 0.137          | 0.156          | 0.194          | 0.182         |
| (60,80)          |                  | 0.179  | 0.161          | 0.152          | 0.137          | 0.128          | 0.123          | 0.123          | 0.134          | 0.126          | 0.142          | 0.152          | 0.182         |
| (80,100]         |                  | 0.179  | 0.160          | 0.149          | 0.142          | 0.130          | 0.127          | 0.124          | 0.124          | 0.108          | 0.119          | 0.104          | 0.144         |
| (100,120]        |                  | 0.187  | 0.163          | 0.159          | 0.144          | 0.135          | 0.132          | 0.134          | 0.138          | 0.136          | 0.133          | 0.138          | 0.138         |
| (120,140]        |                  | 0.177  | 0.158          | 0.145          | 0.140          | 0.136          | 0.131          | 0.126          | 0.125          | 0.131          | 0.141          | 0.130          | 0.127         |
| (140,160]        |                  | 0.174  | 0.151          | 0.142          | 0.137          | 0.136          | 0.130          | 0.127          | 0.116          | 0.107          | 0.109          | 0.170          | 0.155         |
| (160,283]        |                  | 0.187  | 0.158          | 0.143          | 0.137          | 0.130          | 0.123          | 0.121          | 0.117          | 0.107          | 0.154          | 0.120          | 0.187         |
3.5 Conclusion

In this chapter, the volatility structure implied by the XJO options is illustrated by two dimensional graphs of implied volatility against moneyness and three dimensional graphs of implied volatility against both moneyness and time to maturity. It is found that the sneer patterns are pronounced for most of the call and put options implied volatilities, with call (put) option implied volatilities decreasing monotonically as the call (put) goes deeper out of the money (in the money). This result conforms to the general shape hypothesized by Dumas, Flemming and Whaley (1998) for S&P 500 options since the 1987 stock market crash. Some volatility smile patterns are also observed for the put options with T>160 days and T< 60 days. The term structure of implied volatility is nearly flat for at-the-money call and put options, but more fluctuant for deep-in-the-money and deep-out-of-money options.

Overall, these results indicate that the dividend-adjusted Black-Scholes model used in this study tends to overprice in-the-money call options and out-of-the-money put options and underprice in-the-money put options and out-of-the-money call options. At-the-money options are often most actively traded, and hence they are less likely to be mispriced.
Chapter 4  Summary and conclusions

This thesis is concerned with the implied volatility from the S&P/ASX 200 (XJO) index options on the ASX. Two interesting problems on implied volatility have been thoroughly investigated. The relation between implied and realized volatility is first investigated in Chapter 2 and the structure of implied volatility is then considered in Chapter 3.

In Chapter 2, the relation between implied volatility and subsequently realized volatility is investigated in the XJO options market during a period of five years from April 2001 to March 2006. Considering the presence of the EIV problem in the measurement of implied volatility, both OLS and IV estimation methods are employed. The results from both methods are compared and indicate that both implied call and put volatility are better than historical volatility in forecasting future realized volatility, whether assessed by the Adjusted-$R^2$ of each regression or by the magnitude of the regression slope coefficients. Moreover, implied call volatility is nearly an unbiased forecast of future volatility after correcting for the EIV problem by using the IV method.

In Chapter 3, the volatility structure implied by the XJO options is illustrated by two dimensional graphs of implied volatility against moneyness and three dimensional graphs of implied volatility against both moneyness and time to maturity. It is found that the sneer patterns are pronounced for most of the call and put options implied volatilities. This result conforms to the general shape
hypothesized by Dumas, Flemming and Whaley (1998) for S&P 500 options since the 1987 stock market crash, with call (put) option implied volatilities decreasing monotonically as the call (put) goes deeper out of the money (in the money). Some volatility smile patterns are also observed for put options with $T>160$ days and $T<60$ days. The term structure of implied volatility is nearly flat for at-the-money call and put options, but more fluctuant for deep-in-the-money and deep-out-of-money options. These results indicate that the misprice problem of the dividend-adjusted Black-Scholes model used in this study largely exists for in-the-money and out-of-the-money options, but is less severe for at-the-money options.

This thesis makes a number of contributions to the literature on the implied volatility in the Australian index option market setting. First, this thesis is the first one to investigate the relation between realized and implied volatility in the context of Australian S&P/ASX 200 index option market. Second, it is also the first study on the structure of the implied volatility of the XJO options. These findings have important implications to the efficiency of the Australian index options market. In addition, some mathematical derivations on the effects of the EIV problem as well as on the consistent estimator from 2SLS are provided.

One possible extension of this study is to examine how the relation between implied and realized volatility is influenced by different measurement methods of realized volatility, as outlined in Appendix 2. Especially, when realized volatility series is constructed from intraday high frequency data, the forecasting power of implied volatility is expected to be improved largely.
Reference list


Appendices

Appendix 1. Numerical search procedure for implied volatility

The implied volatility was calculated using a VBA (Visual Basic for Applications) function in which an iterative loop procedure is employed. The program is constructed to make the difference between the price calculated from Merton (1973) formulae and the price observed from option market approached to zero. This technique is very similar to the technique used in trial and error. Start with two estimates for the possible volatility. A high estimate of 1 and a low estimate of 0 were used. The average of the initial high and the initial low values, i.e. \((\text{high}+\text{low})/2\), was deemed to be the first estimate and was plugged into the Merton (1973) formulae. If the calculated price was larger than the market price, then the current estimate of \((\text{high}+\text{low})/2\) is too high and the high estimate is replaced by \((\text{high}+\text{low})/2\). If the calculated price was less than the market price, then the current estimate of \((\text{high}+\text{low})/2\) is too low and the low estimate is replaced by \((\text{high}+\text{low})/2\). This procedure was repeated until the difference between the high and the low estimates is less than 0.0001 (or some other desired level of accuracy). Both implied call and put volatilities series are constructed in this manner.
Appendix 2. Several other measure methods of realized volatility

The Parkinson’s (1980) extreme value estimator is

$$\sigma_{t, t} = \sqrt{\frac{252}{n} \sum_{t=1}^{n} \frac{1}{4 \ln 2} (\ln H_t - \ln L_t)^2},$$  \hspace{1cm} (A2.1)$$

where $H_t$ and $L_t$ are the daily high and low prices on date $t$. This estimator does not require the volatility to be constant over the estimation period.

The Yang and Zhang (2000) estimator is a range-based volatility estimator, which incorporates daily high and low as well as daily open and close prices as follows,

$$\hat{\sigma} = \sqrt{252 \times \sqrt{V_0 + kV_c + (1-k)V_{RS}}},$$  \hspace{1cm} (A2.2)$$

where $V_0, V_c$ and $V_{RS}$, are defined as follows:

$$V_0 = \frac{1}{n-1} \sum_{t=1}^{n} (o_t - \bar{\sigma})^2, \quad o_t = \ln O_t - \ln O_{t-1}, \quad \bar{\sigma} = \frac{1}{n} \sum_{t=1}^{n} o_t,$$

$$V_c = \frac{1}{n-1} \sum_{t=1}^{n} (c_t - \bar{c})^2, \quad c_t = \ln C_t - \ln C_{t-1}, \quad \bar{c} = \frac{1}{n} \sum_{t=1}^{n} c_t,$$

$$V_{RS} = \frac{1}{n} \sum_{t=1}^{n} [(\ln H_t - \ln O_t)(\ln H_t - \ln C_t) + (\ln L_t - \ln O_t)(\ln L_t - \ln C_t)]$$

where $O_t$ and $C_t$ are the daily open and close prices on date $t$, and the constant $k$, chosen to minimise the variance of the estimator, is given by:

$$k = \frac{0.34}{1.34 + \frac{n+1}{n-1}}.$$
This volatility estimator explicitly incorporates the opening jump into pricing formula, thus releases the continuous trading assumptions made by other volatility estimators. Since opening jump always happens when unexpected information comes during non-trading time, the incorporation of opening jump may lead to a better measurement of the realized volatility.

The last volatility estimator outlined here is calculated from the intraday returns as following:

$$\hat{\sigma} = \sqrt{\frac{1}{\Delta} \left[ \frac{1}{N} \sum_{i=1}^{N} R_{t+i\Delta}^2 \right]}, \quad (A2.3)$$

where $\Delta$ denotes the annualized fixed interval of trading time, $N$ denotes the number of such intervals in a trading day, and $R_{t+i\Delta}$ denotes the return during the period from $t + (i-1)\Delta$ to $t + i\Delta$. This estimator is named as “integrate volatility” by Andersen (2000). Since the high-frequency data is closer to the continuous data generating process and contains more information in forecasting future volatility, it is natural to expect that the volatility estimator constructed from high frequency data can improve volatility forecast ability. However, the use of high-frequency data exceeds the extent of this thesis and hence can be left to future work.
Appendix 3. Derivation of the effects of ElV problem in Equation (2.19)

This appendix presents a derivation of the probability limit of the OLS slope estimate $\hat{\alpha}$, as shown in Equation (2.19), for the univariate regression model (2.10).

Start by providing an expression for $\hat{\alpha}$ in terms of the regressor and error. Because $rv_t = \alpha_0 + \alpha_1 iv_t + \varepsilon_t$, $rv_t - r\bar{v} = \alpha_1 (iv_t - i\bar{v}) + \varepsilon_t - \bar{\varepsilon}$, so the numerator of the formula for $\hat{\alpha}$ in Equation (2.18) is

$$\sum_{t=1}^{n} (iv_t - i\bar{v})(rv_t - r\bar{v}) = \sum_{t=1}^{n} (iv_t - i\bar{v})[\alpha_1 (iv_t - i\bar{v}) + \varepsilon_t - \bar{\varepsilon}]$$

$$= \alpha_1 \sum_{t=1}^{n} (iv_t - i\bar{v})^2 + \sum_{t=1}^{n} (iv_t - i\bar{v})(\varepsilon_t - \bar{\varepsilon})$$

(A3.1)

Now following the definition of $i\bar{v}$ yields

$$\sum_{t=1}^{n} (iv_t - i\bar{v})(\varepsilon_t - \bar{\varepsilon}) = \sum_{t=1}^{n} (iv_t - i\bar{v})\varepsilon_t - \sum_{t=1}^{n} (iv_t - i\bar{v})\bar{\varepsilon}$$

$$= \sum_{t=1}^{n} (iv_t - i\bar{v})\varepsilon_t - \left[ \sum_{t=1}^{n} iv_t - n i\bar{v} \right] \bar{\varepsilon} = \sum_{t=1}^{n} (iv_t - i\bar{v})\varepsilon_t$$

(A3.2)

Substituting (A3.2) into (A3.1) yields

$$\sum_{t=1}^{n} (iv_t - i\bar{v})(rv_t - r\bar{v}) = \alpha_1 \sum_{t=1}^{n} (iv_t - i\bar{v})^2 + \sum_{t=1}^{n} (iv_t - i\bar{v})\varepsilon_t$$
Substituting this expression into the formula for $\hat{\alpha}_i$ in Equation (2.18) yields

$$\hat{\alpha}_i = \alpha_i + \frac{(1/n) \sum_{t=1}^{n} (iv_t - \bar{v}) \epsilon_t}{(1/n) \sum_{t=1}^{n} (iv_t - \bar{v})^2} \quad (A3.3)$$

Under the least squares assumption, the following probability limits can be obtained:

$$p \lim \left[ \frac{1}{n} \sum_{t=1}^{n} (iv_t - \bar{v})^2 \right] = \sigma_{iv}^2 \quad \text{and}$$

$$p \lim \left[ \sum_{t=1}^{n} (iv_t - \bar{v}) \epsilon_t \right] = \text{cov}(iv_t, \epsilon_t).$$

Substitution of these limits into Equation (A3.3) yields

$$p \lim \hat{\alpha}_i = \alpha_i + \frac{\text{cov}(iv_t, \epsilon_t)}{\sigma_{iv}^2} \quad (A3.4)$$

Because of $\text{cov}(iv_t, \epsilon_t) = -\alpha_i \sigma_{iv}^2$ and $\sigma_{iv}^2 = \sigma_{iv}^2 + \sigma_{iv}^2$, (A3.4) can be reduced to (2.19).
Appendix 4. Derivation of the effects of EIV problem in
Equation (2.23) and (2.24)

This appendix presents derivations of the probability limit of the OLS slope estimates $\hat{\alpha}_i$ and $\hat{\alpha}_h$, as shown in Equation (2.23) and (2.24), for the multiple regression model (2.11).

Start by writing $\hat{\alpha}_i$ and $\hat{\alpha}_h$ in terms of the regressors and error term. Firstly, because

$$rv_i = \alpha_0 + \alpha_i iv_i + \alpha_h hv_{t-1} + \epsilon_i,$$
$$rv_i - r\bar{v} = \alpha_i (iv_i - i\bar{v}) + \alpha_h (hv_{t-1} - h\bar{v}) + \epsilon_i - \bar{\epsilon},$$

$$\sum_{i=1}^{n} (iv_i - i\bar{v})(\epsilon_i - \bar{\epsilon}) = \sum_{i=1}^{n} (iv_i - i\bar{v})\epsilon_i,$$
$$\sum_{i=1}^{n} (hv_{t-1} - h\bar{v})(\epsilon_i - \bar{\epsilon}) = \sum_{i=1}^{n} (hv_{t-1} - h\bar{v})\epsilon_i,$$

$\sum_{i=1}^{n} IV_i RV_i$ and $\sum_{i=1}^{n} HV_{t-1} RV_i$ can be written as following:

$$\sum_{i=1}^{n} IV_i RV_i = \sum_{i=1}^{n} (iv_i - i\bar{v})(rv_i - r\bar{v})$$
$$= \sum_{i=1}^{n} (iv_i - i\bar{v})[\alpha_i (iv_i - i\bar{v}) + \alpha_h (hv_{t-1} - h\bar{v}) + \epsilon_i - \bar{\epsilon}]$$
$$= \alpha_i \sum_{i=1}^{n} (iv_i - i\bar{v})^2 + \alpha_h \sum_{i=1}^{n} (iv_i - i\bar{v})(hv_{t-1} - h\bar{v}) + \sum_{i=1}^{n} (iv_i - i\bar{v})\epsilon_i$$
$$= \alpha_i \sum_{i=1}^{n} IV_i^2 + \alpha_h \sum_{i=1}^{n} IV_i HV_{t-1} + \sum_{i=1}^{n} IV_i \epsilon_i$$

(A4.1)
\[
\sum_{t=1}^{n} HV_{t-1}RV_t = \sum_{t=1}^{n} (hv_{t-1} - h\bar{v})(rv_t - r\bar{v}) \\
= \sum_{t=1}^{n} (hv_{t-1} - h\bar{v})[\alpha_i (iv_t - i\bar{v}) + \alpha_h (hv_{t-1} - h\bar{v}) + \varepsilon_t - \bar{\varepsilon}] \\
= \alpha_i \sum_{t=1}^{n} (iv_t - i\bar{v})(hv_{t-1} - h\bar{v}) + \alpha_h \sum_{t=1}^{n} (hv_{t-1} - h\bar{v})^2 + \sum_{t=1}^{n} (hv_{t-1} - h\bar{v})\varepsilon_t \\
= \alpha_i \sum_{t=1}^{n} IV_tHV_{t-1} + \alpha_h \sum_{t=1}^{n} HV_{t-1}^2 + \sum_{t=1}^{n} HV_{t-1}\varepsilon_t \\
\text{(A4.2)}
\]

Substituting (A4.1) and (A4.2) into (2.23) and (2.24) yields

\[
\hat{\alpha}_i = \frac{\sum_{t=1}^{n} IV_t^2 + \alpha_i \sum_{t=1}^{n} IV_tHV_{t-1} + \sum_{t=1}^{n} IV_t\varepsilon_t \sum_{t=1}^{n} HV_{t-1} - \left( \alpha_i \sum_{t=1}^{n} IV_tHV_{t-1} + \alpha_h \sum_{t=1}^{n} HV_{t-1}\varepsilon_t + \sum_{t=1}^{n} IV_t\varepsilon_t \right) \sum_{t=1}^{n} HV_{t-1}}{\sum_{t=1}^{n} IV_t^2 \sum_{t=1}^{n} HV_{t-1} - \left( \sum_{t=1}^{n} IV_tHV_{t-1} \right)^2} \\
\text{(A4.3)}
\]

\[
\hat{\alpha}_h = \frac{\sum_{t=1}^{n} IV_tHV_{t-1} + \alpha_h \sum_{t=1}^{n} HV_{t-1}\varepsilon_t \sum_{t=1}^{n} IV_t^2 - \left( \alpha_i \sum_{t=1}^{n} IV_tHV_{t-1} + \alpha_h \sum_{t=1}^{n} HV_{t-1}\varepsilon_t + \sum_{t=1}^{n} IV_t\varepsilon_t \right) \sum_{t=1}^{n} IV_tHV_{t-1}}{\sum_{t=1}^{n} IV_t^2 \sum_{t=1}^{n} HV_{t-1} - \left( \sum_{t=1}^{n} IV_tHV_{t-1} \right)^2} \\
\text{(A4.4)}
\]

Under the least squares assumption, the following probability limits can be obtained:

\[ p \lim \sum_{t=1}^{n} IV_t^2 = \sigma_{iv}^2, \]

\[ p \lim \sum_{t=1}^{n} HV_{t-1}^2 = \sigma_{hv}^2, \]

\[ p \lim \sum_{t=1}^{n} IV_tHV_{t-1} = \text{cov}(iv_t, hv_{t-1}) = \text{cov}(iv_t^*, u_t, hv_{t-1}) = \rho \sigma_{iv} \sigma_{hv}, \]
\[ p \lim \sum_{t=1}^{n} IV_t \varepsilon_t = \text{cov}(iv_t, \varepsilon_t) = -\alpha_i \sigma^2_u, \]

\[ p \lim \sum_{t=1}^{n} HV_{t-1} \varepsilon_t = \text{cov}(hv_{t-1}, \varepsilon_t) = 0, \]

where \( \sigma^2_{iv}, \sigma^2_{iv*}, \sigma^2_{hv}, \) and \( \sigma^2_u \) denote the variances of mismeasured implied volatility, true implied volatility, historical volatility, and the measurement error, respectively; and \( \rho \) denotes the correlation between true implied volatility and historical volatility.

Substituting these limits and \( \sigma^2_{iv} = \sigma^2_{iv*} + \sigma^2_u \) into (A4.3) and (A4.4), the probability limits of \( \hat{\alpha}_i \) and \( \hat{\alpha}_h \) can be obtained, as shown in (2.23) and (2.24).
Appendix 5. Proof of consistent estimator from 2SLS

Consider once again the univariate regression model (2.10)

\[ rv_i = \alpha_0 + \alpha_i iv_i + \epsilon_i, \]

and the errors in measurement of implied volatility (2.16)

\[ iv_i = iv_i^* + u_i. \]

Suppose that there exists a variable \( z \) such that \( z \) is correlated with \( iv \) but not with the measurement error \( \epsilon \). That is

\[ \text{cov}(iv_i, z_i) \neq 0, \]

\[ \text{cov}(\epsilon_i, z_i) = 0. \]

Note that

\[ \text{cov}(rv_i, z_i) = \text{cov}(\alpha_0 + \alpha_i iv_i + \epsilon_i, z_i) = \alpha_i \text{cov}(iv_i, z_i) + \text{cov}(\epsilon_i, z_i) = \alpha_i \text{cov}(iv_i, z_i). \]

where the second equality follows from the properties of covariance and third equality follows from the assumption (2.28). Thus, if the instrument \( z \) is valid,

\[ \alpha_i = \frac{\text{cov}(rv_i, z_i)}{\text{cov}(iv_i, z_i)}. \]

That is, the population coefficient \( \alpha_i \) is the ratio of the population covariance between \( rv \) and \( z \) to the population covariance between \( iv \) and \( z \). Utilising a two-stage-least-squares procedure (2SLS), an estimation of \( \alpha_i \) can be obtained. That is
where \( s_{rv,z} \) denotes the sample covariance between \( rv \) and \( z \), and \( s_{iv,z} \) denotes the sample covariance between \( iv \) and \( z \). Because the sample covariance is a consistent estimator of the population covariance, the following probability limits hold,

\[
p \lim s_{rv,z} = \text{cov}(rv_t, z_t)
\]

(A5.6)

\[
p \lim s_{iv,z} = \text{cov}(iv_t, z_t)
\]

(A5.7)

Thus, it follows from Equation (A5.4)-(A5.7) that the TSLS estimator is consistent:

\[
p \lim \alpha_i^{TSL} = \frac{p \lim s_{rv,z}}{p \lim s_{iv,z}} = \frac{\text{cov}(rv_t, z_t)}{\text{cov}(iv_t, z_t)} = \alpha_i
\]

(A5.8)

In the case of multiple regression (2.11), with mismeasured implied volatility and historical volatility as the regressors, a similar derivation can be made but just be cumbersome\(^{29}\).