Bayesian Methods for Estimation, Inference and Forecasting of Flexible Models for Value-at-Risk and Tail Conditional Expectations

Qian Chen

A thesis submitted in partial fullfilment of the requirements

for the degree of Doctor of Philosophy

Discipline of Operations Management and Econometrics

Business School, University of Sydney

October 2011

Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Qian Chen

30.03.2011

Acknowledgements

First and foremost I offer my sincerest gratitude to my supervisor, Dr Richard Gerlach, who has supported me thoughout my thesis with his patience and knowledge whilst allowing me the room to work in my own way. I attribute my publications during the PhD program to his encouragement and effort as a coauthor, and without him this thesis would not have been completed or written. One simply could not wish for a better or friendlier supervisor. I also received warm care from Richard's family, his three lovely boys Thomas, Mathew and James pray for me every night before going to bed. This brings a lot of comfort to me in this long journey in a foreign country.

In my first project and publication, Dr. Zudi Lu provided me with outstanding ideas and contributed as a coauthor and derived the formula and proof for the necessary and sufficient condition for second-order stationarity of the GJR-GARCH-ALD model.

Dr Boris Choy has offered advice and insight throughout my work, and encouraged me to present my work in international conferences, where I was greatly inspired to improve my projects.

I received consistent support from the Discipline of Operations Management and Econometrics that I need to complete the thesis. The Chair of the DOME Professor Eddie Anderson offered research funding for me to present my second project at ASC 2010. Discipline Administrative officers Kandice Cherry and Darae Jung provided me with kind help in their everyday work.

I benefit from the Faculty Research Unit of Business School, from whom I received financial support and beyond to complete the program and thesis, and the travel expense to attend international conferences.

In my daily work I have been blessed with a friendly and cheerful group of fellow students. Susana Sze and En Li shared their experience and information in writing thesis and organizing documents for submission. Min Zhu helped me regain some sort of fitness: healthy body, healthy mind by pushing me to attend the gym and group fitness classes. Yibai Yang helped on the road of LaTeX, including reminding me of updated version and packages.

My friends Qian Wei, Ying Yao and Qiuyu Sun, although doing PhD in other countries, they entertained me with all kinds of jokes and funny things in their PhD journey, and have comforted me with their company and great personalities.

Last but not the least, I thank my parents for understanding my dream and supporting me through the study at the University of Sydney.

Abstract

Forecasting financial risk and risk measurement methods have been of increasing interest for financial market regulators and financial institutions in the past two decades. While the parametric and semi-parametric models have been widely reviewed in the academic literature, the non-parametric methods are popular in practice among the financial institutions. This thesis examines the forecasting models for Value-at-Risk (VaR) and conditional Value-at-Risk for financial return series.

The aims of this thesis are to:

1. Estimate and forecast the potential skewness and dynamics in higher moments for conditional return distributions;

2. Develop flexible parametric models that can accurately forecast the portfolio tail risk levels.

3. Examine the impacts of asymmetry in the volatility and that in the shape of the conditional return distributions on the risk level forecasting.

4. Derive an easily applicable backtesting method for conditional VaR or expected shortfall.

5. Improve the efficiency and accuracy of Bayesian computational schemes for parameter estimation and forecasts.

To achieve the above goals, this thesis first proposes a parametric approach to estimating and forecasting Value-at-Risk (VaR) and Expected Shortfall (ES) for a heteroscedastic financial return series. A GJR-GARCH is used to model the volatility process, capturing the leverage effect. To account for potential skewness and heavy tails, the model assumes an asymmetric Laplace (AL) distribution as the conditional distribution of the financial return series. Furthermore, dynamics in higher moments are captured by allowing for a time-varying shape parameter in this distribution. An adaptive Markov chain Monte Carlo (MCMC) sampling scheme is used for estimation, employing the Metropolis–Hastings (MH) algorithm with a mixture of Gaussian proposal distributions. A simulation

study shows accurate estimation and improved inference of parameters in comparison with a single Gaussian proposal MH method.

We illustrate the model by applying it to forecast return series from four international stock market indices, as well as two exchange rates, and generating one step-ahead forecasts of VaR and ES. We apply standard and non-standard tests to these forecasts, as well as to those from some competing methods, and find that the proposed model performs favourably compared to many popular competitors; in particular, it is the only conservative model of risk among the models considered in this work over the period studied, which includes the recent financial crisis.

However, an AL conditional ditribution may forecast risk too conservatively, and overestimate the risk levels by a factor of two. In other words, the model implies the necessity for financial institutions to set aside up to twice as much regulatory capital as they need. With fixed total capital, the capital available to invest is reduced, leading to a lowered profit potential. To address this dilemma, this study develops and employs a two-sided Weibull (TW) distribution to capture potential skewness and fat-tailed behaviour in the conditional financial return distribution for the purposes of risk measurement and management, specifically focusing on the forecasting of VaR and conditional VaR measures. Four volatility model specifications, including both symmetric and nonlinear versions, are considered to capture heteroscedasticity. An adaptive Bayesian MCMC scheme is devised for estimation, inference, and forecasting. A range of conditional return distributions (TW, AL, symmetric, and skewed Student t) are combined with the four volatility specifications to forecast risk measures.

The study finds that the GARCH-type volatility specification is much less important than that of the conditional distribution and, while the Student t distribution performs particularly well on VaR forecasting, the two-sided Weibull performs at least equally well for VaR, but the most favourably for conditional VaR forecasting, both prior to as well as during and after the recent financial crisis.

Nonetheless, the TW distribution can be bimodal, while the conditional distribution of real financial return series are known to be uni-modal. To address this issue, this study develops a partitioned distribution, combining the Weibull tails with a uni-modal AL

centre. The proposed distribution is combined with the GJR-GARCH volatility model, to estimate and forecast the VaR and Conditional VaR. The estimation is via an adaptive MCMC sampling scheme and the MH algorithm, with a more general and flexible mixture of Student t proposal distributions. A simulation study demonstrates the estimation is marginally closer to the true values than the mixture of Gaussian proposal distributions. The model is illustrated via application to real financial return series, generating one-day-ahead forecasts and is compared with several competing models. The forecasts are evaluated by formal and non-formal backtesting methods. The model-fitting performances are demonstrated by a range of residual tests. We find the partitioned distribution forecasts financial tail risks slightly less accurately than the TW, but is most favoured by the residual tests.

Contents

Statement of Originality **2**

Acknowledgement **1**

Abstract **1**

Contents **4**

List of Figures and Tables **9**

1 Introduction **16**

1.1 Financial Risk Measurement Methods 18

1.2 Models for Value-at-Risk and Conditional Value-at-Risk 19

1.2.1 Volatility models 20

1.2.2 The return distributions 22

1.3 Bayesian Inference and Forecasting 26

1.3.1 Monte Carlo Markov chain for GARCH models 27

1.4 Methodology for Evaluating the Forecast Performance 28

1.4.1 Back-text based on violations 28

1.4.2 Diagnostic hypothesis tests 29

1.5 Structure of the Thesis 30

1.6 Summary 31

2 Bayesian Value-at-Risk and Expected Shortfall Forecasting via the Asymmetric Laplace Distribution **33**

2.1 Model Specification 34

2.1.1 VaR and ES 34

2.1.2 Asymmetric Laplace distribution . . . 35
2.1.3 GJR-GARCH-ALD model with fixed shape . . . 39
2.1.4 Dynamic higher moments . . . 40
2.1.5 GJR-GARCH AL model with dynamic shape . . . 41
2.2 Bayesian MCMC Methods . . . 41
2.2.1 Bayesian methods . . . 41
2.2.2 Adaptive MCMC sampling using MH methods . . . 42
2.3 Simulation Study . . . 45
2.3.1 Simulation setup . . . 46
2.3.2 Simulation results . . . 48
2.4 Empirical Study . . . 48
2.4.1 Data . . . 48
2.4.2 Risk forecasting study . . . 54
2.4.2.1 One-step-ahead VaR and ES forecasting . . . 54
2.4.2.2 Ten-step-ahead VaR and ES forecasting . . . 56
2.4.3 Backtesting one-day-ahead VaR forecast models . . . 57
2.4.3.1 Overall forecast period performance . . . 58
2.4.3.2 Pre-financial-crisis and post-financial-crisis VaR performance 62
2.4.4 Theoretical expected shortfall and backtesting for 1-day-ahead ES forecast models . . . 64
2.4.4.1 Overall forecast period performance . . . 67
2.4.4.2 Pre-financial-crisis and post-financial-crisis performance . 68
2.4.5 Further model assessment . . . 70
2.4.5.1 Market risk charges . . . 70
2.4.5.2 Absolute deviation of violating returns . . . 72

2.4.5.3 Loss function . . . 75
2.4.6 Results summary and discussion for one-step-ahead risk forecasts . 77
2.4.7 A brief discussion of ten-step-ahead risk forecasts . . . 81
2.5 Conclusion . . . 85

3 The Two-sided Weibull Distribution and Forecasting Financial Tail Risk 87

3.1 A Two-sided Weibull Distribution . . . 88
3.1.1 Standardised Two-sided Weibull distribution . . . 89
3.1.2 VaR and tail conditional expectations for the two-sided Weibull . . 93
3.2 Model Specification . . . 94
3.2.1 Volatility models . . . 95
3.3 Estimation and Forecasting Methodology . . . 96
3.3.1 Bayesian estimation methods . . . 96
3.3.1.1 Simulation Results . . . 99
3.3.2 VaR and ES forecasts . . . 99
3.4 Empirical Study . . . 101
3.4.1 Data . . . 101
3.4.2 VaR forecast comparison . . . 108
3.4.3 Expected Shortfall forecast comparison . . . 110
3.4.4 Pre-financial-crisis and post-financial-crisis forecast performance . . 114
3.4.5 Loss function . . . 116
3.5 Conclusion . . . 117

4 Forecasting Tail Risk via a Partitioned Distribution 130

4.1 A Partitioned Distribution . . . 131

4.1.1 Standardized partitioned distribution of the Weibull and asymmetric Laplace . . . 133
4.1.2 VaR and tail conditional expectations for the standardized PWAL . 135
4.2 Model Specification and Estimation Methodology . . . 135
4.2.1 Volatility model . . . 135
4.3 Estimation Methodology and Simulation . . . 137
4.3.1 Bayesian estimation methodology . . . 138
4.3.1.1 Adaptive MCMC and AdMit . . . 138
4.3.2 Simulation study . . . 139
4.3.2.1 Simulation Study for Mixture of Gaussian and Mixture of Student's t . . . 140
4.3.2.2 Simulation Study for GJR-GARCH-SPWAL . . . 142
4.3.3 VaR and ES forecasts . . . 143
4.4 Empirical Study . . . 146
4.4.1 Data . . . 146
4.4.2 VaR forecast comparison . . . 147
4.4.3 Expected Shortfall forecast comparison . . . 152
4.4.4 Pre-financial-crisis and post-financial-crisis forecast performance . . 157
4.4.5 Loss function . . . 161
4.4.6 Further evaluations of model performance . . . 163
4.4.6.1 Log predictive likelihood . . . 163
4.4.6.2 Diagnostic tests . . . 164
4.5 Conclusion . . . 170

5 Discussion **174**

Bibliography **178**

Appendices **188**

Appendix 1 188

Appendix 2 189

Appendix 3 193

Appendix 4 195

List of Figures and Tables

List of Figures

1 Chapter 2: AL, Student t, and skewed Student t densities and log densities. 37

2 Chapter 2: Gaussian, Student t, and mixture of Gaussian densities and log densities. 44

3 Chapter 2: Plots of adaptive MCMC scheme in burn-in and sampling periods for β_1 in the GJR-GARCH-AL. 47

4 Chapter 2: Plots and histograms of the Australian market for both the initial learning period and the forecasting period. 50

5 Chapter 2: Plots and histograms of the US market for both the initial learning period and the forecasting period. 50

6 Chapter 2: Plots and histograms of the UK market for both the initial learning period and the forecasting period. 51

7 Chapter 2: Plots and histograms of the Hong Kong market for both the initial learning period and the forecasting period. 51

8 Chapter 2: Plots and histograms of the Australian dollar to US dollar for both the initial learning period and the forecasting period. 52

9 Chapter 2: Plots and histograms of the Euro to US dollar for both the initial learning period and the forecasting period. 52

10 Chapter 2: Dynamic shape and skew from GJR-GARCH-AL across six markets from May 1999 to April 2007. 55

11 Chapter 2: Australian stock market: 1-day-ahead 1% VaR forecasts from various models. 59

12 Chapter 2: Australian stock market: 1% 1-day-ahead ES forecasts from various models. 69

13 Chapter 2: Forecasts of constant and dynamic shape parameter from the GJR-AL model. 80

14 Chapter 2: Dynamic shape estimates from GJR-AL in the Australian market from May 1999 to May 2009, plus 95% credible intervals. 81

15 Chapter 2: 1% VaR and ES forecast over 10-day horizon from various in the Australian market. 82

16 Chapter 3: Ranges for skewness and kurtosis from TW distribution with $k_1 = k_2$. 91

17 Chapter 3: Some standardised two-sided Weibull densities. 92

18 Chapter 3: STW, AL, skewed Student t and Student t densities and log densities. 93

19 Chapter 3: MCMC plots in the burn-in period with six different starting positions for λ_1 in the TGARCH-TW model. 101

20 Chapter 3: Plots and histograms of the Australian market for both the initial learning period and the forecasting period. 102

21 Chapter 3: Plots and histograms of the US market for both the initial learning period and the forecasting period. 102

22 Chapter 3: Plots and histograms of the UK market for both the initial learning period and the forecasting period. 103

23 Chapter 3: Plots and histograms of the HK market for both the initial learning period and the forecasting period. 103

24 Chapter 3: Plots and histograms of AU/US for both the initial learning period and the forecasting period. 104

25 Chapter 3: Plots and histograms of Euro/US for both the initial learning period and the forecasting period. 104

26 Chapter 3: Plots and histograms of IBM for both the initial learning period and the forecasting period. 105

27 Chapter 3: Estimated $\frac{\lambda_1}{k_1}$ and p for TGARCH-TW and TGARCH-AL from January 2006 to January 2010. 106

28 Chapter 3: 1% VaR forecasts from GJR-n, GJR-t, GJR-skt, GJR-ALCP, and GJR-TW. 109

29 Chapter 3: 1% ES forecasts from GJR-n, GJR-t, GJR-skt, GJR-ALCP, and GJR-TW. 111

30 Chapter 3: The ratios of VRate/α at $\alpha = 0.01, 0.05$ for the pre-crisis and post-crisis periods. 115

31 Chapter 3: The ratios of ESRate/δ_α at $\alpha = 0.01, 0.05$ for the pre-crisis and post-crisis periods. 115

32 Chapter 3: Average loss functions of VaR and ES forecasts from various distributions across various volatility models. 117

33 Chapter 4: SPWAL, TW, and AL densitites and log densities. 136

34 Chapter 4: Estimates of the GJR-GARCH-AL parameters using AdMit and MiG proposals.: 400 replications 141

35 Chapter 4: Simulation study of the GJR-GARCH-PWAL model using AdMit. 144

36 Chapter 4: Estimates from 100 replications of the GJR-GARCH-PWAL model using AdMit. 145

37 Chapter 4: Ratios of $\hat{\alpha}/\alpha$ at $\alpha = 0.01, 0.05$. 150

38 Chapter 4: 1% VaR forecasts from AS CAViaR, GJR-t, GJR-AL, and GJR-PWAL. 151

39 Chapter 4: 1% ES forecasts from RiskMetrics, GJR-t, GJR-AL, and GJR-PWAL. 153

40 Chapter 4: ESRatios $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$. 155

41 Chapter 4: Pre-crisis and post-crisis VRate of $\hat{\alpha}/\alpha$ at $\alpha = 0.01, 0.05$. 158

42 Chapter4: Pre-crisis and post-crisis: ESRates of $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$. . . 158

43 Chapter 4: Loss function for VaR and ES forecasts. 162

44 Chapter 4: Log predictive likelihood for six models across five markets. . . 166

45 Chapter 4: Q-Q plots of residuals from the RiskMetrics and the GJR-GARCH-n model fitting of the return series of the Australian market. . . . 171

46 Chapter 4: Q-Q plots of residuals from the GJR-GARCH-t and the GJR-GARCH-AL model fitting of the return series of the Australian market. . . 171

47 Chapter 4: Q-Q plots of residuals from the GJR-GARCH-TW and the GJR-GARCH-PWAL model fitting of the return series of the Australian market. 172

List of Tables

1 Chapter 2: Summary statistics for parameter estimates from 400 simulated data sets from the GJR-GARCH-AL model. 48

2 Chapter 2: Precision and coverage for parameter estimates from 400 simulated data sets from the GJR-GARCH-AL model. 49

3 Chapter 2: Summary statistics of six return series. 53

4 Chapter 2: GJR-GARCH-AL parameter estimates and standard errors (bracketed). 53

5 Chapter 2: Ratios of $\hat{\alpha}/\alpha$ for 1-day-ahead VaR at $\alpha = 0.01, 0.05$. 61

6 Chapter 2: Counts of for 1-day-ahead VaR model rejections, at the 5% significance level, at risk level $\alpha = 0.01, 0.05$. 62

7 Chapter 2: Before the financial crisis: ratios of $\hat{\alpha}/\alpha$ for 1-day-ahead VaR at $\alpha = 0.01, 0.05$. 64

8 Chapter 2: Post-Financial Crisis: ratios of $\hat{\alpha}/\alpha$ for 1-day-ahead VaR at $\alpha = 0.01, 0.05$. 65

9 Chapter 2: ES quantile level function for the Gaussian and AL distributions. 66

10 Chapter 2: Nominal levels δ_α for 1-day-ahead ES for the Gaussian and AL distributions. 67

11 Chapter 2: Estimated nominal ES levels δ_α for 1-day-ahead ES at $\alpha = 0.01, 0.05$. 67

12 Chapter 2: ESRatios $\hat{\delta}_\alpha/\delta_\alpha$ for 1-day-ahead ES at $\alpha = 0.01, 0.05$. 67

13 Chapter 2: Counts of 1-day-ahead ES model rejections at $\alpha = 0.01, 0.05$. 70

14 Chapter 2: Before the financial crisis: $\hat{\delta}_\alpha/\delta_\alpha$ for 1-day-ahead ES at $\alpha = 0.01, 0.05$. 71

15 Chapter 2: After the financial crisis: $\hat{\delta}_\alpha/\delta_\alpha$ for 1-day-ahead ES at $\alpha = 0.01, 0.05$. 71

16 Chapter 2: Average daily market risk charge for 1-day-ahead VaR at $\alpha = 0.01$. . 72

17 Chapter 2: AD mean of violating 1-day-ahead returns at $\alpha = 0.01, 0.05$. 74

18 Chapter 2: AD maximum of violating for 1-day-ahead returns at $\alpha = 0.01, 0.05$. 75

19 Chapter 2: AD mean for 1-day-ahead ES at $\alpha = 0.01, 0.05$. 76

20 Chapter 2: Maximum absolute deviation for 1-day-ahead ES at $\alpha = 0.01, 0.05$. . 76

21 Chapter 2: Loss function for 1-day-ahead VaR at $\alpha = 0.01, 0.05$. 77

22 Chapter 2: Loss function for 1-day-ahead ES at $\alpha = 0.01, 0.05$. 78

23 Chapter 2: Ratios of $\hat{\alpha}/\alpha$ for 10-day-ahead VaR at $\alpha = 0.01, 0.05$. 84

24 Chapter 2: ESRates for 10-day-ahead ES at $\alpha = 0.05$. 85

25 Summary statistics for parameter estimates from 100 simulated data sets from the TGARCH-TW model. 100

26 Chapter 3: Summary statistics of seven return series. 107

27 Chapter 3: TGARCH-TW parameter estimates and standard errors (bracketed). 107

28 Chapter 3: Ratios of $\hat{\alpha}/\alpha$ at $\alpha = 0.01, 0.05$. 119

29 Chapter 3: Counts of model rejections for VaR models at $\alpha = 0.01, 0.05$. 120

30 Chapter 3: Estimated nominal ES levels δ_α at $\alpha = 0.01, 0.05$. 121

31 Chapter 3: ESRatios $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$. 122

32 Chapter 3: Counts of model rejections for ES models at $\alpha = 0.01, 0.05$. 123

33 Chapter 3: Before the financial crisis: ratios of $\hat{\alpha}/\alpha$ at $\alpha = 0.01, 0.05$. 124

34 Chapter 3: Before the financial crisis: $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$. 125

35 Chapter 3: Post-financial crisis: ratios of $\hat{\alpha}/\alpha$ at $\alpha = 0.01, 0.05$. 126

36 Chapter 3: Post-Financial Crisis: Ratios of $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$. 127

37 Chapter 3: Loss function for VaR at $\alpha = 0.01, 0.05$. 128

38 Chapter 3: Loss function for ES at $\alpha = 0.01, 0.05$. 129

39 Chapter 4: Summary statistics for parameter estimates from 400 simulated data sets from the GJR-GARCH-AL model. 142

40 Chapter 4: Summary statistics for parameter estimates from 100 simulated data sets from the GJR-GARCH-PWAL model. 143

41 Chapter 4: Summary statistics of five return series. 147

42 Chapter 4: GJR-GARCH-SPWAL parameter estimates and standard errors (bracketed). 148

43 Chapter 4: Ratios of $\hat{\alpha}/\alpha$ at $\alpha = 0.01, 0.05$. 149

44 Chapter 4: Counts of model rejections, at the 5% significance level, at risk level $\alpha = 0.01, 0.05$. 152

45 Chapter 4: Estimated nominal ES levels δ_α at $\alpha = 0.01, 0.05$. 154

46 Chapter 4: ESRatios $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$. 156

47 Chapter 4: Counts of ES model rejections at $\alpha = 0.01, 0.05$ 156

48 Chapter 4: Before the financial crisis: ratios of VRate/α at $\alpha = 0.01, 0.05$. . . . 159

49 Chapter 4: Before the financial crisis: $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$. 159

50 Chapter 4: Post-financial crisis: ratios of $\hat{\alpha}/\alpha$ at $\alpha = 0.01, 0.05$. 160

51 Chapter 4: Post-financial crisis: ratios of $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$. 160

52 Chapter 4: Loss function for VaR at $\alpha = 0.01, 0.05$. 161

53 Chapter 4: Loss function for ES at $\alpha = 0.01, 0.05$. 163

54 Chapter 4: Log predictive likelihood 164

55 Chapter 4: Shapiro-Wilk test. 167

56 Chapter 4: Lilliefors test. 167

57 Chapter 4: Chi-square goodness-of-fit test. 168

58 Chapter 4: Jacque-Bera test. 168

59 Chapter 4: Ljung-Box test on residuals: a lag of 20. 169

60 Chapter 4: Ljung-Box test on squared residuals: 20 lags. 169

61 Chapter 4: ARCH test: a lag of 10. 170

1 Introduction

The recent Global Financial Crisis (GFC) has once again called into question Financial Risk Management (FRM) methods and practice, as noted in Chen, Gerlach, Lin and Lee (2011). Many regulatory changes were introduced to FRM after the major stock-market crash ("Black Monday") of October 1987. In order to better control the risk and protect financial institutions against large unexpected losses, the group of G-10 countries agreed in 1988 to sponsor and subsequently form the original Basel Capital Accord (http://www.bis.org/publ/bcbsca.htm). However, in the 1990s, the occurrence of further financial crises (e.g. Orange County, US, lost $1.6 billion in 1994; in 1995 Barings Bank, UK, lost $1.4 billion; the Long-term Capital Management disaster in 1998) spurred regulators and financial institutions to establish a new benchmark for assessing market risks.

While the stock prices (and financial returns in general) cannot be predicted in the short term or long term, it is possible to forecast risk levels given certain scenarios and investment stances (that is, amount of leverage, long versus short the market, investments in derivatives and options, and so forth). With knowledge of a correctly forecasted risk level (given a set of circumstances), financial entities can manage their capital in a manner that enables them to be as prepared for specific levels of risk as possible. Therefore, loss can be reduced during volatile market movements and risk can be hedged to ensure that these institutions remain profitable.

The fundamental task of this thesis is to develop flexible parametric models for the most well-known and popular modern risk measurement methods, the Value-at-Risk (VaR) and the conditional VaR. The study aims to:

1. Estimate and forecast the potential skewness and dynamics in higher moments for conditional return distributions;

2. Accurately forecast the VaR and conditional VaR.

3. Examine the impacts of asymmetry in the volatility and that in the shape of the conditional return distributions on the VaR and conditional VaR forecasting.

4. Derive an easily applicable backtesting method for conditional VaR or expected shortfall.

5. Improve the efficiency and accuracy of Bayesian computational schemes for parameter estimation and forecasts.

The major contributions of this work are that it has:

1. Proposed a new model with a conditional asymmetric Laplace return distribution, which repeatedly provides conservative VaR and conditional VaR forecasts.

2. Derived the first four moments, the probability density function (pdf), the cumulative density function (cdf), inverse cdf, as well as VaR and conditional VaR functions for an asymmetric Laplace distribution.

3. Derived the function to locate the quantile that the conditional VaR lies at for an asymmetric Laplace distribution, given confidence level α.

4. Developed a more general and flexible distribution, a two-sided Weibull for a conditional return distribution, which accurately forecasts the financial tail risks.

5. Derived the first four moments, the pdf, cdf, inverse cdf, VaR and conditional VaR functions, and the ES quantile function for a two-sided Weibull ditribution.

6. Derived the ES function and ES quantile level function for the skewed Student t of Hansen (1994).

7. Addressed the 'Capital Charge Puzzle' by accurately forecasting risk levels via a two-sided Weibull distribution.

8. Proposed a partitioned distribution to address the bimodal problem of the two-sided Weibull ditribution, with better model-fitting performance.

9. Designed an easily comprehensible and applicable backtesting method for ES based on the actual quantile that the conditional VaR falls at.

10. Improved the accuracy of parameter estimation and speed of convergence in the Monte Carlo Markov chain, by employing a mixture of Gaussian proposals.

11. Further improved the accuracy and efficiency of Bayesian inference via a mixture of Student t proposals.

12. Found that the tails of the conditional return distribution play a more important role in VaR and ES forecasting than do the GARCH-type volatility specifications.

13. Derived necessary and sufficient conditions for second-order stationarity of the GJR-GARCH-AL and GJR-GARCH-TW models.

1.1 Financial Risk Measurement Methods

VaR was pioneered in 1993, as a part of the "Weatherstone 4:15pm" daily risk assessment report, in the RiskMetrics model at J.P. Morgan. In 1996, the Basel Committee on Bank Supervision advised commercial banks to determine a minimum regulatory capital, possibly by using an appropriate internal model that calculated VaR thresholds. VaR represents the market risk as one number: the maximum loss expected on an investment over a given time period at specific levels of confidence. Basel II recommends a backtesting procedure for evaluating and comparing VaR models based on the number of observed violations; that is, when actual losses exceed the VaR, in a hold-out sample period. As an important and popular method, the VaR approach is frequently investigated; see Duffie and Pan (1997), Dowd (1998), Jorion (2000), Dempster (2002), Allen (2003), and Holton (2003), among many others.

Although it is widely used by financial institutions, the practicality of using VaR was questioned at least by 1999, when the Bank of International Settlements (BIS) Committee on the Global Financial System pointed out that extreme market movement events "were in the 'tail' of [return] distributions, and that VaR models were [thus] useless for measuring and monitoring market risk." While this statement stood as an 'extreme' view, it is true that VaR does not measure the magnitude of loss for violations, illustrated well by Yamai and Yoshiba (2005). Further, Artzner, Delbaen, Eber, and Heath (1999) found that

VaR does not satisfy the sub-additivity property, that is, the sum of the risks of each asset in a portfolio is no greater than the total risk of the portfolio. Therefore VaR is not a 'coherent' risk measure. Consequently, the use of VaR can (sometimes) lead to portfolio concentration rather than diversification. Artzner et al. (1997, 1999) proposed an alternative, coherent measure, called Expected Shortfall (ES) (or conditional VaR, tail VaR, average value at risk, AVaR, or tail conditional expectation, TCE), which gives the expected loss (magnitude) conditional on exceeding a VaR threshold. However, VaR is recommended in Basel II, while ES is not (yet). Hence, both risk measures are considered here.

Basel II recommends a backtesting procedure for evaluating and comparing VaR models based on the number of observed violations; that is, when actual losses exceed the VaR, for a hold-out sample period of at least one year. Under-estimation of VaR (and ES) levels can result in institutions setting aside insufficient regulatory capital and thus suffering losses leading to insolvency during extreme market movements. Ewerhart (2002) argued that prudent financial institutions tend to hold unnecessary, excessive regulatory capital to ensure their reputation and quality, while Bakshi and Panayotov (2007) called this the "Capital Charge Puzzle". Intuitively, overstated VaR will lead financial institutions to allocate excessive amounts of capital, which may be attractive in the market after the most recent global financial crisis. However, as the goals of financial institutions are to meet the regulatory and capital requirements and to maximize profits and attract investors, such capital over-allocation represents an investment opportunity cost. Thus, although regulators may prefer smaller violation numbers in case of excessive losses, investors favour models that accurately predict risk, instead of overpredicting or underpredicting it. The goal of this study is to find a model that accurately predicts risk for the period before the recent global financial crisis as well as during and after the crisis.

1.2 Models for Value-at-Risk and Conditional Value-at-Risk

The recent global financial crisis showed that international financial markets are subject at times to quickly changing volatility and risk levels. Therefore, the crisis called into question, once again, risk measurement and risk management practices. With the in-

auguration of Basel III in 2011, fundamental questions are being raised and examined concerning how to measure risk, or even if, its level can be forecast accurately. There are three categories of methods for VaR and ES estimation according to McNeil and Frey (2000, p. 272):

1. Non-parametric historical simulation,
2. Parametric methods with fully specified volatility processes and conditional distributions, and
3. Semi-parametric methods based on the Extreme Value Theory (EVT) or quantile regression.

This thesis focuses on the parametric methods of risk forecasting.

In the academic literature, much interest has focused on conditional asset return distributions (which could help forecast financial risk levels accurately if well specified), with particular attention on two aspects: (i) the time-varying nature of the distribution, for example, volatility; and (ii) the shape and form of the standardised conditional distribution itself, for example, Gaussian; Tsay (2010, Chapter 7) provides an extensive list of models and methods on VaR.

1.2.1 Volatility models

Parametric VaR and ES measures rely heavily on volatility estimation, since this aspect is important for financial returns (well-known since at least the publication by Engle, 1982). Substantial literature now exists on the time-varying, persistent, and nonlinear properties of conditional return volatility. One stream of volatility models are the GARCH specifications. Five highly popular and well-known volatility models are: the Autoregressive Conditionally Heteroscedastic (ARCH) model of Engle (1982); the Generalised ARCH (GARCH) of Bollerslev (1986); the EGARCH model of Nelson (1991) and the GJR-GARCH (GJR) model of Glosten et al. (1993), both of which capture the well-known asymmetric volatility effect of Black (1976); and finally the RiskMetrics (RM) IGARCH

model developed by J.P. Morgan (1996). McAleer (2005) discussed these models in detail. More recently, fully nonlinear GARCH models have been specified, including the Threshold (T)-GARCH of Zakoian (1994); the Double Threshold (DT)-ARCH of Li and Li (1996); the DT-GARCH of Brooks (2001), and the Smooth Transition (ST-)GARCH of Gonzalez-Rivera (1998) and Gerlach and Chen (2008). A more general Dynamic Asymmetry (DA-)GARCH model, allowing for the dynamics in the multi-threshold structure, is proposed by Caporin and McAleer (2006). Many other models have been proposed that are far too numerous to mention.

The other stream of volatility models are the stochastic volatility (SV) models, which is based on a continuous-time stochastic process. Substantial literature exists for the SV model: it was first introduced by Taylor (1986) as an alternative to GARCH volatility models; Ghysels et al. (2002) and Shephard (2005) among many other authors have well reviewed the SV models. The SV model with Student t errors is widely adopted to model the heavy tails in the return series. However, practical studies found this model still inadequate to capture the heavier than normal tails, consequently Eraker et al. (2003) and Nakajima and Omori (2008) suggested using a jump component in the SV model to account for the tail fatness of returns. For more details of SV-jump model comparisons, refer to Chernov et al. (2003), Raggi and Bordignon (2006).

More recently, realized volatility model, based on high frequency intraday data, becomes a quite popular observable proxy for the latent volatility. The RV models are subject to various realized volatility measures and assumption on the dynmaics of return mean and volatility. Intraday squared returns are an intuitive realized volatility measure, while Andersen, Bollerslev, Diebold and Labys (2001) proposed to use intraday data as a more accurate proxy for the latent volatility. This non-parametric volatility model can be used to model leverage effect, jump processes and mixed jump diffusions for the return series.

Furthermore, Taylor (1986) suggested the absolute returns is a long memory series, with slow decaying autocorrelation coefficient. However, GARCH model is short memory model based on squared returns. For more studies on the relation of long memory volatility and option pricing, refer to the studies of Bollerslev and Mikkelsen (1996, 1999), Taylor

(2000) and Ohanissian, Russel and Tsay (2003). Bollerslev, Diebold and Labys (2003) built a vector autoregressive model with long distributed lags on realized volatility (RV) model for exchange rates. Fulvio Corsi (2004) developed a Heterogeneous Autoregressive model of the Realized Volatility (HAR-RV) to account for the long memory, tail fatness and self-similarity of volatility.

On the other hand, compared to the popular GARCH model, SV model is difficult to estimate, while the RV models have no specification on the shapes of the conditional return distribution. As the latter is one of the major focus of the thesis, this paper mainly consider the GARCH volatility models. Particularly, IGARCH is able to capture the long memory feature of return volatility, GJR-GARCH is able to model the leverage effect, non-linear GARCH models are able to examine the regime-switching process. The SV, RV and long memory models with focus on big shocks in the volatility will be of interest for future research.

Chapter 2 focuses on three of the most important and often-used models: GARCH, GJR, and RM IGARCH. In Chapter 3, to account for the nonlinearity and asymmetry in the volatility process, four specific GARCH models of the many enumerated above are reviewed: GARCH, GJR-GARCH, T-GARCH, and ST-GARCH. In Chapter 4, since the GJR-GARCH model did as well as, if not better than the other volatility models in Chapter 3, and with focus on the return distributions, the classic GJR-GARCH model again is adopted to specify the volatility process.

1.2.2 The return distributions

Another important aspect for VaR and ES estimation is the shape of the conditional return distribution. There is considerable empirical evidence that daily asset returns are fat-tailed (or leptokurtic) and also mildly negatively skewed, both unconditionally and conditionally, and thus cannot be adequately characterized solely by mean and variance (see for example Poon and Granger, 2003, among many others). Furthermore, during extreme market movement, the commonly-used GARCH model with Gaussian errors, is usually not able to fully capture the fat-tails (Thavaneswaran, Peiris, and Singh, 2008), influencing the proposal of the GARCH model with Student t error model by Bollerslev

(1987). In two papers, Harvey and Siddique (1999, 2000) argued that skewness in returns may be critical in investment decisions; for example, taking all else as constant, risk-averse investors should prefer portfolios that are right-skewed (that is, fewer but larger payoffs) to those that are left-skewed (that is, fewer but larger losses). Sadly, the recent global financial crisis highlights the danger of a left-skewed asset portfolio: one large loss can wipe out all previous gains. Ait-Sahalia and Brandt (2001) pointed out the importance of including skewness in financial decision making. For other reviews on skewness of financial returns, see Kraus and Litzenberger (1976), Friend and Westerfield (1980), Lim (1989), Richardson and Smith (1993), and Chen (2001). Ait-Sahalia and Brandt (2001) stated the difficulty of applying (forecast) skewness into financial practice.

There have been attempts to overcome this difficulty by adopting non-Gaussian distributions in finance. Mandelbrot (1963) and Fama (1965) pioneered the use of non-Gaussian distributions in finance, investigating the stable Paretian and power laws, while Mittnik and Ratchev (1989) also considered the Weibull, log-normal (separately for positive and negative returns), and Laplace distributions as unconditional return distributions. Fama (1965) and Barnea and Downes (1973) also considered mixtures of Gaussian ditribution in this context.

Subsequent to the first-generation ARCH and GARCH models, and because the kurtosis allowed by Gaussian errors does not often fully capture fat tails in returns, Bollerslev (1987) proposed the GARCH with conditional Student t error model; McCulloch (1985) used a simplified ARCH-type structure with a conditional stable distribution, updated to GARCH by Liu and Brorsen (1995); Nelson (1982) employed the generalised exponential as a conditional distribution in his E-GARCH; Vlaar and Palm (1993) used a mixture of Gaussian errors in a GARCH model; and Haas, Mittnik, and Paolella (2009) developed an asymmetric multivariate version.

For the dynamics in the skewness and kurtosis, Hansen (1994) developed a skewed Student t distribution, combining it with a GARCH model, also allowing both conditional skewness and kurtosis to change over time. More recently, Zhu and Galbraith (2009) extended the idea of a skewed Student t by using a generalized asymmetric Student t conditional distribution, with separate parameters in each tail; Li, Villani, and

Kohn (2009) modelled the conditional distribution with a smooth mixture of asymmetric Student t densities.

For other distributions used in financial modelling, Griffin and Steel (2006) and Jensen and Maheu (2010) employed Dirichlet process mixtures, while Aas and Haff (2006) used a generalised hyperbolic conditional return distribution.

Two particular distributions draw our attention for taking the skewness and heavy tails into account. One is the asymmetric Laplace (AL) distribution, first introduced in Hinkley and Revankar (1977) and applied to modelling financial data by Madan and Seneta (1990) and other researchers. The advantage of the AL distribution (ALD) is that, apart from the flexible excess kurtosis and skewness properties, dynamics in these higher moments can be allowed via a time-varying shape parameter, as proposed in Lu, Huang, and Gerlach (2010), who employed an exponential smoothing procedure and frequentist methods for that purpose. While Lu et al. (2010) considered an exponential smoothed conditional standard deviation model driven by absolute returns, we favour a more conventional GJR-GARCH model, driven by squared returns, but with AL errors; this model borrows existing specifications and combines them into a somewhat new Bayesian model. Unlike in Lu et al. (2010), under a Bayesian framework, simultaneous inference for all model parameters is possible; besides, valid finite sample inference can be obtained under parameter constraints required for the usual stationarity and positivity conditions of GARCH models; parameter uncertainty can also be accounted for in all forecast distributions. The empirical study shows that this model performs well for the time periods before, during, and after the recent global financial crisis, for VaR and ES forecasting, in comparison with a range of popular and often-used risk models/methods. Furthermore, the proposed model is the only one that remains conservative in its risk forecasts during and after the global financial crisis.

Although the regulators may favour this model because it repeatedly over-estimated risk levels and thus was a conservative risk model, especially for the period during the recent global financial crisis, investors may prefer a model that estimates risk levels more accurately. We achieve this by employing a natural and more flexible extension of the Laplace distribution, the Weibull, and subsequently developing a two-sided Weibull (TW)

distribution. After developing this distribution and its properties, we found that Sornette et al. (2000) had developed a symmetric, two-sided modified Weibull, subsequently used in Maleverge and Sornette (2004), as an unconditional distribution for asset returns, in combination with a Gaussian copula, in order to form efficient portfolios; an asymmetric modified Weibull was also briefly discussed.

We propose a slightly more flexible asymmetric TW distribution to use as a conditional return distribution in this study. The shape parameters of this TW distribution take value from a wider range, thus allowing more flexible shapes. After employing this more flexible model, we found that the Student t distribution is very accurate for VaR forecasting; however, the two-sided Weibull distribution is at least its equal for VaR forecasting, and works best for conditional VaR forecasting for the periods before, during, and after the recent global financial crisis.

The TW distribution accurately models the financial tail risks. However, from the empircal study we find the TW with appropriate tails can be potentially bimodal, while it is common knowledge that both the unconditional and conditional distribution of real financial return series are uni-modal. An intuitive and natural solution to this problem is to develop a partitioned distribution with Weibull tails and a uni-modal distribution in the centre. The distribution in the centre can be either symmetric or asymmetric. This partitioned distribution is able to capture the potential skewness and heavy tails in the conditional financial return distribution as well as better to specify the centre.

Several examples have been detailed in the literature of the combination of different distributions to form a new distribution. Behrens, Lopes, and Gamerman (2004) developed a truncated Gamma with a generalised Pareto distribution in the right tail for estimating the threshold of extreme events. Zhao, Scarrott, Oxley, and Reale (2010) considered a two-stage model, with a GARCH-type model in the first stage, and a normal distribution plus two generalised Pareto distributions in two tails modelling residuals in the second stage. So and Chan (2011) developed a mixture of distributions under a GARCH framework to model the tail asymmetry in financial time series. This mixture of distribution approach partitions the whole conditional distribution into three parts, with a normal distribution and two generalised Pareto distributions for the two tails.

The aforementioned model in So and Chan (2011) is constructed in a manner that is similar to the manner for the partitioned distribution model in this study; namely, both models use distributions from the stable distribution family to capture the heavy tails. Since the conditional distribution of real financial return series has a higher degree of kurtosis (that is, fat tails) than does a normal distribution, we tends to select a more general distribution with a wider range of kurtosis than normal, and with flexibility to be symmetric and asymmetric. Therefore, we choose the asymmetric Laplace distribution, which itself is an example of a partitioned distribution, with one break point and exponential distributions on each side. This is unlike the normal assumption that So and Chan (2011) made for the centre of the distribution.

Finally, very few studies have considered how to test ES models for adequacy or accuracy; see McNeil and Frey (2000). A method is presented in this thesis, based on a suggestion by Kerkhof and Melenberg (2004), using the trait that the ES occurs at specific, constant quantile levels, regardless of the other distribution parameters, for both the Gaussian and AL distribution, depending only on the VaR quantile level α. Thus, for these models the standard tests for correct quantile violation rates and independence can be applied to test for ES violations. Furthermore, we extend the conditional coverage test of Christoffersen (1998), which is a lag 1 test only, to apply as well to multiple lags.

1.3 Bayesian Inference and Forecasting

The Bayesian approach can easily incorporate uncertainty in all unknown factors and for any prior information, thus is a natural choice of inference and forecasting for VaR and ES models. The Bayesian method is widely used to deal with the challenges of capturing various unobservable dynamics and non-linearity, and generating accurate forecasts for the models in all economic activity, including finance. In this thesis, the likelihood functions of most models are of non-standard forms, thus it is difficult to estimate parameters via the Mamixum Likelihood (ML) method. Meanwhile, the moment conditions of the model parameters and the data are difficult to obtain, thus the parameter is unable to be estimated via the Generalized Method of Moments (GMM). Unlike the ML or GMM, Bayesian methods allow simultaneous inference on all parameters in a model, as well as

parameter uncertainty with inclusion of minimum parameter constraints. Therefore, this thesis employs Bayesian methods for parameter inference and forecasting.

1.3.1 Monte Carlo Markov chain for GARCH models

To execute Bayesian methods, the Markov chain Monte Carlo (MCMC) sampling scheme is adopted to sample from the joint posterior density of models and model parameters. A typical Markov chain is constructed by using Metropolis-Hastings (MH) steps with properly chosen proposal distributions. For a certain GARCH-type model, we divide the parameters into two groups, the GARCH parameters and error distribution parameters, and they are sampled in turn as a block in each MCMC iterate. Applications of the MCMC strategy for GARCH models can be found in the work of Muller and Pole (1995), Bauwens and Lubrano (1998), and Vrontos, Dellaportas, and Politis (2000), among many others. Particularly, the griddy Gibbs sampler (GGS) proposed by Bauwens and Lubrano (1998) is able to capture the shape of the posterior by using a comparatively smaller MCMC sample than the classic Gibbs sampler.

This study takes a Bayesian approach for estimation, extending the adaptive MCMC method of Chen and So (2006). We design a novel mixture of Gaussian proposal distributions for the required MH algorithm. Instead of the usual Gaussian random-walk proposal, this proposal will speed up mixing and convergence by reducing the risk of the chain getting stuck in local modes or mixing too slowly. A similar idea was considered by Hoogerheide, Kaashoek, and Van Dijk (2007) using mixtures of Student t proposal distributions targeted to the shape of the posterior. Giordani and Kohn (2008) improved the sampling efficiency by initiating a frequent updating process of the mixture of normal distributions at a very early stage of the chain. Sampling properties are studied and the mixture proposal approach has seen improved results of MCMC estimates. Estimation under the AL distribution in a quantile regression setting has been considered by Yu and Moyeed (2001), and Yu et al. (2003). Meanwhile, Geraci and Bottai (2007) developed MCMC methods in the semi-parametric quantile setting; this setting is different to the parametric case that this thesis primarily considers.

1.4 Methodology for Evaluating the Forecast Performance

1.4.1 Back-text based on violations

Basel II recommends backtesting VaR models to assess forecast accuracy. One of the most commonly used method is the violation rates (VRate), which allow an intuitive evaluation of VaR forecasts. VRate is the proportion of observations whose return exceeds the VaR level (called violations):

$$\text{VRate} = \frac{1}{M} \sum_{t=N+1}^{N+M} I(y_t < \text{VaR}_t) \tag{1}$$

where n is the in-sample size, M the forecast sample size. If a model correctly specifies the return quantiles, its true VRate should equal the nominal level α. The risk level α is usually chosen as 1% and 5%.

Three formal testing methods are also employed: **1.** The unconditional coverage (UC) test of Kupiec (1995): a likelihood ratio (LR) test that VRate = α; **2.** The conditional coverage (CC) test of Christoffersen (1998): a joint LR test for independence of violations and UC; **3.** The Dynamic Quantile (DQ) test of Engle and Manganelli (2004): an alternative joint test of independence and UC. The DQ test has been found to be more powerful than the lag 1 CC test; see Berkowitz et al. (in press).

As Basel II requires financial institutions reserve capital as insurance to extreme loss of assets, evaluating the size of the reserved capital, is also of interest. The capital set aside is called capital charge or market risk charge, it is set to be the larger of either current estimate of 1% VaR, or the average 1% VaR over the last 60 days multiplied by a penalty factor. The penalty factor is determined by the number of violating returns in the last 250 forecast trading days (approximately the number of trading days in a year). The expected number of losses exceeding 1% VaR in 250 forecast trading days should be 2.5. Banks are punished for using risky models, which have more than 4 violations per year. In this context, the model performance is categorized into three zones as follows:

$$\text{M}_t = \begin{cases} 3 & , \ N_{viol} \leq 4 \ \ \text{green} \\ 3 + 0.2\,(N_{viol} - 4) & , \ 5 \leq N_{viol} \leq 9 \ \ \text{yellow} \\ 4 & , \ 10 < N_{viol} \ \ \text{red} \end{cases}$$

where green indicates good models, yellow indicates less acceptable, and red indicates unacceptable models and requires immediate remedy to improve the risk management. Generally, models with lower MRC are more favoured for the purpose of risk management.

Furthermore, McAleer and da Veiga (2008) recommended the absolute deviation (AD) of returns exceeding VaR threshold for comparing the magnitude of loss. Smaller AD values indicate smaller size of loss for violations. However, it is noted that, models can have smaller AD mean and maximum, if they frequently under-estimate risk and end up with large number of violations. Chen, Gerlach, Lin and Lee (2011) pointed out that both the MRC and the AD methods should be combined with violation rates to evaluate the model performance. Thus, models with small MRC or AD is preferable only when they have appropriate violation rates at risk level α. Since the VRate and UC, CC and DQ tests are more effective, MRC and AD tests are only demonstrated in Chapter 2.

Loss functions are also applicable to assess general quantile forecasts. The applicable loss function is the criterion function, minimised in quantile regression estimation, e.g. as in Koenker and Bassett (1978), and can be written as:

$$LF = \sum_{t=N+1}^{N+M} (y_t - R_t)\,(\alpha - I_t). \tag{2}$$

where I_t is the indicator variable of a violation (i.e. $y_t < R_t$), R_t is the risk forecast, here we use VaR_t for each model/method. The best risk forecasts in terms of accuracy should minimise this loss function.

1.4.2 Diagnostic hypothesis tests

It is popular to evaluate the forecasting performance for models, adopting the idea of maximizing the predictive likelihood. The predictive likelihood measures the ability of forecasting (see e.g., Bjornstad, 1990, Davison, 1986 and Butler, 1986 for more details of this forecasting likelihood measure), thus is a suitable measure of fit for out-of-sample forecasts instead of in-sample model fitting. Particularly, Geweke and Keane (2007) and Li, Villani and Kohn (2009) employed predictive likelihood for comparing financial time series models.

Moreover, the normalized predictive residuals can be tested via a range of diagnostic

hypothesis tests. Smith (1985) first examined several diagnostic tests on nonnormal and non-standard models, whose residuals are not normal. Gerlach, Carter and Kohn (1999) computed the test statistics by using MCMC and importance sampling. For more details of diagnostic methods, that are widely applied to evaluate financial time series models, refer to Diebold, Gunther and Tay (1998) and Diebold, Hahn and Tay (1999).

In Chapter 4, both the forecasting performance and model-fitting performance are the focus of empirical studies. Thus only in Chapter 4, the predictive likelihood and diagnostic tests on the normalized predictive residuals are demonstrated.

1.5 Structure of the Thesis

In Chapter 2, the well-known GJR-GARCH form is used to model the volatility process, capturing the leverage effect; this is accomplished by combining an asymmetric Laplace form as the conditional distribution, accounting for potential skewness and heavy tails in the series. The dynamics at higher moments are specified by the time-varying shape parameter in this distribution. The ES formula is derived for the asymmetric Laplace distribution. A mixture of Gaussian proposal distributions is adopted in the MH algorithm for MCMC samples. The posterior distributions of the unknown parameters are estimated from these samples, hence the estimates of parameters are obtained by summarizing the estimated posterior distributions. The model is applied to forecast VaR and ES for real financial time series. Its performance is compared with popular models from non-parametric, semi-parametric, or parametric methods. Models are evaluated in periods before and after the recent global financial crisis, as this is of great interest to the financial community.

In Chapter 3, the scope of the study is expanded to include more general models and methods for VaR and ES. We examine a range of conditional distributions combined with four volatility specifications to forecast the tail risk in real financial return series. Asymmetry in both the volatility process and conditional distributions are considered. Specifically, a TW distribution, of which the AL distribution is a special case, is proposed to model the conditional return distribution. The first four moments, pdf, cdf, inverse cdf

and VaR, as well as the ES and ES quantile functions, are derived for the TW distribution. In the empirical study, one of the findings is that the tails of the conditional return distribution are much more important than the volatility model in VaR and ES forecasting. Furthermore, the TW outperforms other distributions, including the AL distribution, by providing the most accurate risk assessment, both before, during and after the recent global financial crisis.

Chapter 4 addresses the bimodal problem of the TW distribution. A partitioned distribution with Weibull tails and an AL distribution at the centre is proposed, alongside attempts to adequately model the conditional return distribution. The advantage of using Weibull distributions in tails for forecasting tail risk in financial series is proven in Chapter 3. A GJR-GARCH model specifies the volatility process. For parameter estimation, an adaptive MCMC sampling scheme is adopted, employing an MH algorithm with a mixture of Student t proposal distributions. The components of the mixture are determined through the adaptive mixture of the Student t (AdMit) approach of Hoogerheide et al. (2007). In empirical study, the forecasts from the GJR-GARCH-PWAL model, together with a range of competing models, are evaluated via backtesting methods. A number of formal goodness-of-fit tests are employed to evaluate the normalized residuals from the forecasts to evaluate the model fitting. In general, the proposed model ranks second or third in forecasting VaR and ES, closely following the GJR-GARCH-TW model; however, it has the smaller number of rejections than all the other competing models and ranked first in fitting the five series of returns in the empirical study.

Chapter 5 concludes.

1.6 Summary

In a world saturated with information, can quantitative methods make risk forecasting easier? Is there a model to adequately and accurately forecast the risk levels, not being too conservative, yet able to provide acceptable risk coverage? Which component is more important for a parametric VaR model, the correct volatility models, or the correct conditional return distributions?

This thesis addresses the above questions with novel models developed for adequate VaR and ES forecasting. As the goals of financial institutions are to meet the regulatory and capital requirements and in the meantime to maximize profits and attract investors, neither under-estimation of risk nor over-allocation of regulatory capital is desirable. The proposed models include a conservative and two less conservative models, to adapt to various purposes of risk measurements. The conservative model may be favoured for regulatory purposes. The other two models adequately and accurately predicting risk, rather than over-predicting or under-predicting risk; thus they may be favoured by investors due to their accuracy.

This thesis highlights some non-Gaussian distributions, namely asymmetric Laplace (AL), two-sided Weibull (TW), and partitioned distribution of the Weibull and AL (PWAL). These distributions capture the potential skewness and heavy tails in the financial series. The main characteristics of these distributions and their VaR and ES are derived. Various distributions are assessed under the GARCH frameworks, using both symmetric and asymmetric volatility specifications. As revealed by the empirical study, the correct conditional return distribution (or, more specifically, the tails of the distribution), is of prime importance for forecasting the tail risk in financial time series.

2 Bayesian Value-at-Risk and Expected Shortfall Forecasting via the Asymmetric Laplace Distribution

The goals of this chapter are to:

1. Propose a parametric model for estimating and forecasting financial risk to provide sufficent risk coverage.
2. Capture the potential skewness and heavy tails in the conditional return distribution.
3. Design a backtesting method for ES that can be easily understood and applied.
4. Improve the accuracy of Bayesian inference for parameters and speed up the covergence in the MCMC procedure.

To achieve the above goals, the proposed model in this chapter employs an AL conditional return distribution, capturing the potential skewness and heavy tails. By allowing for a time-varying shape parameter for the AL distribution, the dynamics in skewness, kurtosis, and higher moments can be characterised. A classic GJR-GARCH model is used to account for the leverage effect in the volatility process. For parameter estimation and forecast, an adaptive MCMC sampling scheme, employing an MH algorithm with a mixture of Gaussian proposal densities, is adopted to improve the efficiency and accuracy. An ES backtesting method is derived based on the quantile that the ES lies at. The model is applied to produce one-step-ahead and ten-step-ahead VaR and ES forecasts for six return series, covering a two-year period, which includes the global fiancial crisis in 2008. We extended the conditional coverage test of Christoffersen (1998) to allow for more than one lags. Formal and informal tests indicate that the GJR-GARCH-AL model outperformed a number of popular alternative models. The proposed model is the only model providing always conservative risk forecasts before the crisis, as well as during and after the crisis.

The contributions in this chapter are described as follows. First of all, the pdf, cdf, inverse cdf, VaR and ES functions, ES quantile function, and first four moments for the AL

distribution are derived. An ES backtesting method based on the actual ES quantile for level α is easy to understand and to apply for parametric models. Then, AL distribution with a dynamic shape parameter is compared with a constant shape parameter, and empirical evidence shows the assumption of dynamic skewness is necessary. In terms of estimation methodology, simluation studies sees improved estimation performance of MH algorithm with a mixture of Gaussian proposals. Last but not least, evalution of risk forecast performance in different crisis periods is novel as well.

A shortened version of the material in this chapter has been accepted for publication and will appear as Chen, Q., Gerlach, R. and Lu, Z. (2011), "Bayesian Value-at-Risk and expected shortfall forecasting via the asymmetric Laplace distribution," *Computational Statistics & Data Analysis*, in press.

2.1 Model Specification

This section reviews the statistical concepts of VaR and ES and addresses the problem of dynamic quantile and conditional expectation estimation via a GARCH-type volatility model with an AL error distribution.

2.1.1 VaR and ES

A general definition of VaR is defined in McNeil, Frey and Embrechts (2005). Given some confidence level $\alpha \in (0,1)$, the VaR of the portfolio at the confidence level α is given by the smallest number l such that the probability that the loss L exceeds l is no larger than $1-\alpha$. Formally,

$$\text{VaR}_\alpha = \inf\{l \in \Re : Pr(L > l) \leq 1-\alpha\} = \inf\{l \in \Re : F_L(l) \geq \alpha\}. \tag{3}$$

where F_L is the cdf of conditional return distribution. Thus, in this term, VaR is a quantile of loss function.

For a continuous conditional return distribution X, the ES of the portfolio given a loss L with $E(|L|) < \infty$ at confidence level α is defined as:

$$\text{ES}_\alpha = E[X|X \leq \text{VaR}_\alpha].$$

Therefore, the ES or the conditional VaR can be written as:

$$\begin{aligned} \mathrm{ES}_\alpha &= \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_u\left(F_L\right)du \\ &= \int_{-\infty}^{VaR_\alpha} xf\left(x|x<VaR_\alpha\right)dx \\ &= \int_{-\infty}^{VaR_\alpha} xf\left(x\right)/\mathrm{P}\left(x<VaR_\alpha\right)dx \\ &= \frac{1}{\alpha}\int_{-\infty}^{VaR_\alpha} xf\left(x\right)dx. \end{aligned} \tag{4}$$

There are three main categories of dynamic quantile and ES estimators: (i) non-parametric: where minimal assumptions are made, e.g. sample quantiles (often called historical simulation, HS); (ii) semi-parametric: e.g. quantile regression or the Conditional Autoregressive Value at Risk (CAViaR) model (Engle and Manganelli, 2004); and (iii) fully parametric: which specifies the dynamics and error distribution completely. In the empirical section in this chapter we consider examples from all three categories. However, this section focuses on a proposed parametric model: a GJR-GARCH model with AL distributed errors.

2.1.2 Asymmetric Laplace distribution

Guermat and Harris (2001) used the symmetric Laplace distribution with GARCH volatility to model short-horizon asset returns. Lu et al. (2010) extended this to allow skewness via the ALD. If a random variable X has an ALD, denoted $X \sim \mathrm{AL}(\gamma,\tau,p)$, the density function is

$$f(x|\gamma,\tau,p) = \frac{b_p}{\tau}\exp\left[-\frac{b_p}{\tau}|x-\gamma|\left(\frac{1}{p}I(x<\gamma)+\frac{1}{1-p}I(x>\gamma)\right)\right], \tag{5}$$

where $b_p = \sqrt{p^2+(1-p)^2}$ and γ,τ and p are location, scale and shape parameters, respectively. This parameterization is different to that in Yu and Zhang (2005, pg. 1867) for improved interpretability. Here, the shape parameter is defined such that $p = Pr(X < \gamma)$, and γ is the mode. Further, the scale factor b_p is included so that $\mathrm{Var}(X) = \tau^2$.

Volatility models employ i.i.d. error distributions standardized to have mean 0 and variance 1. The standard $\mathrm{AL}(0,1,p)$ distribution has density function

$$f(x|p) = b_p\exp\left[-b_p|x|\left(\frac{1}{p}I(x<0)+\frac{1}{1-p}I(x>0)\right)\right], \tag{6}$$

where the variance is 1, but the mean is $E(X) = \frac{1-2p}{b_p}$. Thus $Z = X - \frac{1-2p}{b_p}$ has an ALD, with shape parameter p, mean 0, and variance 1. This is the proposed error distribution for the GJR-GARCH model. Lu et al. (2010) used the AL in a time-varying parameter standard deviation GARCH model; we slightly simplify and modify their model.

Some other properties of the AL$(0, 1, p)$ are

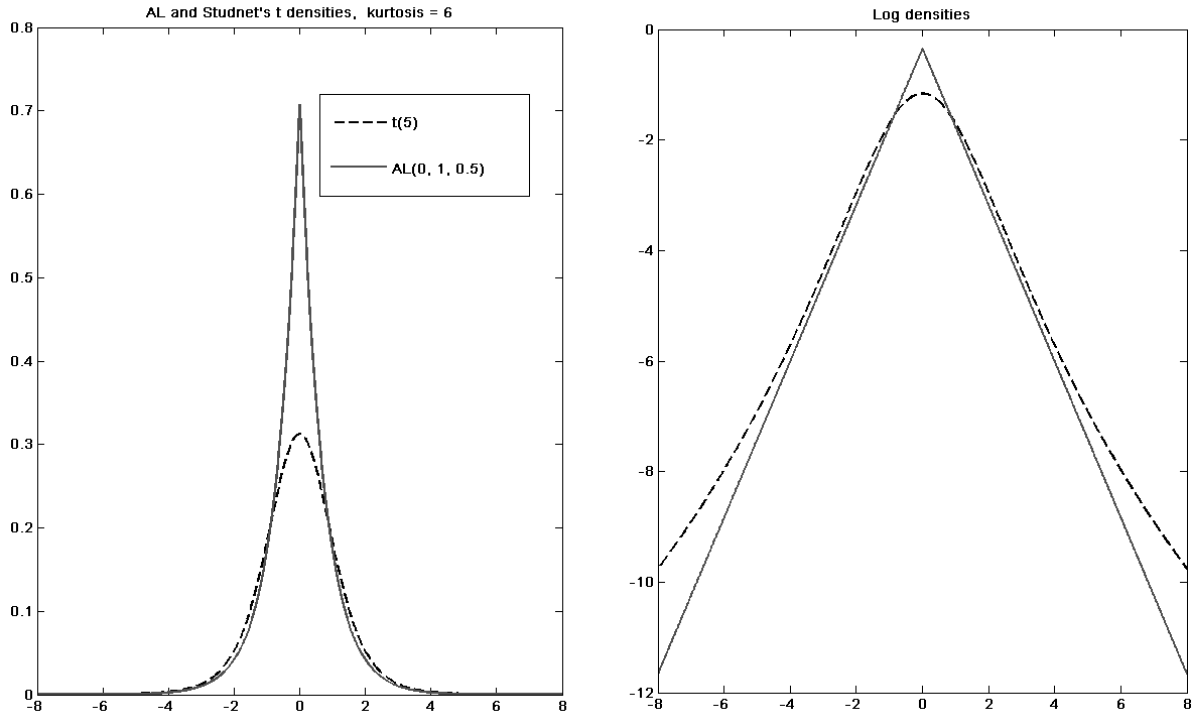
$$\begin{aligned} S(X) &= \frac{2\left((1-p)^3 - p^3\right)}{b_p^3}, \\ K(X) &= \frac{9(1-p)^4 + 6(1-p)^2p^2 + 9p^4}{b_p^4}, \end{aligned}$$

where S and K refer to the usual Pearson measures of skewness and kurtosis. Here $p = 0.5 \equiv S(X) = 0$ and thus symmetry while $p = 0.5 \equiv K(X) = 6$. Skewness is a monotonic function of p, ranging from $[-2, 2]$ as p ranges from 1 to 0; while kurtosis ranges from 6 to 9, with its maximum at 9 as $p \to 0$ or 1. Thus p tunes the shape of the ALD, controlling the skewness, kurtosis, and all higher moments. Specifically, if $p < 0.5$, the density is skewed to the right; while the opposite applies for $p > 0.5$.

Fat tails or excess kurtosis in conditional financial return distributions are often modelled by a Student t distribution. As a comparison, for $p = 0.5$, both the ALD and standardized Student t distribution with 5 degrees of freedom (DoF) have kurtosis of 6. However, the standardized t density function is lower than that for an AL$(0, 1, 0.5)$, from 2 to 6 standard deviations away from the mean. In this region, the symmetric Laplace density has fatter tails. Further, a skewed Student t was proposed by Azzalini and Capitanio (2003). For an AL$(0, 1, 0.55)$, the skewness and kurtosis are approximately -0.4 and 6.2, respectively. A skewed t ($\alpha = 0.5, DoF = 5.875$) with the same skewness and kurtosis has its left tail heavier than the AL density, between 2 and 6 standard deviations from the mean. In particular, $p = 0.55$ was a common value found in our empirical studies in Section 5. Figure 1 demonstrates the densities and log densities of AL, Student t, and skewed Student t. The skewed Student t distribution is not considered for conditional return distribution in this Chapter, but is included in Chapter 3 for a more extensive comparison. The tails of the ALD, fatter compared to the Student t distribution, may allow risk measures to be more accurate and effective during a financial crisis.

Parametric VaR and ES calculations require the cdf and inverse cdf, which can be

AL and Student t.



AL and skewed Student t.

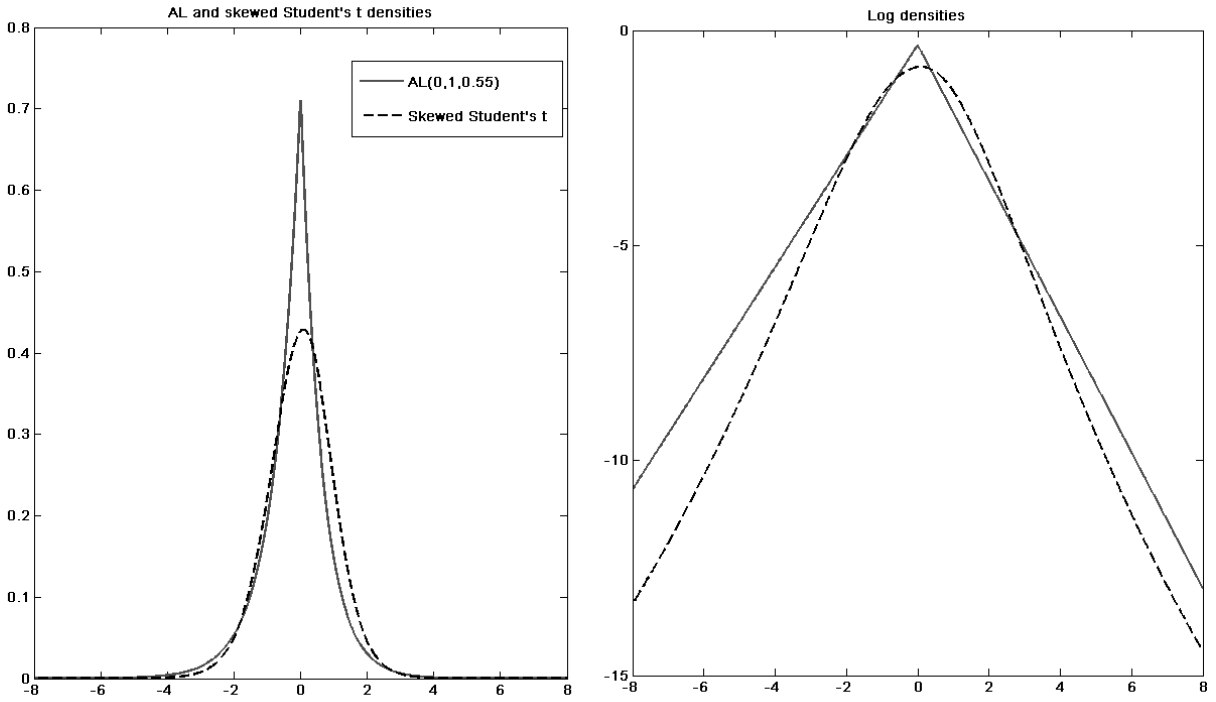


Figure 1: Chapter 2: AL, Student t, and skewed Student t densities and log densities.

obtained directly from (5). Since, in practice, p is usually close to 0.5, while risk management concerns only α levels well away from 0.5, i.e. extreme tails, and thus usually focuses only on the cases $\alpha \leq 0.05$. Hence only the case $\alpha < p$ is relevant here; results below are mostly shown only for this case. The $\mathrm{AL}(0,1,p)$ cdf is

$$\begin{aligned} F(x|p) &= \int_{-\infty}^{x} f(u)du \\ &= \begin{cases} \int_{-\infty}^{x} b_p \exp\left(\frac{b_p u}{p}\right) du \ ; \ x < 0 \\ \int_{0}^{x} b_p \exp\left(-\frac{b_p u}{1-p}\right) du + \int_{-\infty}^{0} b_p \exp\left(\frac{b_p u}{p}\right) du \ ; \ x \geq 0 \end{cases} \\ &= \begin{cases} p \exp\left(\frac{b_p x}{p}\right) \ ; \ x < 0 \\ p + (1-p)\left[1 - \exp\left(-\frac{b_p x}{1-p}\right)\right] \ ; \ x \geq 0 \end{cases} \\ &= \begin{cases} p \exp\left(\frac{b_p x}{p}\right) \ ; \ x < 0 \\ 1 - (1-p)\exp\left(\frac{-b_p x}{1-p}\right) \ ; \ x \geq 0. \end{cases} \end{aligned} \tag{7}$$

The inverse cdf of the $\mathrm{AL}(0,1,p)$ can be easily derived from (7), thus is:

$$F^{-1}(\alpha|p) = \begin{cases} \frac{p}{b_p} \log\left(\frac{\alpha}{p}\right) \ ; \ 0 \leq \alpha < p \\ -\frac{(1-p)}{b_p} \log\left(\frac{1-\alpha}{1-p}\right) \ ; \ p \leq \alpha < 1. \end{cases} \tag{8}$$

The expectation of an $\mathrm{AL}(0,1,p)$ for the long position (holding an asset), conditional on being below a quantile level α, is

$$\mathrm{ES}_\alpha = \int_{-\infty}^{VaR_\alpha} x f\left(x | x < VaR_\alpha\right) dx,$$

where $f\left(x|x < VaR_\alpha\right)$ is the conditional density function, which becomes

$$\begin{aligned} \mathrm{ES}_\alpha &= \int_{-\infty}^{VaR_\alpha} \frac{1}{\alpha} x b_p \exp\left(\frac{b_p x}{p}\right) dx \\ &= \frac{p}{\alpha}\left[x \exp\left(\frac{b_p x}{p}\right)|_{-\infty}^{VaR_\alpha} - \int_{-\infty}^{VaR_\alpha} \exp\left(\frac{b_p x}{p}\right) dx\right] \\ &= \frac{p}{\alpha}\left[VaR_\alpha \exp\left(\frac{b_p VaR_\alpha}{p}\right) - \frac{p}{b_p}\exp\left(\frac{b_p VaR_\alpha}{p}\right)\right]. \end{aligned}$$

By substituting the VaR_α with the corresponding formula in (8), the ES becomes:

$$\begin{aligned} \mathrm{ES}_\alpha &= \frac{p}{\alpha}\left(\frac{p}{b_p}\log\left(\frac{\alpha}{p}\right) - \frac{p}{b_p}\right)\left(\frac{b_p}{p}\frac{p}{b_p}\log\left(\frac{\alpha}{p}\right)\right) \\ &= \frac{p}{b_p}\left[\log\left(\frac{\alpha}{p}\right) - 1\right] \ ; \ 0 \leq \alpha < p \\ &= \left(1 - \frac{1}{\log\left(\frac{\alpha}{p}\right)}\right)\mathrm{VaR}_\alpha \ ; \ 0 \leq \alpha < p. \end{aligned} \tag{9}$$

Similarly, the ES for for $AL(0,1,p)$ for the short position (selling an asset) can be derived and is of the following form:

$$\begin{aligned} \mathrm{ES}_\alpha &= -\frac{1-\alpha}{\alpha}\frac{1-p}{b_p}\left[\log\left(\frac{1-\alpha}{1-p}\right)-1\right] \ ; \ p \le \alpha < 1 \qquad (10) \\ &= \frac{1-\alpha}{\alpha}\left(1-\frac{1}{\log\left(\frac{1-\alpha}{1-p}\right)}\right)\mathrm{VaR}_\alpha \ ; \ p \le \alpha < 1 \qquad (11) \end{aligned}$$

2.1.3 GJR-GARCH-ALD model with fixed shape

The GJR-GARCH model is a highly popular volatility model allowing for volatility persistence, heavy tails, and the leverage effect of Black (1976). The first model proposed is a GJR-GARCH model with standardised ALD error:

$$\begin{aligned} y_t &= (\epsilon_t - \mu_\epsilon)\sigma_t \qquad (12) \\ \epsilon_t &\overset{\text{i.i.d.}}{\sim} AL(0,1,p) \\ \sigma_t^2 &= \alpha_0 + \left[\alpha_1 + \alpha_2 I(y_{t-1}<0)\right] y_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \end{aligned}$$

Here, $\mu_\epsilon = \frac{1-2p}{b_p}$ and $\epsilon_t - \mu_\epsilon$ has mean 0 as standard; the indicator $I(\cdot)$ distinguishes negative from positive lagged returns. The leverage effect is measured by α_2. While the mean return here is fixed at 0, the methods here extend easily to any standard mean equation, such as autoregressions, exogenous regressions, nonlinear models, etc. However, since returns are unpredictable in mean, we chose a mean of 0 here. Sufficient conditions for positivity and stationarity (also necessary) are

$$\begin{gathered} \alpha_0 > 0 \ ; \ 0 \le \alpha_1 + \beta_1 + \alpha_2 c_p < 1 \\ \alpha_1, \beta_1, \alpha_1 + \alpha_2 \ge 0, \end{gathered}$$

where c_p is derived and given in Appendix 1. The usual value, $c_p = 0.5$, applies only when $p = 0.5$, i.e. for symmetry. Using (8) and (9) above, it is straightforward to show that

$$\begin{aligned} VaR_{\alpha,t+1} &= \sigma_{t+1}\frac{p}{b_p}\log\left(\frac{\alpha}{p}\right) - \mu_\epsilon \sigma_{t+1} \\ ES_{\alpha,t+1} &= \left(1-\frac{1}{\log\left(\frac{\alpha}{p}\right)}\right)\mathrm{VaR}_{\alpha,t+1} \ ; \ 0 \le \alpha < p. \qquad (13) \end{aligned}$$

This model is denoted as GJR-ALCP in Section 5.

2.1.4 Dynamic higher moments

The recent GFC may induce changes in the process that could be captured by allowing parameters to change dynamically. Since volatility is already heteroscedastic, we choose instead to focus on the higher moments, e.g. skewness and kurtosis. Hansen (1994) proposed a GARCH model with a skewed Student t error distribution, further allowing the skewness and degrees of freedom parameters to be dynamic, as later adopted by Jondeau and Rockinger (2006). Further, Zhu and Galbraith (2009) extended this idea, using a generalized asymmetric Student t distribution with separate parameters to control skewness and the thickness of each tail.

Instead, we follow Lu et al. (2010), who employed ALD errors and developed a dynamic smoothing equation for the shape parameter p. They showed that the maximum likelihood estimator of p is given by

$$\hat{p} = \frac{1}{1+\sqrt{\frac{u}{v}}},$$

where

$$u = \frac{1}{n}\sum_{x_t>0} |x_t| \; ; \quad v = \frac{1}{n}\sum_{x_t<0} |x_t|,$$

with $x_1, x_2, \ldots, x_n$ an i.i.d. sample from the AL$(0, 1, p)$. A dynamic, exponentially weighted moving average (EWMA) estimate for each of u and v was proposed as

$$\begin{aligned} u_t &= (1-\lambda)|x_{t-1}|I(x_{t-1}>0) + \lambda u_{t-1} \\ v_t &= (1-\lambda)|x_{t-1}|I(x_{t-1}<0) + \lambda v_{t-1} \end{aligned}$$

where $0 \leq \lambda \leq 1$ is the exponential smoothing parameter, and thus

$$p_t = \frac{1}{1+\sqrt{\frac{u_t}{v_t}}} \; ; \; b_{p[t]} = \sqrt{p_t^2 + (1-p_t)^2}$$

was proposed to allow a dynamic shape. This specification, which introduces three extra parameters, namely, the decaying parameter λ, and the initial values u_0, v_0 for u, v, to be estimated, allows all higher moments to change over time, in a manner directly influenced by the standardized data sample x, where from (12) $x_t = \epsilon_t = y_t/\sigma_t - \mu_{\epsilon_t}$. u_0, v_0 are very similar to the initial level in a standard EWMA model: their effect will die out over time.

2.1.5 GJR-GARCH AL model with dynamic shape

The full model, extended from (12) to allow for dynamic shape, is

$$\begin{aligned} y_t &= (\epsilon_t - \mu_{\epsilon_t})\sigma_t \\ \epsilon_t|\Omega_{t-1} &\sim AL(0,1,p_t) \\ \sigma_t^2 &= \alpha_0 + (\alpha_1 + \alpha_2 I(y_t<0))\, y_{t-1}^2 + \beta_1\sigma_{t-1}^2, \end{aligned} \tag{14}$$

where $\mu_{\epsilon_t} = (1-2p_t)/b_{p_t}$ and the previous definitions and restrictions apply. Using (8) and (9) again,

$$\begin{aligned} VaR_{\alpha,t+1} &= \sigma_{t+1}\frac{p_{t+1}}{b_{p[t+1]}}\log\left(\frac{\alpha}{p_{t+1}}\right) - \mu_{\epsilon_{t+1}}\sigma_{t+1} \\ ES_{\alpha,t+1} &= \left(1 - \frac{1}{\log\left(\frac{\alpha}{p_{t+1}}\right)}\right)\mathrm{VaR}_{\alpha,t+1}\ ;\ 0 \le \alpha < p_{t+1}, \end{aligned} \tag{15}$$

where again only the relevant $0 \le \alpha < p_{t+1}$ case is shown.

Let $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta_1, \alpha_2); \boldsymbol{\eta} = (u_0, v_0)$, and $y^{1,t-1} = \{y_1, \ldots, y_{t-1}\}$, the likelihood function from model (14) is

$$L(\boldsymbol{y}|\boldsymbol{\theta}, \lambda, \eta) = \prod_{t=2}^{n} f(y_t|y^{1,t-1}, \boldsymbol{\theta}, \lambda, \boldsymbol{\eta}) \tag{16}$$

where

$$f(y_t|y^{1,t-1}, \boldsymbol{\theta}, \lambda, \boldsymbol{\eta}) = \frac{b_{p[t]}}{\sigma_t}\exp\left[-b_{p[t]}|\epsilon_t|\left(\frac{1}{p_t}I(\epsilon_t<0) + \frac{1}{1-p_t}I(\epsilon_t>0)\right)\right].$$

2.2 Bayesian MCMC Methods

This section reviews and details the Bayesian methods and MCMC procedures of Chen and So (2006) for estimating parameters from Chen and So (2006), and specifies details for the proposed adaptive mixture distribution proposal for the MH methods.

2.2.1 Bayesian methods

For a standard Bayesian analysis, the likelihood function, specified in (16), and a prior distribution are required. We assume a priori independence in this context, and choose

an uninformative prior over relevant ranges for each parameter, as follows:
For the GJR-GARCH parameters $\boldsymbol{\theta}$:

$$\pi(\boldsymbol{\theta}) \propto I\left(\alpha_0 > 0, \alpha_1 + \beta_1 + c_p\alpha_2 < 1, \alpha_1, \beta_1, \alpha_1 + \alpha_2 \geq 0\right).$$

Then, for the exponential decay parameter λ:

$$\pi(\lambda) \propto I\left(0 \leq \lambda \leq 1\right).$$

Finally,

$$\pi(\eta) \propto I\left[0 < u_0, v_0 < \max\left(\frac{\boldsymbol{y}}{\boldsymbol{\sigma}}\right)\right].$$

Using (16) and the above priors, the joint posterior density kernel for $\boldsymbol{\theta}, \lambda, \eta|\boldsymbol{y}$ is simply the likelihood times the multiplication of the specified priors:

$$p(\boldsymbol{\theta}, \lambda, \eta|\boldsymbol{y}) \quad \propto \quad p(\boldsymbol{\theta}, \lambda, \eta)L(\boldsymbol{\theta}, \lambda, \eta; \boldsymbol{y}). \tag{17}$$

2.2.2 Adaptive MCMC sampling using MH methods

The posterior in (17) is of a non-standard form in the parameters; thus, an MCMC sample is drawn by employing MH algorithms (Metropolis et al., 1953; Hastings, 1970). We adopt an adaptive sampling scheme by combining a random-walk Metropolis (RWM) and independent kernel Metropolis-Hastings (IKMH) algorithms to speed up convergence and mixing in the chain, extending that in Chen and So (2006).

Specifically, the MH algorithms are embedded within a Gibbs chain. As a special case of MCMC, Gibbs sampler (introduced by Geman and Geman 1984), enables sampling from a conditional distribution rather than from marginal distribution by integrating the joint distribution of a multivariate distribution. Thus, this sampling scheme is simpler and easier to use. Particularly, if we want to generate J samples of $Z = (\boldsymbol{\theta}, \lambda, \eta)$ from the joint density $p\left(\boldsymbol{\theta}, \lambda, \eta\right)$. Denote the jth sample as $Z^{(j)} = \left(\boldsymbol{\theta}^{(j)}, \lambda^{(j)}, \eta^{(j)}\right)$, $j = \{1, \ldots, J\}$, the Gibbs sampling scheme is as follows:

1. Begin with some initial values $Z^{(0)}$ for each variable;

2. For the jth sample, sample $\boldsymbol{\theta}^{(j)}$ from the conditional distribution $p\left(\boldsymbol{\theta}^{(j)}|\lambda^{(j-1)}, \eta^{(j-1)}, \boldsymbol{y}\right)$;

3. Then sample $\lambda^{(j)}$ from the conditional distribution $p\left(\lambda^{(j)}|\boldsymbol{\theta}^{(j)}, \eta^{(j-1)}, \boldsymbol{y}\right)$;

4. Sample $\eta^{(j)}$ from the conditional distribution $p\left(\eta^{(j)}|\boldsymbol{\theta}^{(j)}, \lambda^{(j)}, \boldsymbol{y}\right)$.

By this means, each variable is sampled from the distribution of that variable conditioned on all the other variables and the data. The Gibbs sampler use the most recent values and update the variable with latest proposed value immediately after it has been sampled. The samples are approximation of the joint distribution of all variables. Thus, the estimated value of variables can be obtained by taking average over all the samples.

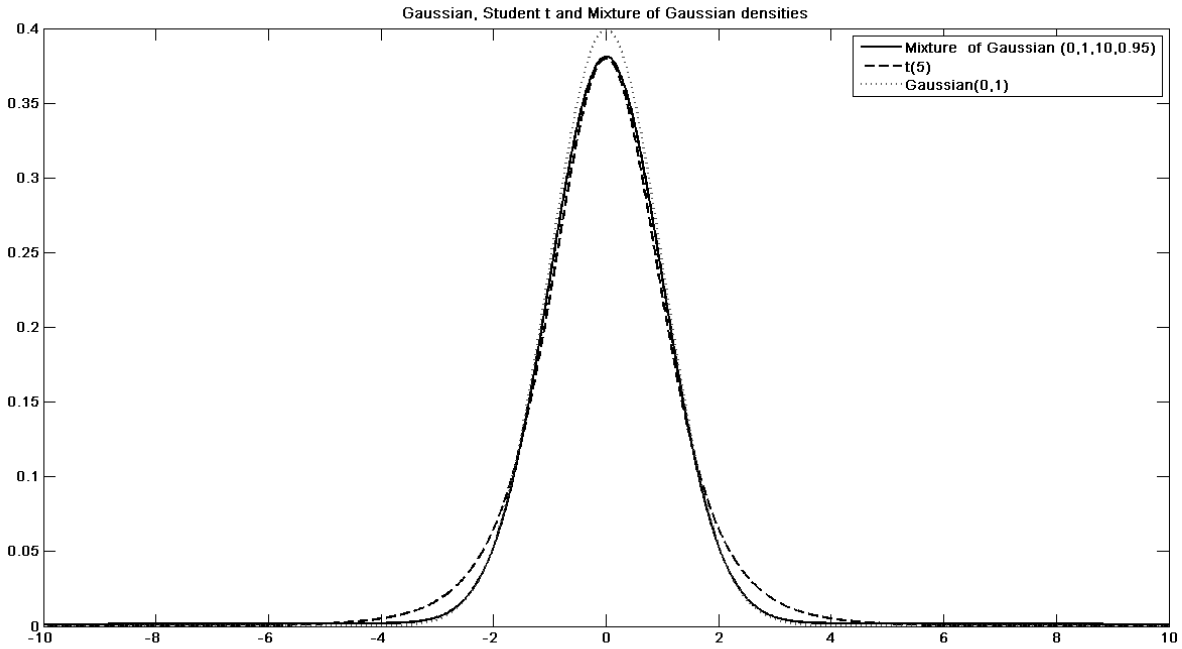
For each Gibbs sample, mixture of Gaussian proposal density is proposed for the parameter groupings $\boldsymbol{\theta}$ and λ in both the RWM and IKMH sampling schemes. In this study, a mixture of Gaussian random variable is defined as:

$$x \sim \rho N(\mathbf{0}, \text{diag}\{\boldsymbol{c_1}\}) + (1-\rho)N(\mathbf{0}, \omega\text{diag}\{\boldsymbol{c_1}\}),$$

and a good choice for ρ can be within $(0.8, 1)$, which allows for the occasional jump to propose values from the second component; the choice for ω should be large, for example 100, and c_1 should be chosen matching the scale of the parameters. Figure 2 shows the density and log density plots of a mixture of N(0,1) and N(0,10) with weight on the first component equalling 0.95, together with a N(0,1) and a Student t with 5 degrees of freedom; the latter two are a common choice for the RWM algorithm. The mixture of Gaussian distributions has much fatter tails than the Gaussian and Student t distribution, though the Student t has fatter tails until 5 standard deviations, thus can propose values in a wider range. Mixture of Gaussian distributions has flexible shapes, which can be controlled by the choice of ρ and number of components. The choice of $\rho = 0.95$ and a bivariate mixture here, though seems arbitrary, is acceptable, as long as it can generate appropriate proposed values (and occasionally distant enough from the mean) without the knowledge the exact distribution of parameters. This proposal will help to speed up mixing of the Markov chain, since it is designed to allow occasional large jumps, perhaps away from local modes, allowing a greater chance that the chain will cover the posterior distribution and not get 'stuck'. This is vitally important, since the IKMH part of the algorithm completely relies on such coverage, as below. The sampling scheme for $\boldsymbol{\theta}$ follows:

***Step*1.** RWM with mixture proposal during the burn-in period: $\boldsymbol{\theta}|\boldsymbol{y}, \lambda, \boldsymbol{\eta}$. At MCMC

Density plots.



Log densities.

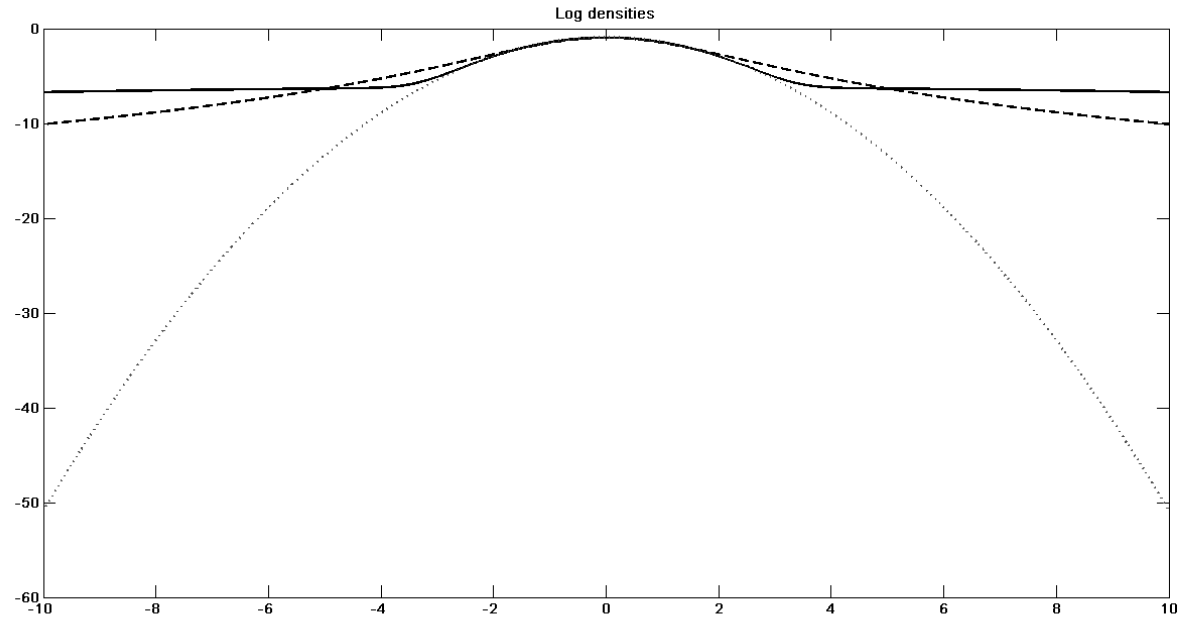


Figure 2: Chapter 2: Gaussian, Student t, and mixture of Gaussian densities and log densities.

iteration j,

$$\boldsymbol{\theta}^p = \boldsymbol{\theta}^{(j-1)} + \boldsymbol{\varepsilon} \ ; \ \ \boldsymbol{\varepsilon} \sim \rho N(\mathbf{0}, \text{diag}\{\boldsymbol{c_1}\}) + (1-\rho)N(\mathbf{0}, \omega \text{diag}\{\boldsymbol{c_1}\}),$$

where $\rho = 0.95$, $\omega = 100$, and $\boldsymbol{c_1}$ is a vector of positive numbers, tuned so that the observed acceptance rate lies $\in (0.15, 0.5)$, as recommended by Gilks et al (1996). Tuning is done every 200 iterations, by increasing each element in $\boldsymbol{c_1}$ when the acceptance rate in the last 200 iterations is higher than 0.5, or decreasing $\boldsymbol{c_1}$ when that acceptance rate is lower than 0.15. The proposed $\boldsymbol{\theta}^p$ is accepted as $\boldsymbol{\theta}^{(j)}$ with the usual RWM probability. Step 1 is iterated, at least until convergence is achieved, for $j = 1, \ldots, J$, and these iterates make up the burn-in MCMC sample.

The MCMC sampling period employs an IKMH algorithm. First, the sample mean and sample covariance matrix of the burn-in MCMC iterates for $\boldsymbol{\theta}$ are calculated and denoted as $\boldsymbol{\mu}$ and Σ.

Step2. Re-start the MCMC using an IK mixture of Gaussian proposal MH method for $\boldsymbol{\theta}|\boldsymbol{y}, \lambda, \boldsymbol{\eta}$. At MCMC iterate j :

$$\boldsymbol{\theta}^p = \boldsymbol{\mu} + \Sigma^{0.5}\boldsymbol{\varepsilon} \ ; \ \ \boldsymbol{\varepsilon} \sim \rho N(0, I) + (1-\rho)N(0, \omega I),$$

where I is the identity matrix and $\Sigma^{0.5}$ is the lower triangular Cholesky decomposition of $\Sigma = (\Sigma^{0.5})'\Sigma^{0.5}$. Then accept $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^p$ with the usual IKMH probability. This step is iterated for $j = J+1, \ldots, J+K$, where K is chosen to be sufficiently large for reliable inference.

The adaptive MCMC sampling scheme for each of $\boldsymbol{\eta}|\boldsymbol{y}, \boldsymbol{\theta}, \lambda$ and $\lambda|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\eta}$ is equivalent.

2.3 Simulation Study

A small-scale simulation study is carried out to examine the effectiveness of the proposed MCMC sampling procedure and compare its' sampling properties with the method of Chen and So (2006).

2.3.1 Simulation setup

Sample series of size $n = 2000$, the same size as the real data we consider, are simulated from the following model:

$$\begin{aligned}
u_t &= 0.002|\epsilon_{t-1}|I(\epsilon_{t-1} > 0) + 0.998u_{t-1}, \\
v_t &= 0.002|\epsilon_{t-1}|I(\epsilon_{t-1} < 0) + 0.998v_{t-1}, \\
p_t &= \frac{1}{1+\sqrt{\frac{u_t}{v_t}}} \; ; \; \epsilon_t \overset{\text{i.i.d.}}{\sim} AL(0,1,p_t), \\
\sigma_t^2 &= 0.05 + (0.05 + 0.05I(y_{t-1} < 0))\, y_{t-1}^2 + 0.85\sigma_{t-1}^2 \\
y_t &= (\epsilon_t - \mu_{\epsilon_t})\sigma_t,
\end{aligned}$$

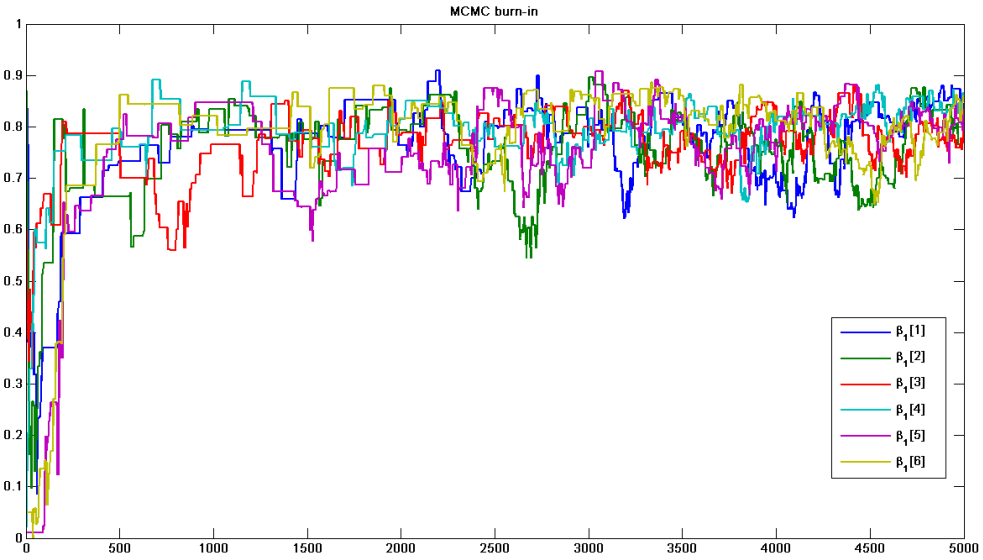
where true parameter values are chosen as similar to parameter estimates for this model from the real data considered. The initial values u_0, v_0 are chosen randomly in $(0, 1)$ for each data set, while the first 500 observations are discarded to lessen the effects of these initial settings (the 2000 observations simulated after the 500th are used).

The adaptive MCMC method in Section 2.2.2 is used to estimate the parameters for each of 400 data replications. For each replicate the MCMC sample has 15 000 iterations in total: the first 5000 are the burn-in period using Step 1, and are discarded; while Step 2 is used to generate the final sample of 10 000 iterates, used for estimation and inference.

The adaptive MCMC scheme for estimating the GJR-GARCH-AL model is realized in Matlab. I derived the code for the adaptive MCMC scheme and can be contacted if the code is of interest. The typical MCMC running time for $n = 2000$ with single data set for GJR-GARCh-AL model is around 1.5 minutes, the typical MH acceptance rates are between 5% and 17% for the burn-in period, and between 50% and 68% in the sampling period.

MCMC convergence is assessed via extensive examination of time series plots of MCMC iterates from multiple and varied starting positions for each parameter, from which the burn-in and sampling period sizes were chosen: in all simulated and real data examples we observed MCMC convergence well inside 5000 iterations. Figure 3 shows that the adaptive MCMC scheme performs well with fast convergence in both burn-in and sampling periods from varied starting positions for GJR-GARCH parameter $\boldsymbol{\theta}$.

MCMC plots in burn-in with six different starting points.



MCMC plots in sampling period with various starting points.

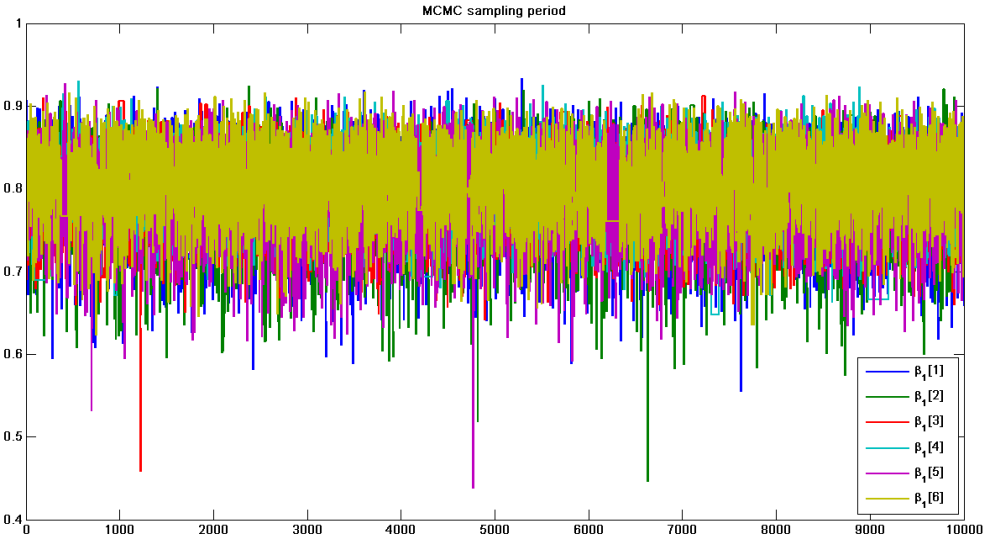


Figure 3: Chapter 2: Plots of adaptive MCMC scheme in burn-in and sampling periods for β_1 in the GJR-GARCH-AL.

2.3.2 Simulation results

For each replicate data set, the model was also estimated using the single-Gaussian-proposal adaptive method in Chen and So (2006) for the parameters in $\boldsymbol{\theta}$. Table 1 and Table 2 summarize the results obtained over the 400 data sets. The averages of the 400 posterior mean estimates (Mean) and median estimates (Median), the standard deviation of the 400 estimates, and the observed coverage rate for the 400 95% credible intervals are shown. Values for the best method are made bold in each case, in terms of the 'Mean' being closest to the true value, minimum standard error, and the 'Coverage' closest to the nominal 95%. Clearly, results favour the suggested mixture proposal MH method in this case, especially in terms of precision and coverage.

Table 1: Chapter 2: Summary statistics for parameter estimates from 400 simulated data sets from the GJR-GARCH-AL model.

		Mean			Median		
par.	True	Gaussian	$t(5)$	Mixture	Gaussian	$t(5)$	Mixture
λ	0.998	0.995	**0.996**	**0.996**	**0.997**	**0.997**	0.996
α_0	0.05	0.085	0.073	**0.072**	**0.061**	0.067	0.066
α_1	0.05	0.066	**0.064**	**0.064**	**0.058**	0.062	0.062
α_2	0.05	0.059	0.056	**0.055**	**0.052**	0.054	0.053
β_1	0.85	**0.804**	0.802	**0.804**	**0.823**	0.813	0.812

Note: The bold indicates the estimates closer to the true values.

The parameter estimates are similar among models with the above three different proposal distributions. The precision of estimates should be included as well; the coverage rates of estimates are recorded in Table 2.

2.4 Empirical Study

2.4.1 Data

The model is illustrated by applying it to daily series from four international stock-market indices: the S&P 500 (US); FTSE 100 (UK); AORD All Ordinaries Index (Australia); the HANG SENG Index (Hong Kong), as well as two exchange rates: the AU dollar to the US dollar and the Euro to the US dollar. The data are from Yahoo! Finance, covering ten

Table 2: Chapter 2: Precision and coverage for parameter estimates from 400 simulated data sets from the GJR-GARCH-AL model.

	Standard error			Coverage		
par.	Gaussian	$t(5)$	Mixture	Gaussian	$t(5)$	Mixture
λ	0.025	0.004	**0.002**	0.863	0.865	**0.878**
α_0	0.126	0.030	**0.025**	0.763	**0.845**	0.840
α_1	0.059	**0.026**	**0.026**	0.835	0.955	**0.953**
γ_1	0.099	0.038	**0.034**	0.743	0.830	**0.838**
β_1	0.063	0.056	**0.050**	0.820	0.930	**0.940**

Note: The bold indicates the estimates of better precision and higher coverage rate.

years, May, 1999–May, 2009. We consider the daily percentage log-return series, defined by $y_t = (ln(P_t) - ln(P_{t-1})) \times 100$, where P_t is the closing price index on day t.

The sample is divided into two periods: the first 2000 returns, from May 1999 to approximately April 2007, are an initial learning period. The remaining sample, approximately the 500 to 550 observations from May 2007 to May 2009 in each market, are used as a forecasting period. The variations from market to market in sample size are due to different trading day holidays, etc. Summary statistics for the six return series presented in Table 3 and Figures 4 to 9 show the separate plots and histograms from the learning and forecast samples. While not too much should be inferred from this, the forecast period was more volatile than the learning period in each market, likely due to the GFC. The parameter estimates for learning period and standard errors (in brackets) using the GJR-GARCH-AL model are given in Table 4.

The estimation results in each market are mostly as expected and well-documented in the literature: e.g. high volatility persistence ($\alpha_1 + \beta_1 + \alpha_2/2$ close to 1); the leverage effect implies that $\alpha_2 > 0$: this occurred in all four stock markets; it is well-known that exchange rates usually do not exhibit significant leverage: our results agree. At the start of the learning, sample p is estimated less than 0.5, since it is estimated that $u_0 < v_0$, so skewness is estimated as negative in all markets except the EUR/US rate. Figure 10 shows the estimated dynamic p, and the corresponding skewness estimates in each market, during the entire sample period. The two dates in the plots correspond to two large single-day drops in the Australian index value.

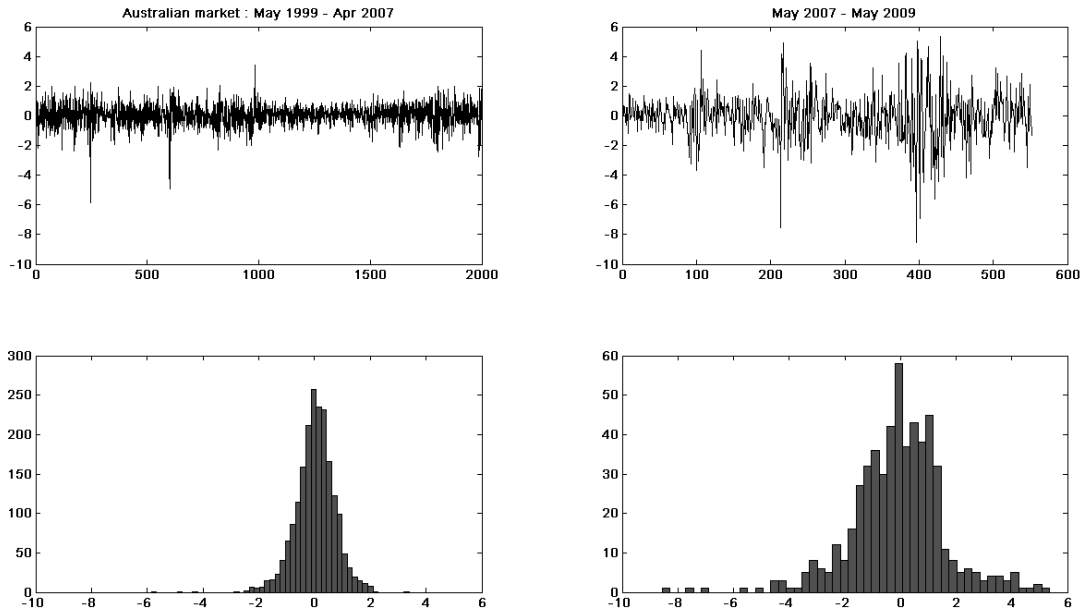


Figure 4: Chapter 2: Plots and histograms of the Australian market for both the initial learning period and the forecasting period.

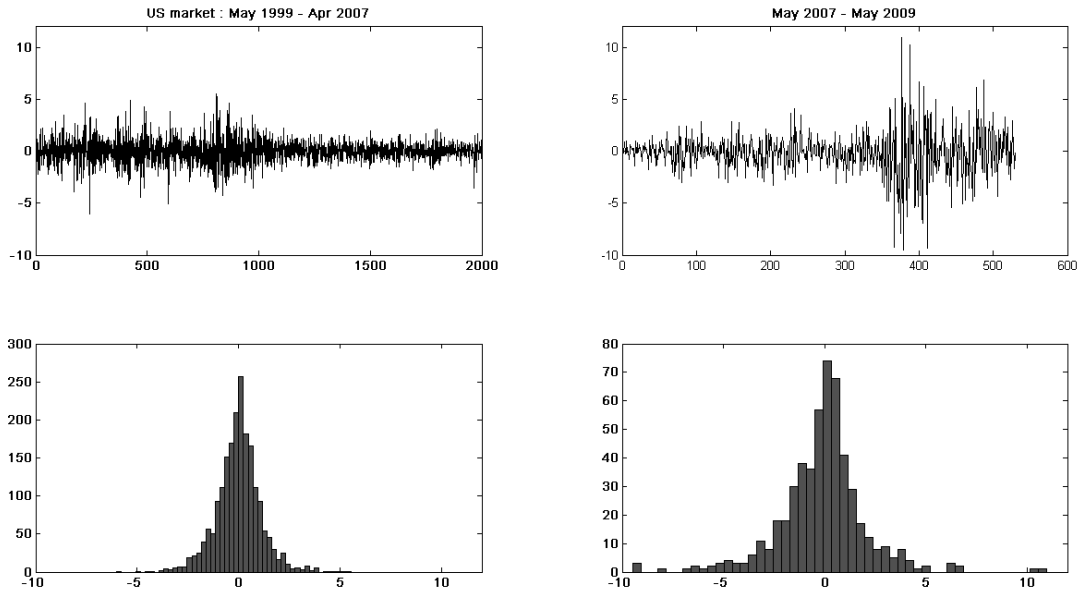


Figure 5: Chapter 2: Plots and histograms of the US market for both the initial learning period and the forecasting period.

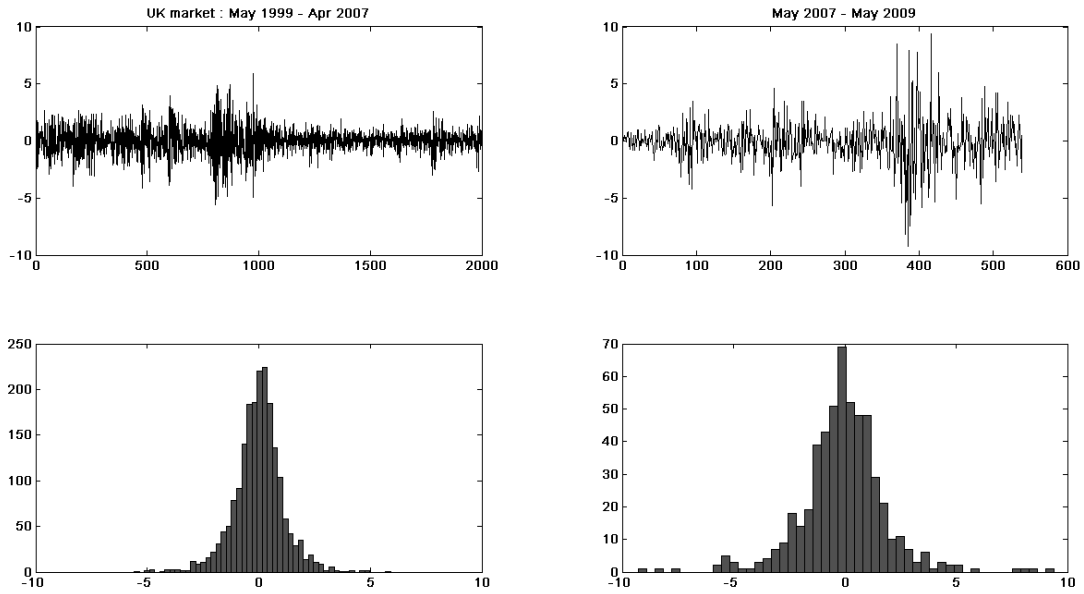


Figure 6: Chapter 2: Plots and histograms of the UK market for both the initial learning period and the forecasting period.

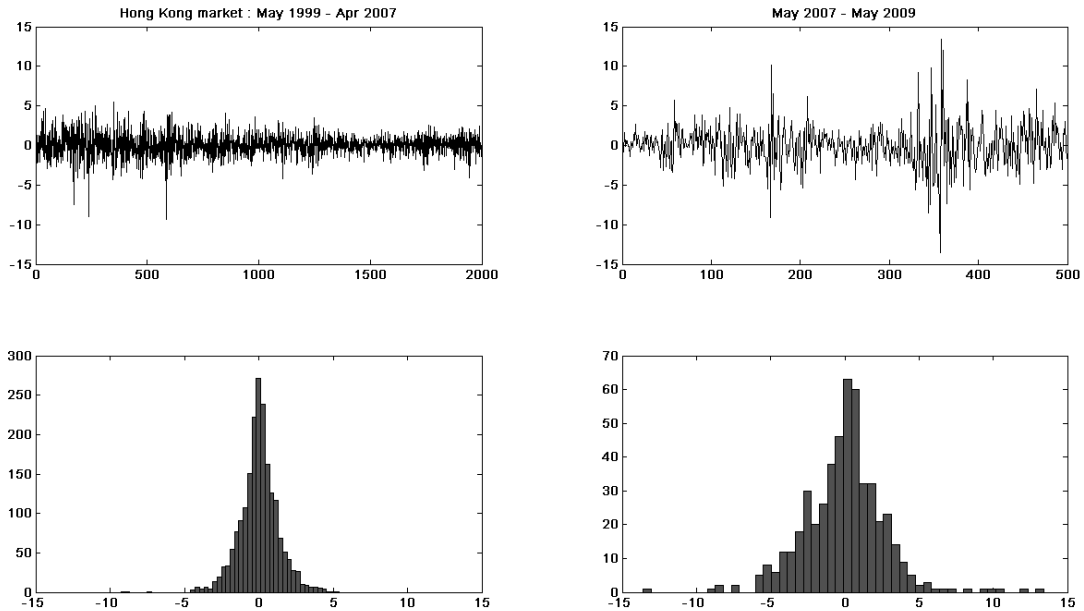


Figure 7: Chapter 2: Plots and histograms of the Hong Kong market for both the initial learning period and the forecasting period.

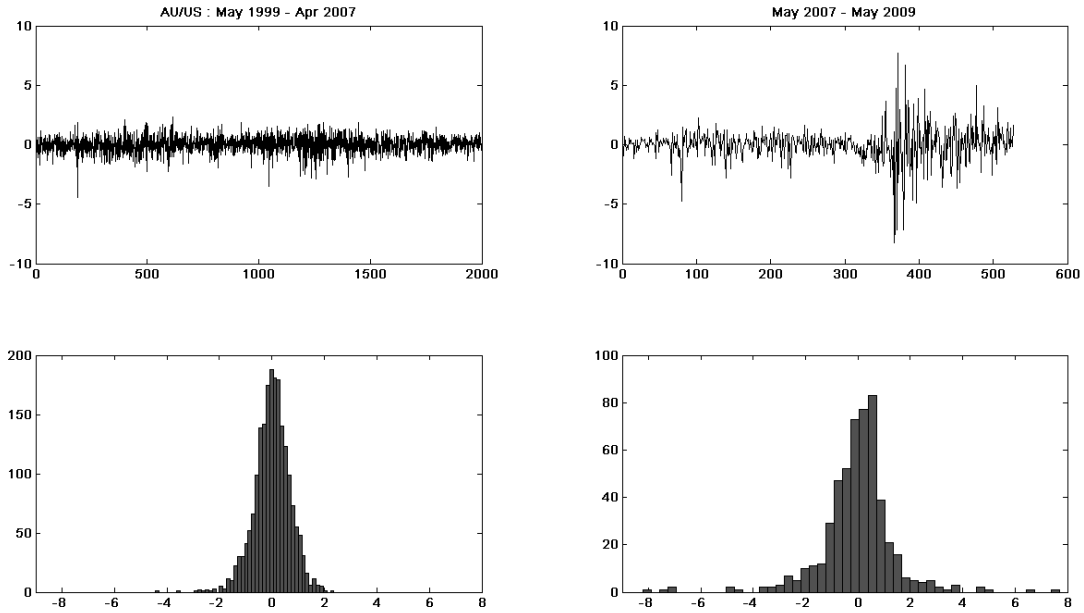


Figure 8: Chapter 2: Plots and histograms of the Australian dollar to US dollar for both the initial learning period and the forecasting period.

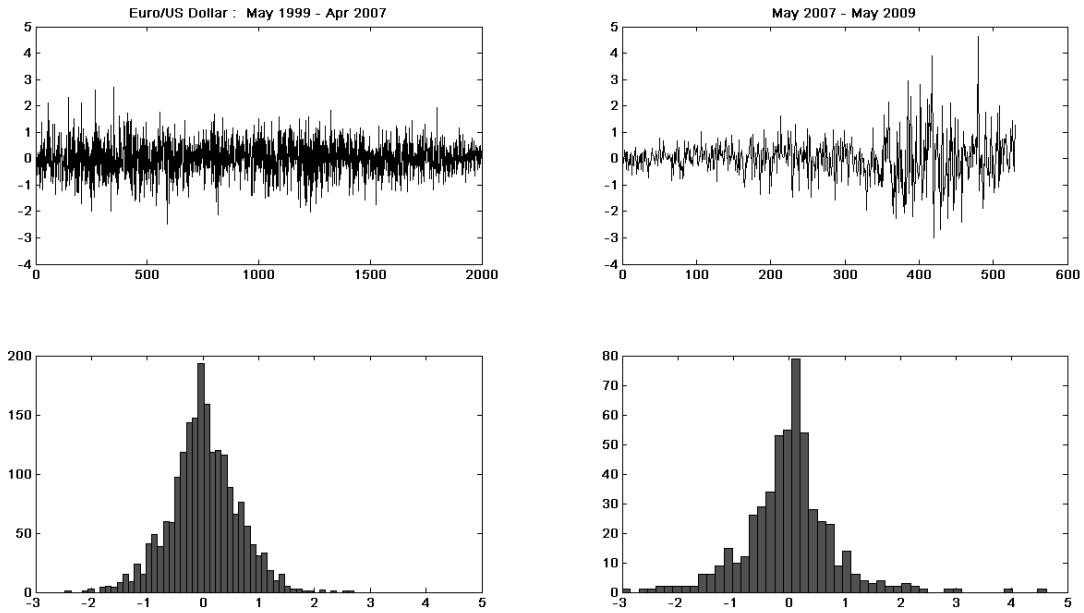


Figure 9: Chapter 2: Plots and histograms of the Euro to US dollar for both the initial learning period and the forecasting period.

Table 3: Chapter 2: Summary statistics of six return series.

Index	Period	Mean	Std	Skewness	Kurtosis	Min	Max
Aus	May 99 - Apr 07	0.036	0.71	-0.68	7.47	-5.85	3.39
	May 07 - May 09	-0.11	1.74	-0.37	5.38	-8.55	5.36
US	May 99 - Apr 07	0.006	1.11	0.07	5.58	-6.00	5.57
	May 07 - May 09	-0.11	2.21	-0.06	7.20	-9.47	10.96
UK	May 99 - Apr 07	-0.003	1.12	0.17	5.93	-5.90	5.59
	May 07 - May 09	0.083	2.01	-0.04	6.87	-9.38	9.26
HK	May 99 - Apr 07	0.022	1.32	-0.34	6.56	-9.29	5.43
	May 07 - May 09	-0.037	2.73	0.16	6.59	-13.58	13.41
AU/US	May 99 - Apr 07	0.012	0.68	-0.48	4.91	-4.45	2.36
	May 07 - May 09	-0.025	1.32	-0.85	15.10	-8.21	7.70
EUR/US	May 99 - Apr 07	0.026	0.57	-0.21	3.65	-2.47	1.94
	May 07 - May 09	-0.009	0.95	0.48	5.58	-3.00	4.62

Table 4: Chapter 2: GJR-GARCH-AL parameter estimates and standard errors (bracketed).

	Aus	US	UK	HK	AU/US	EUR/US
α_0	0.020	0.017	0.030	0.017	0.012	0.004
	(0.007)	(0.005)	(0.007)	(0.006)	(0.006)	(0.002)
α_1	0.013	0.006	0.010	0.018	0.041	0.034
	(0.011)	(0.005)	(0.009)	(0.010)	(0.018)	(0.012)
α_2	0.160	0.141	0.201	0.062	0.015	0.002
	(0.036)	(0.026)	(0.035)	(0.020)	(0.023)	(0.016)
β_1	0.895	0.921	0.885	0.947	0.937	0.961
	(0.022)	(0.014)	(0.019)	(0.012)	(0.018)	(0.013)
λ	0.999	0.999	0.998	0.999	0.999	0.999
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
u_0	1.251	1.251	1.146	2.043	1.675	2.744
	(0.465)	(0.443)	(0.441)	(0.690)	(0.565)	(0.898)
v_0	2.359	2.004	1.967	2.366	2.442	2.407
	(0.779)	(0.675)	(0.736)	(0.779)	(0.760)	(0.805)

These plots illustrate that: (i) dynamic estimates of p remained on the same side of 0.5 throughout the sample period in each market; (ii) estimated skewness was negative in each market over the sample period, except for the EUR/US rate; (iii) in the Australian and UK markets, shape and skewness changed the most compared to other markets, e.g. skewness ranged from ≈ -0.7 to ≈ -0.5 in these two markets; in contrast, estimated shape

and skewness seemed closer to a constant in the other four series for this sample period; and (iv) the GFC was notable in the Australian market on January 22nd and October 10, 2008. These large negative return days are marked and had a clear downward effect on skewness estimates. That is, the estimated value of shape parameter p suddenly increased on these extreme days, and the estimated skewness subsequently had a sharp drop, which suggests the conditional return distribution was more left skewed.

2.4.2 Risk forecasting study

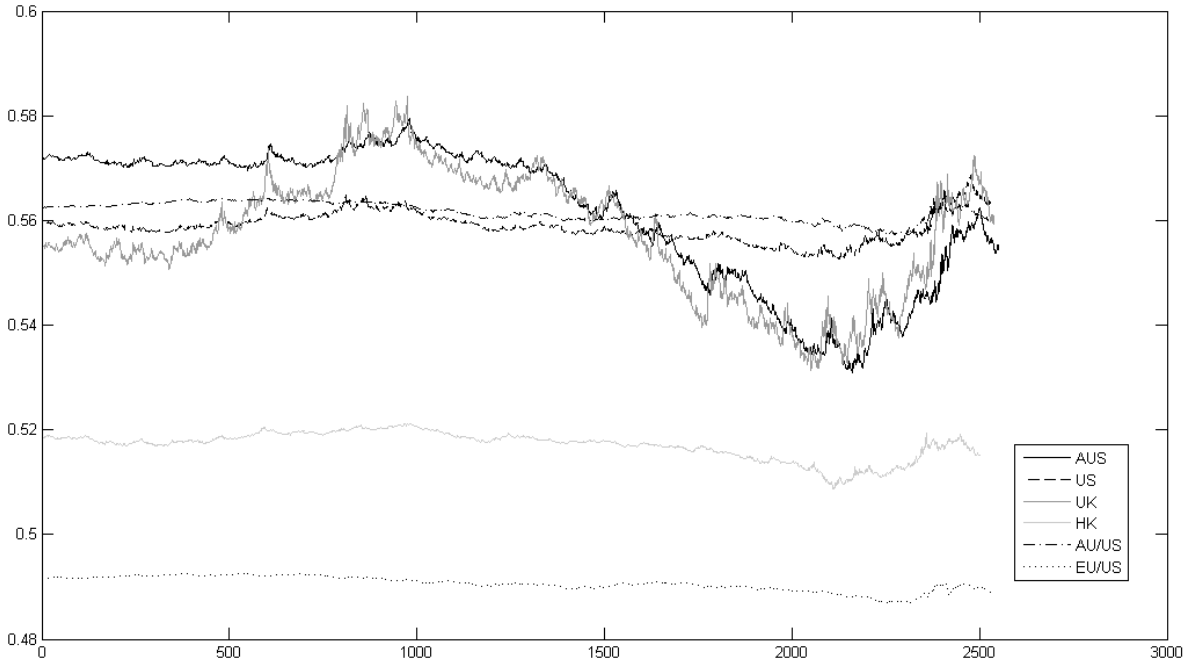
2.4.2.1 One-step-ahead VaR and ES forecasting

VaR and ES forecasts are generated using a range of competing methods for each day in the forecast sample for each market. The one-day-ahead forecasts are generated using a moving window of 2000 returns for estimation, and all parameters are re-estimated for each new forecast day and each method.

The proposed GJR-GARCH-AL model with dynamic shape is compared with: the GARCH (1,1) model with Gaussian (denoted as G-n) and Student t error (denoted as G-t); the GJR-GARCH (1,1) with Student t error (denoted as GJR-t); a semi-parametric nonlinear model called the asymmetric slope CAViaR (denoted as CAV; see Engle and Manganelli, 2004); the popular RiskMetrics model (denoted as RM; see Morgan, J. P., 1996) and two non-parametric Historical Simulation methods: short-term (25 days; ST) and medium-term (100 days; LT) sample percentiles. These two ad hoc methods are used by some financial institutions.

Unlike the parametric models used in this chapter, the CAViaR model directly models the quantile behaviour, and make minimum assumptions on the returns, such as the i.i.d. returns or the shape of whole conditional return distribution. The CAViaR model uses the regression quantile framework of Koenker and Bassett (1978) and allows the quantile vary over time under an autoregression specification. The principles of this model is that, increase the VaR immediately whenever a violation occurs, and decrease it by a marginal amount when it is not exceeded by the returns. Furthermore, the asymmetric slope CAViaR model allows different response to positive and negative returns, which is

Estimated dynamic shape p.



Estimated dynamic skewness.

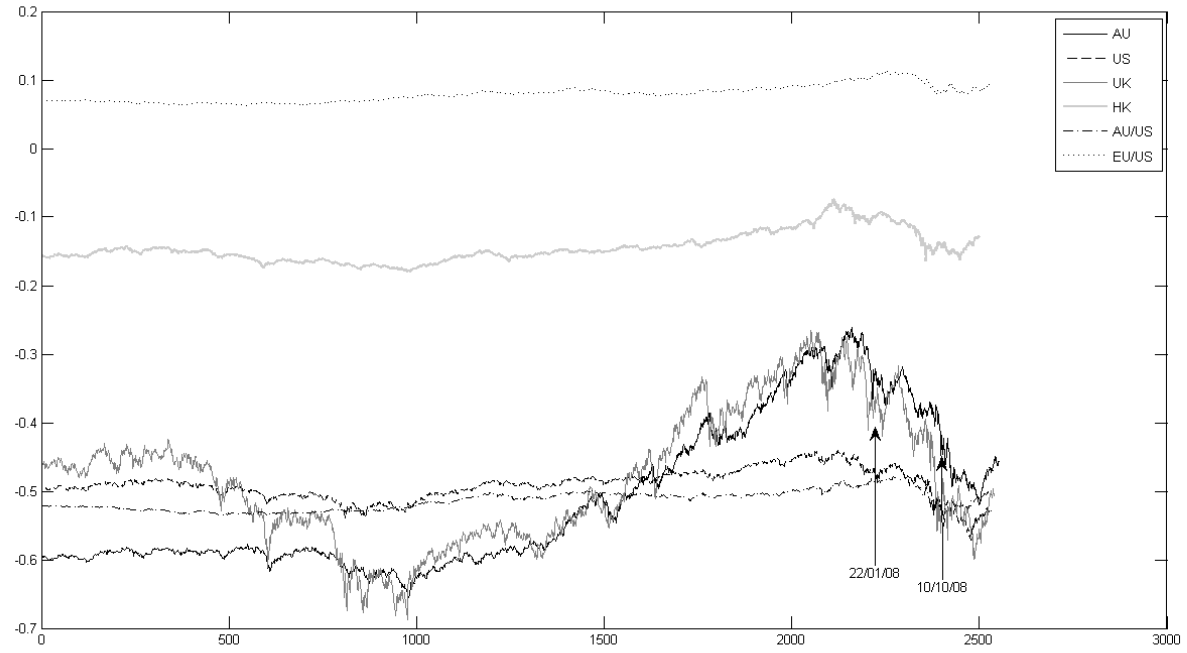


Figure 10: Chapter 2: Dynamic shape and skew from GJR-GARCH-AL across six markets from May 1999 to April 2007.

comparable to the leverage effect in the GJR-GARCH model and the potential skewness captured by the AL distribution. The CAV model can be written as

$$VaR_t = \phi_0 + \phi_1 VaR_{t-1} + (\phi_2 + \phi_3 I(y_{t-1} < 0))|y_{t-1}|, \tag{18}$$

where the parameters are estimated by minimizing the standard quantile loss function; see Engle and Manganelli (2004) for details.

The RM model is an integrated GARCH model, written as

$$y_t = \sigma_t \epsilon_t \ ; \ \epsilon_t \sim i.i.d.N(0,1) \tag{19}$$

$$\sigma_t^2 = 0.06 y_{t-1}^2 + 0.94 \sigma_{t-1}^2, \tag{20}$$

which is an EWMA variance estimate, with a fixing smoothing parameter of 0.94. A GJR-GARCH-AL model with constant shape p, denoted as GJR-ALCP, is also included for comparison to examine whether dynamic shape is important or required.

In MCMC methods, at each stage the entire parameter vector, denoted $\boldsymbol{\theta}$, has values simulated for it from the posterior, combining to give a Monte Carlo sample $\boldsymbol{\theta}^{[1]}, \ldots, \boldsymbol{\theta}^{[K]}$, where K is the MC sample size. Each of these iterates provides a one-step-ahead forecast of h_t, which can be combined with the VaR and ES formula (e.g., (8) and (9) for AL errors), to give MC iterate forecasts of VaR and ES, i.e. $\text{VaR}^{[i]}$, $\text{ES}^{[i]}$ for $i = 1, \ldots, N$, for each model. These are simply averaged over the iterates in the sampling period of the MCMC scheme, to give a one-step-ahead forecast of VaR and ES for each model.

2.4.2.2 Ten-step-ahead VaR and ES forecasting

Ten-day-ahead 1% VaR forecast is recommended by Basel II to measure the risk exposure of banks. Under a dynamic volatility model, given the information set Ω_t at time t, the conditional one-day-ahead return $y_{t+1}|\Omega_t \sim D\left(0, \sigma_{t+1}^2\right)$, i.e., $\epsilon_t \sim i.i.d.D(0,1)$. However, the distribution of l-day return y_{t+l} is not D, when $l > 1$. Hence, the ten-day-ahead return, defined as $y_t[10] = \sum_{k=1}^{h} y_{t+k}|\Omega_t$, does not follow the distribution D as well, in the GARCH-type volatility model.

As the conditional forecast distribution of $y_t[10]$ is of unknown form, Monte Carlo sampling scheme is used to directly sample from the distribution of $y_t[10]$ as follows:

***Step*1.** At jth MCMC iteration, calculate $\sigma_{t+1}^2|\Omega_t$ using current parameter estimates.

***Step*2.** Simulate a single $\epsilon_{t+1} \sim D(0,1)$ (and μ_ϵ for AL distribution).

***Step*3.** Calculate $y_{t+1} = \epsilon_{t+1}\sigma_{t+1}$ (or $y_{t+1} = (\epsilon_{t+1} - \mu_\epsilon)\,\sigma_{t+1}$ for AL distribution). Then y_{t+1} is a single observation sample from $P\left(y_{t+1}|\Omega_t\right)$.

***Step*4.** Calculate $\sigma_{t+2}^2|y_{t=1}, \Omega_t$ using y_{t+1} simulated in the above step.

***Step*5.** Simulate a single $\epsilon_{t+2} \sim D(0,1)$ (and μ_ϵ for AL distribution).

***Step*6.** Obtain $y_{t+2} = \epsilon_{t+2}\sigma_{t+2}$ (or $y_{t+2} = (\epsilon_{t+2} - \mu_\epsilon)\,\sigma_{t+2}$ for AL distribution). This is a single observation sample from $P\left(y_{t+1}|y_{t+1}, \Omega_t\right)$.

***Step*7.** Repeat ***Step*4** to ***Step*6** to obtain $y_{t+3}|\left\{y_{t+1}, y_{t+2}\right\}, \Omega_t, \ldots, y_{t+10}|\left\{y_{t+1}, \ldots, y_{t+9}\right\}, \Omega_t$ using simulated sample returns from previous steps.

***Step*8.** Calculate $y_t[10]|\Omega_t$ by summing up the MC realizations $y_{t+1}, \ldots, y_{t+10}$.

***Step*9.** Repeat the above steps for each MCMC iteration, then a Monte Carlo sample from the conditional forecast distribution for $y_t[10]|\Omega_t$ can be obtained.

Then non-overlapping ten-step-ahead $\alpha\%$ VaR forecast can be obtained by taking corresponding sample quantile of the forecasted ten-step-ahead returns. Relevant ES is the mean of the violating ten-step-ahead returns.

Specifically, for CAViaR model in (18), y_{t-1} is replaced by the sum of non-overlapping ten-day returns $y_{t-1}[10]$.

The estimation of the above models is realized in Matlab. I wrote the codes for GJR-ALCP, RM, and two HS methods, and can be contacted if the codes are of interest. The G-n, G-t, and GJR-t models are estimated by using the GARCH Toolbox in Matlab. The code for the CAViaR model is provided by my supervisor, Dr. Richard Gerlach.

2.4.3 Backtesting one-day-ahead VaR forecast models

VaR forecasts at the 1% and 5% (α) risk levels are obtained and formal and informal assessment measures are applied. As standard, the ratio VRate/α is used to compare and rank the competing models, where models with VRate/α closer to 1 are preferred.

Further, in the event of ties, conservative models (i.e. risk over-estimated and hence VRate/$\alpha < 1$) are favoured.

Three formal testing methods, i.e., UC, CC and DQ tests are also employed. We extended the CC test to allow more than one lag in testing.

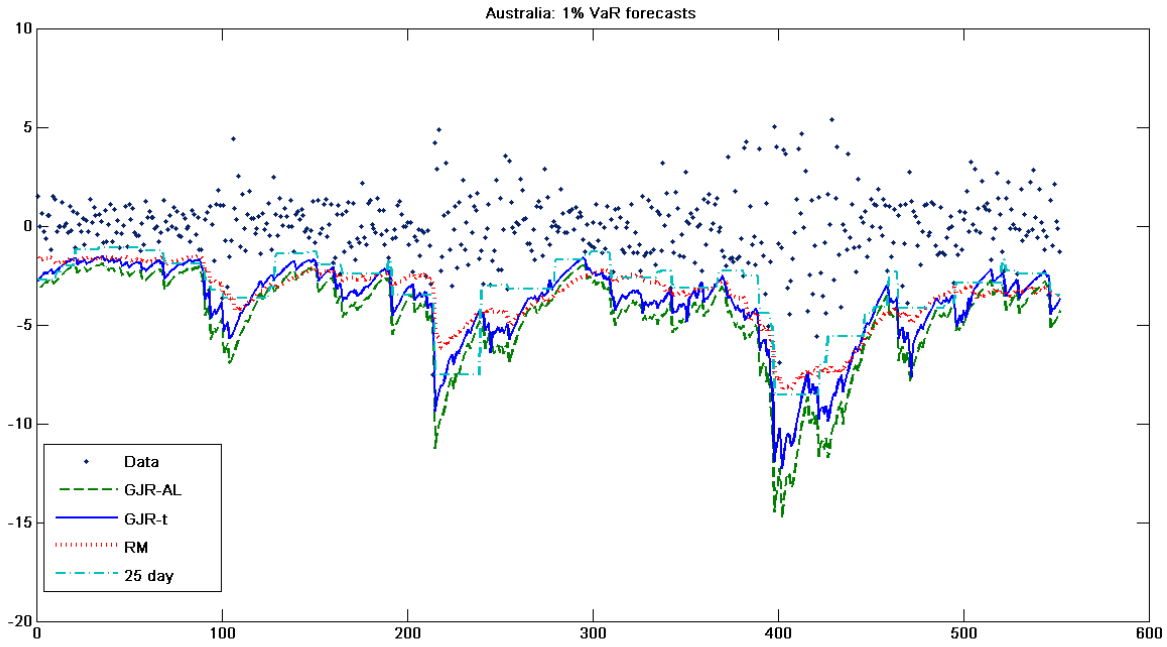
2.4.3.1 Overall forecast period performance

Table 5 presents the observed ratios VRate/α, when $\alpha = 0.01, 0.05$ for the six return series, over the forecast period for each model. Each market has ≈ 500 returns in the forecast sample, so e.g. VRate/0.01 = 5.01 for the ST method in the UK indicates that there were 5.01 times the expected violations for this method: the number of violations was 25, instead of the expected $0.01 \times 500 \approx 5$ violations. Bold ratios indicate a VRate significantly different to α via the UC test (at a 5% significance level); boxes indicate the best model, with ratio closest to 1, in that market. Clearly, both sample percentile (ST, LT) methods, and also popular methods such as RM and G-n, frequently and significantly under-estimated risk levels from May, 2007 to May, 2009, especially at the 1% risk level, as did the CAViaR model. The under-estimation was less in magnitude at the 5% risk level for these models. However, RM did much better at 5% risk, ranking first in HK, and was only rejected by the UC test once.

Selected VaR forecast series for the Australian stock market are shown in Figure 11 for 1% risk. These plots indicate that the two GJR-AL models, being quite similar, both repeatedly estimate a higher level of risk than the other models, followed by the GARCH-t and GJR-t models. The RM (which is very similar to the G-n, not shown) and ST methods repeatedly estimate a much reduced risk level. The CAViaR model seems to be similar to the AL models in extreme return periods, but subsequently and quickly reverts to similar or lower levels of risk to the RM model.

Table 5 summarizes the ratios of $\hat{\alpha}/\alpha$ across markets with an average ('Mean') and a deviation ('Std') measure; Std is the standard deviation of the observed ratios away from the expected ratio of 1. The VRate from a model for series j is denoted as $VRate_j$, and

GJR-AL, GJR-t, RiskMetrics and 25 day sample quantile (ST).



GJR-AL, GJR-ALCP, GARCH-t and CAViaR.

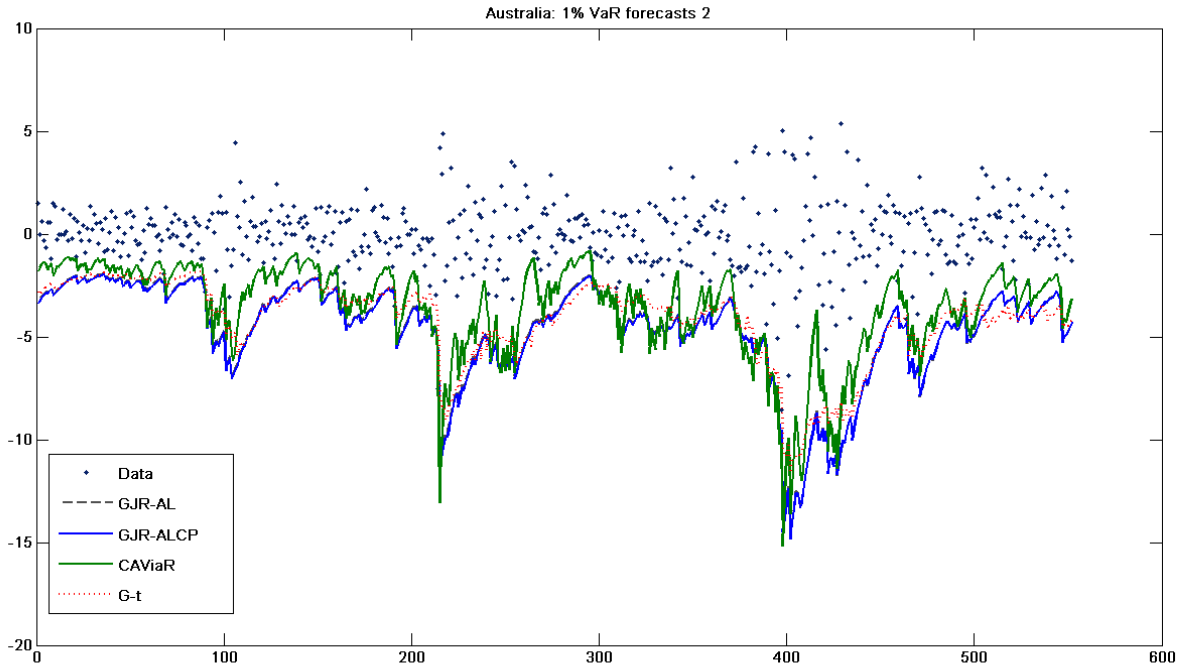


Figure 11: Chapter 2: Australian stock market: 1-day-ahead 1% VaR forecasts from various models.

Std for this model is calculated by:

$$\text{Std} = \sqrt{\frac{\sum_{j=1}^{M}(VRate_j - 1)^2}{M}}$$

where M is the number of series. Models with smaller Std tend to be more consistent in risk level forecasting, given variation in different series. The mean closest (furthest from) 1 and lowest (maximum) deviation from 1 are boxed (made bold). In terms of averages, the GARCH-t model is closest to the desired ratio of 1; however, only the two GJR-GARCH-AL models have ratios that are always below 1, with mean ratios only marginally further away from 1 than the GARCH-n and GJR-t models. The models with AL errors are the only ones that are always conservative: all others are always risky, at the 1% level; all models are always risky at the 5% level. Further, the dynamic shape GJR-AL model has the lowest deviation of ratios from 1, closely followed by the GJR-ALCP model, at the 1% level. At the 5% level, the GJR-ALCP model ranks first in both average ratio, where GARCH-t ranks second, and deviation away from 1 across markets, where the dynamic p GJR-AL ranks second.

Overall, the two GJR-AL specifications rank highly in terms of average VRate ratio and minimum deviation from 1, and are the only always conservative models at the 1% risk level. The GARCH-t and GJR-t models also rank well, though they always under-estimat risk at both levels. Ad hoc sample percentile approaches, as well as the GARCH-n and CAViaR approaches, frequently under-estimat risk in this two-year data period. The RM approach performs similarly to these methods at 1% but improved at the 5% level, though still ranking behind the models with Student t and AL errors.

Table 6 shows the counts of rejections for each model, across the six markets, for the UC, CC, and DQ test. Engle and Manganelli (2004) suggest the lag $q = 4$ for the DQ; we also ran the test for $q = 1, 2, 3$, and found minimal sensitivity to the choice of q. Further, we extended the CC test, which is a lag 1 test only, to also apply for multiple lags. This test examines a two-state (violation or non-violation) model with Markov (lag 1) dependence: i.e. given the time $t-1$ state, a Markov probability exists for the transition to the next state at time t. Our extension simply considers the previous state as the maximum of the states for the last q lags, $t-1, t-2, \ldots, t-q$, and runs the equivalent CC test, which is now a multiple-lag test. We also consider $q = 1, 2, 3, 4$, but report only

Table 5: Chapter 2: Ratios of $\hat{\alpha}/\alpha$ for 1-day-ahead VaR at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	Mean	Std.
ST	**3.99**	**4.53**	**5.01**	**3.80**	**3.04**	**4.15**	**4.34**	**3.68**
LT	1.81	**2.83**	1.86	**2.60**	1.52	**3.02**	2.27	1.53
CAViaR	**3.08**	**3.96**	**2.78**	**2.00**	**3.04**	1.70	2.55	1.80
RM	**2.17**	**3.21**	**2.97**	**2.20**	**3.04**	1.70	2.76	2.09
G-n	**2.72**	**3.40**	**2.78**	**2.20**	**3.23**	**2.26**	2.73	1.97
G-t	1.45	1.13	**2.04**	0.60	1.33	1.13	1.28	0.56
GJR-t	1.63	1.13	**2.41**	0.80	1.33	1.13	1.41	0.72
GJR-AL	0.54	0.57	0.56	0.80	0.76	**0.19**	0.57	0.52
GJR-ALCP	0.54	0.57	0.56	0.60	0.76	**0.19**	0.54	0.54
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	Mean	Std.
ST	**1.52**	**1.43**	1.26	**3.40**	1.37	1.21	1.37	0.43
LT	1.20	**1.58**	**1.48**	**1.48**	**1.40**	1.32	1.41	0.47
CAViaR	**1.74**	**2.97**	**2.20**	**3.55**	1.37	**1.70**	1.57	0.66
RM	1.20	**1.58**	1.37	1.16	1.21	1.25	1.28	0.34
G-n	**1.41**	**1.89**	1.30	**1.56**	1.37	**1.51**	1.41	0.47
G-t	1.09	1.32	1.19	0.88	1.10	1.17	1.12	0.20
GJR-t	1.30	1.28	1.22	1.00	1.06	1.32	1.20	0.26
GJR-AL	1.12	1.21	1.15	1.20	1.06	1.13	1.15	0.17
GJR-ALCP	1.05	1.21	1.11	1.16	1.06	1.06	1.11	0.14

Note: Boxes indicate that the ratio is closest to 1 in that market, bold indicates that the model is rejected by the UC test (at a 5% level), for each market.

$q = 4$ results since sensitivity was again minimal.

The sample percentile, ST and LT, CAViaR, GARCH-n, and RM models are rejected in all six markets by the DQ test, and in all or most markets by the UC and CC tests, at the 1% risk level and in four markets at the 5% risk level. The RM method does marginally better at the 5% risk level, rejected in three markets by the DQ test. These models significantly under-estimate risk and simply cannot capture the changing dynamic risk profiles in these markets in this period. The G-t model, which has displayed favourable VRate estimates, also fails to capture the risk dynamics, as measured by the DQ test, rejecting in four of the markets at risk level 1% and two markets at 5%. Though ranking fourth on most summary measures concerning VRate, the GJR-t model seems better than GARCH-n at 1% and similar at the 5% risk level, at capturing the dynamic risk in these markets: it is only rejected once by the CC and DQ tests at the 1% level, though twice

by the UC test, and three times by the DQ test at the 5% risk level.

Table 6: Chapter 2: Counts of for 1-day-ahead VaR model rejections, at the 5% significance level, at risk level $\alpha = 0.01, 0.05$.

	$\alpha = 0.01$				$\alpha = 0.05$			
Method	UC	CC	DQ	Total (out of 6)	UC	CC	DQ	Total (out of 6)
ST	**6**	**6**	**6**	**6**	3	0	4	4
LT	3	4	**6**	**6**	4	4	**6**	**6**
CAViaR	5	4	**6**	**6**	4	2	3	4
RM	5	3	**6**	**6**	1	0	3	3
G-n	**6**	4	**6**	**6**	3	1	4	4
G-t	1	0	4	4	0	1	2	2
GJR-t	2	1	1	2	0	1	3	3
GJR-AL	1	0	0	1	0	0	0	0
GJT-ALCP	1	0	0	1	0	0	1	1

Note: Boxes indicate the favoured model, bold indicates the least favoured model.

The dynamic shape GJR-AL model is favoured by all three tests at both risk levels. The GJR-ALCP ranks equally best at 1% and second at the 5% risk level. The models with Student t errors do next best. When combined with the results from the VRates, the GJR-AL models are clearly the most favoured models overall for forecasting VaR at both 5% and 1% risk levels.

2.4.3.2 Pre-financial-crisis and post-financial-crisis VaR performance

The forecast sample period covers the well-known GFC that, by all accounts, started in 2008. The performance of the models may vary between the pre-financial-crisis period and the post-financial-crisis period (which contains returns during the crisis and post-crisis). We thus present the pre- and post-crisis comparison of the models' VaR forecasting performance.

A specific date for the start of the crisis must be chosen, but there is little reason for this date to be exactly the same in each market. From news media accounts and Wikipedia, it is largely agreed that the effects of the crisis are initially apparent during September and/or October, 2008. We choose dates for each market based on maximizing the sample variance in the post-crisis period among possible days in September/October

2008. The dates thus chosen for each market were: Australia: 22nd September; the US: 19th September; the UK: 10th September; HK: 18th September; AU/US: 23rd September; and EUR/US: 23rd September, all in 2008. The sample up to the day before these dates is the pre-crisis period, while from these dates up to May 2009, is the post-crisis period. For each market, there are approximately 330 days in the pre-crisis period and approximately 170 days in the post-crisis sample.

Tables 7 and 8 show the ratios of VRate/α at $\alpha = 0.01, 0.05$ for the pre-crisis and post-crisis periods. The results for the pre-crisis sample are highly consistent with those for the whole forecast sample, no doubt influenced by the larger overlapping sample size (compared to the post-crisis case). Clearly, models with Student t and AL errors do best in terms of accurately forecasting risk levels, assessed informally and under formal UC tests. All models without AL errors always under-estimate risk, significantly so for the ad hoc and CAViaR models at both risk levels and for the G-n and RM methods at the 1% risk level. The GJR-AL models are again the only regularly conservative risk forecasters, ranking 2nd and 3rd and then 1st and 2nd for average VRate ratios and 1st and 2nd for deviation in ratios from 1, at the 1% and 5% risk levels, respectively. In this case, the GARCH-t and GJR-AL models could not be rejected by the UC tests in any market. Once again the GJR-AL models are marginally favoured for forecasting VaR levels, here prior to the financial crisis.

Results for the post-crisis period tell a slightly different story. At the 1% risk level, the GARCH-t model ranks first in average VRate ratio and in deviation of the ratios from 1, though closely followed by the RM and GJR-GARCH-AL models in deviation. At 5% risk level the GARCH-t ranks 1st in average VRate ratio, this being < 1 for the first time for this model, and 2nd in deviation from 1. Further, the RM method now ranks 3rd in VRate ratio average and 2nd in deviation from 1 at risk level 1%, and ranks 2nd and 1st in average ratio and deviation at risk level 5%, during the post-crisis period. While still being the only models to be always conservative in risk forecasting at the 1% risk level, the GJR-AL models over-estimate risk by a larger magnitude post-crisis, having only one or zero return violations in each market. Here, the expected number of violations was ≈ 1.7 in each market, for $\alpha = 0.01$, so zero or one violations is not unlikely. Once again in

Table 7: Chapter 2: Before the financial crisis: ratios of $\hat{\alpha}/\alpha$ for 1-day-ahead VaR at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	Mean	Std.
ST	**3.93**	**4.72**	**5.42**	**4.24**	**4.76**	**4.44**	**4.59**	**3.96**
LT	1.57	**3.06**	1.90	**2.42**	1.40	**3.06**	2.23	1.53
CAViaR	**2.88**	**4.72**	2.17	**2.73**	**2.52**	1.67	3.06	2.42
RM	**2.36**	**4.17**	**3.25**	**2.73**	**3.92**	1.94	2.78	2.21
G-n	**2.88**	**4.17**	**2.98**	**2.42**	**3.92**	**2.78**	3.19	2.50
G-t	1.47	1.39	2.17	0.91	1.40	1.11	1.38	0.60
GJR-t	1.57	1.39	**2.44**	1.21	1.40	1.11	1.52	0.74
GJR-AL	0.52	0.56	0.54	0.61	0.84	0.28	0.56	0.52
GJR-ALCP	0.52	0.56	0.54	0.91	0.84	0.28	0.61	0.49
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	Mean	Std.
ST	**1.57**	**1.56**	1.41	**1.58**	1.40	1.33	1.47	0.53
LT	1.26	**1.67**	**1.73**	**1.58**	**1.51**	1.22	1.49	0.58
CAViaR	**1.73**	**2.00**	1.19	**1.58**	1.23	**1.67**	**1.57**	**0.69**
RM	1.20	**1.61**	1.46	1.27	1.29	1.33	1.36	0.42
G-n	1.41	**1.72**	1.41	1.39	1.40	**1.56**	1.41	0.54
G-t	1.05	1.44	1.30	1.03	1.12	1.22	1.19	0.26
GJR-t	1.20	1.44	1.19	1.15	1.06	1.33	1.23	0.28
GJR-AL	1.05	1.33	1.08	1.33	1.01	1.11	1.15	0.22
GJR-ALCP	0.89	1.33	1.03	1.21	1.01	1.01	1.08	0.18

Note: Boxes indicate that the ratio is closest to 1 in that market, bold indicates that the model is rejected by the UC test (at a 5% level), for each market.

this case, the GARCH-t and GJR-AL models could not be rejected by the UC tests in any market. The GJR-AL models are conservative and quite close in deviation from 1 in their VRate ratios, ranking equal 3rd and 4th at risk levels 1% and 5%, respectively. However, it seems that the GARCH-t model should be marginally favoured for VaR forecasting during the post-crisis period, while the RM method also did very well at the 5% risk level.

2.4.4 Theoretical expected shortfall and backtesting for 1-day-ahead ES forecast models

Various formal backtesting methods do exist for ES: e.g. the censored Gaussian method of Berkowitz (2001), the functional delta approach of Kerkhof and Melenberg (2004), the

Table 8: Chapter 2: Post-Financial Crisis: ratios of $\hat{\alpha}/\alpha$ for 1-day-ahead VaR at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	Mean	Std.
ST	**4.12**	**4.12**	**4.12**	**3.53**	**3.53**	**3.53**	**3.82**	**3.11**
LT	**2.35**	**2.35**	1.76	**2.94**	1.76	**2.94**	2.35	1.57
CAViaR	**3.53**	**2.35**	**3.53**	0.59	**4.12**	1.76	2.65	2.24
RM	1.76	1.18	**2.35**	1.18	1.18	1.18	1.47	0.71
G-n	**2.35**	1.76	**2.35**	1.18	1.76	1.18	1.76	0.99
G-t	1.76	0.59	1.76	0.00	1.18	1.18	1.08	0.69
GJR-t	1.76	0.59	**2.35**	0.00	1.18	1.18	1.18	0.85
GJR-AL	0.59	0.59	0.59	0.00	0.59	0.00	0.39	0.73
GJR-ALCP	0.59	0.59	0.59	0.00	0.59	0.00	0.39	0.73
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	Mean	Std.
ST	1.41	1.18	0.94	1.41	0.94	0.94	1.14	0.28
LT	1.06	1.41	0.94	1.29	1.18	**1.53**	1.24	0.34
CAViaR	**1.76**	**1.65**	**1.53**	**1.53**	**1.65**	1.29	**1.57**	**0.64**
RM	1.18	1.29	1.18	0.94	1.06	1.06	1.12	0.18
G-n	1.41	1.29	1.18	1.06	1.29	1.41	1.27	0.33
G-t	1.18	1.06	0.94	0.59	1.06	1.06	0.98	0.21
GJR-t	1.53	0.94	1.29	0.71	1.06	1.29	1.14	0.33
GJR-AL	1.41	0.94	1.29	1.18	1.18	1.18	1.20	0.27
GJR-ALCP	1.41	0.94	1.29	1.06	1.18	1.06	1.16	0.24

Note: Boxes indicate that the ratio closest to 1 in that market, bold indicates that the model is rejected by the unconditional coverage test (at a 5% level), for each market.

saddle point techniques of Wong (2008), and the mean ES residual method of McNeil and Frey (2000). However, these tests tend to be difficult to implement and perhaps overcomplicated: certainly it is harder to back-test an ES than a VaR model. Kerkhof and Melenberg (2004) suggested, as a useful and simpler alternative, comparing ES, and other risk measurement methods, on an equal quantile basis, in the same way as VaR models are back-tested by VRate. This method relies on calculating the specific quantile that the ES should fall at, and hence can only be used for fully parametric ES models. We do not consider this a problem, since the three non-parametric or semi-parametric methods performed the worst of all methods for VaR forecasting, and we assume they would also do poorly for quantiles further out in the distribution tail, where ES is located.

McNeil et al. (2005) report the closed-form solutions for ES for the Gaussian and Student t distributions, while ES for the GJR-AL model is given in (13) and (15). For

the Gaussian and AL distributions, as shown in Table 9 that the quantile level of the ES depends only on α and not on the mean and variance for the Gaussian, or the location, scale, and shape parameters for the AL distribution. The independence of the AL ES quantile from p is surprising, but it is simple to show that $F(ES_\alpha|p) = \frac{\alpha}{e}$ where e is the natural exponent. Thus, for the Gaussian and AL error models, tests equivalent to the UC, CC, and DQ ones can be run to examine violations from ES forecasts, using the nominal levels (denoted as δ_α) at which the ES is located, as given in Table 10. The ES quantile level for the Student t distribution does depend on the degrees of freedom parameter ν. As the estimate for ν will necessarily change for each forecast day, so will the expected ES quantile level. However, as an approximation, we use the quantile levels corresponding to the average of the 500 estimates of ν, so that approximate versions of the tests and informal assessment of ES violation rates can be conducted for the Student t models.

Table 9: Chapter 2: ES quantile level function for the Gaussian and AL distributions.

Distribution	ES_α
$N(0,1)$	$\Phi\left(-\frac{1}{\alpha}\phi\left(\Phi^{-1}(\alpha)\right)\right)$
t_ν^*	$T_\nu\left(-\frac{1}{\alpha}t_\nu\left(T_\nu^{-1}(\alpha)\right)\left(\frac{\nu+(T_\nu^{-1}(\nu))^2}{\nu-1}\right)\sqrt{\frac{\nu-2}{\nu}}\right)$
$AL(0,1,p_t)$	$\frac{\alpha}{\exp(1)}\ ;\ 0 \leq \alpha < p_t$

Note: $\phi(\cdot)$, $\Phi(\cdot)$ and $\Phi(\cdot)^{-1}$ are the pdf, cdf and inverse cdf of normal distribution, respectively; t_ν, T_ν and T_ν^{-1} are the pdf, cdf and inverse cdf of standardized Student t distribution.

The approximate quantiles for ES from the G-t and GJR-t models, using the average of the 500 estimates for ν for each model, are presented in Table 11 for each of the six markets. Hence, the violation rate for ES, denoted as ESRate, is defined as

$$\text{ESRate} = \frac{1}{m}\sum_{t=n+1}^{n+m} I(y_t < \text{ES}_t),$$

and good models should have ESRate closer to the nominal δ_α.

Table 10: Chapter 2: Nominal levels δ_α for 1-day-ahead ES for the Gaussian and AL distributions.

α	$\Phi(ES_\alpha)$	$F(ES_\alpha \vert p)$
α	$N(0,1)$	AL
0.01	0.0038	0.0037
0.05	0.0196	0.0184

Table 11: Chapter 2: Estimated nominal ES levels δ_α for 1-day-ahead ES at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US
G-t	0.0037	0.0036	0.0034	0.0035	0.0033	0.0036
GJR-t	0.0038	0.0039	0.0035	0.0036	0.0033	0.0037
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US
G-t	0.0184	0.0184	0.0185	0.0182	0.0177	0.0187
GJR-t	0.0187	0.0190	0.0187	0.0182	0.0178	0.0187

Table 12: Chapter 2: ESRatios $\hat{\delta}_\alpha/\delta_\alpha$ for 1-day-ahead ES at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	Mean	Std.
RM	**3.77**	**4.41**	**4.34**	**4.16**	**4.93**	**1.96**	3.93	3.37
G-n	**5.18**	**3.92**	**4.82**	**4.16**	**4.44**	**2.94**	**4.24**	**3.64**
G-t	**2.42**	**2.09**	**4.32**	0.57	**4.06**	0.52	2.33	2.19
GJR-t	**2.37**	1.46	**4.27**	1.10	**4.05**	0.51	2.29	2.11
GJR-AL	0.49	0.00	0.50	0.00	1.54	0.00	0.42	0.87
GJR-ALCP	0.49	0.00	0.50	0.00	1.54	0.00	0.42	0.87
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	Mean	Std.
RM	**2.22**	**2.31**	**1.80**	**1.74**	**1.94**	1.45	1.91	1.05
G-n	**2.04**	**2.51**	**1.99**	**1.84**	**2.42**	**1.74**	**2.09**	1.23
G-t	1.15	1.46	1.43	0.90	1.23	1.03	1.84	0.98
GJR-t	1.31	1.59	1.91	0.85	1.37	1.08	1.97	1.44
GJR-AL	0.79	0.51	1.01	1.20	0.72	0.51	0.79	0.36
GJR-ALCP	0.59	0.51	1.01	0.98	0.72	0.51	0.72	0.38

Note: Boxes indicate a ratio closest to 1 in that market, bold indicates the model is rejected by the UC test (at a 5% level), for each market.

2.4.4.1 Overall forecast period performance

Table 12 presents the ratios of $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$. At both risk levels, there is a slightly different story to the corresponding VaR forecast performance. The GJR-AL

models are the only always conservative ES risk level forecasters, have the average ESRate ratios closest to 1, and the lowest deviation away from a ratio of 1 across markets. Given the nominal quantile levels for AL ES in Table 10, there is expected to be 1.5 violations for 500 observations in the forecast period. Thus, zero violations and subsequently the zero VRate ratio in Table 12 for GJR-AL models is acceptable. The G-t, GJR-t, RM, and G-n models all significantly under-estimate risk in most markets (all markets for RM and G-n) at 5% risk, while the RM and G-n models significantly under-estimate ES risk levels for 1% risk also. At 1% risk, the G-t and GJR-t models also frequently under-estimate risk levels with average ESRate ratios well above 1. The GJR-AL models have dominated the other models for ES forecasting across the entire forecast sample in these markets.

The ES forecasts from selected models for the Australian stock market are shown in Figure 12 for $\alpha = 0.01$, confirming the evidence from the earlier VaR figure. The AL-based models are forecasting a higher level of risk than the other models. The Student t forms forecast lower levels of risk, while the RM and GARCH-n have the lowest risk forecastings.

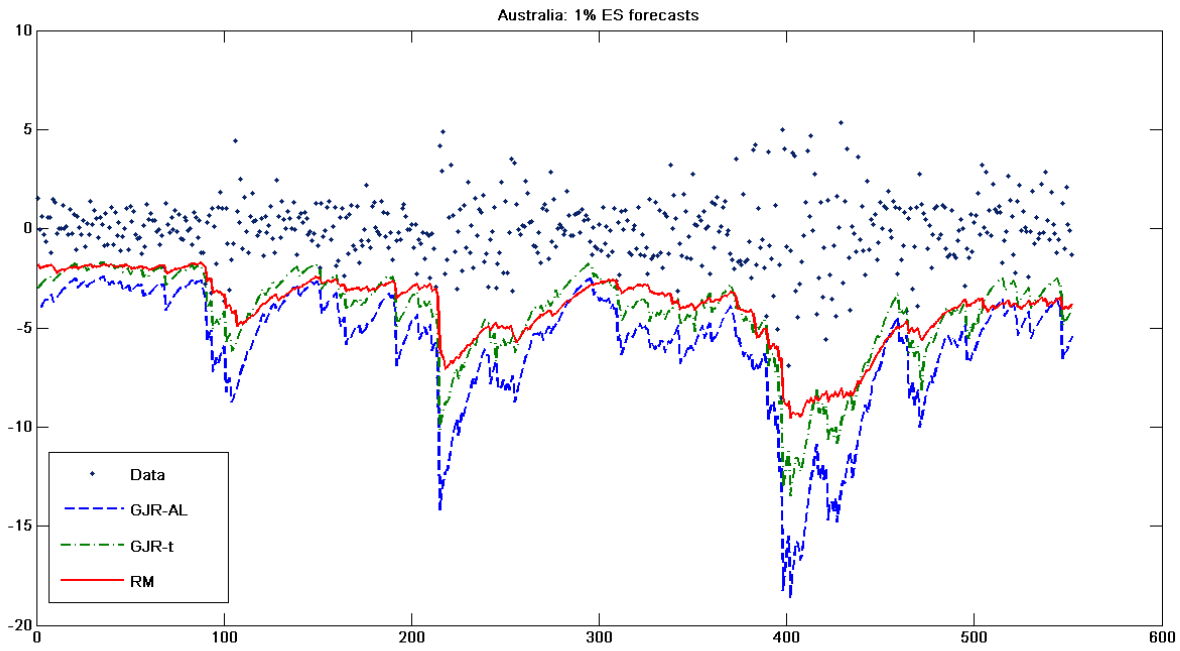
UC, CC, and DQ tests are now applied to the violations from each model's ES forecasts, using the nominal levels for each model and market in Tables 10 and 11. The numbers of model rejections over the six markets are presented in Table 13; the CC and DQ tests again used four lags. Results are similar to those for the tests applied to the VaR forecasts for each model. The G-n and RM methods are rejected in five or six markets; the DQ test captures the poor risk dynamics of the G-t and GJR-t models; the GJR-AL models do best for each test and cannot be rejected at all for capturing risk dynamics.

In summary, for the overall sample, the dynamic shape GJR-AL seems marginally the preferred choice for ES forecasting. It is frequently close to the expected number of violations, both in average and deviation from 1, and fares the best under the formal tests of unconditional and conditional coverage across the six markets.

2.4.4.2 Pre-financial-crisis and post-financial-crisis performance

Again it is informative to examine ES forecasting during the pre-crisis and post-crisis periods. Tables 14 and 15 present the relevant ESRate ratios in each period. Pre-crisis,

GJR-AL, GJR-t and RiskMetrics.



GJR-AL, GJR-ALCP, GARCH-t and GARCH-n.

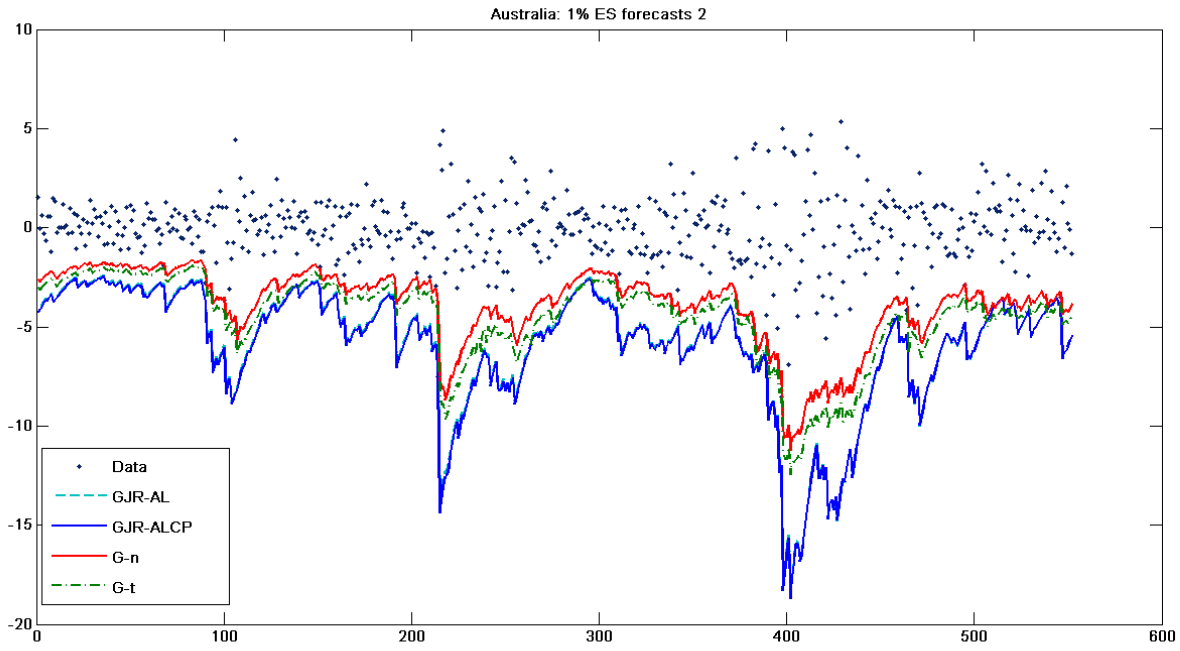


Figure 12: Chapter 2: Australian stock market: 1% 1-day-ahead ES forecasts from various models.

Table 13: Chapter 2: Counts of 1-day-ahead ES model rejections at $\alpha = 0.01, 0.05$.

	$\alpha = 0.01$				$\alpha = 0.05$			
Method	UC	CC	DQ	Total (out of 6)	UC	CC	DQ	Total (out of 6)
RM	5	5	5	5	5	3	5	5
G-n	**6**	5	**6**	**6**	**6**	4	**6**	**6**
G-t	2	1	3	3	4	2	4	4
GJR-t	2	2	2	2	4	3	**6**	**6**
GJR-AL	1	0	0	1	0	0	0	0
GJR-ALCP	2	0	0	2	0	0	0	0

Note: Boxes indicate the favoured model, bold indicates the leaset favoured models (at a 5% level).

the GJR-AL models again dominate the other models. Average ESRate ratios are much closer to 1 in average and deviation and they are the only conservative models of risk. Again, in most markets the RM, G-n, G-t, and GJR-t models significantly under-estimate the ES risk levels at both the 5% and 1% levels, and are mostly rejected by the UC test. The AL error models cannot be rejected once again. Finally, in the post-crisis period much the same story applies. The GJR-AL models are still conservative and rank first in terms of average ratio and deviation at $\alpha = 0.01, 0.05$; all other models frequently and often significantly under-estimate the ES risk levels and have ratios > 1.

2.4.5 Further model assessment

Three assessment methods are considered to further evaluate the risk model performance. Besides VRate, the market risk charges and absolute deviation of violating returns are also recommended by Basel II to help financial institutions form better evaluation of VaR models.

2.4.5.1 Market risk charges

The function of capital charge is given by Jorion (2002) as follows:

$$\text{MRC}_t = sup\left(\text{VaR}_t, M_t \times \sum^{6} 0_{i=1} VaR_{t-i}\right), \tag{21}$$

where VaR_t is the current estimate of 1% VaR, M_t is the penalty factor in Table 2, based on the violation numbers at 1% risk level in 250 trading days.

Table 14: Chapter 2: Before the financial crisis: $\hat{\delta}_\alpha/\delta_\alpha$ for 1-day-ahead ES at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	Mean	Std.
RM	**4.13**	**5.85**	**4.28**	**4.78**	**5.90**	**2.92**	**4.64**	**4.15**
G-n	**4.82**	**5.12**	**4.99**	**4.78**	**5.16**	**2.92**	4.63	4.07
G-t	**2.74**	**2.31**	**4.06**	0.88	**4.41**	0.77	2.53	2.27
GJR-t	2.01	1.40	**3.99**	1.64	**4.40**	0.76	2.37	2.10
GJR-AL	0.71	0.00	0.73	0.00	1.51	0.00	0.49	0.83
GJR-ALCP	0.71	0.00	0.73	0.00	1.51	0.00	0.49	0.83
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	Mean	Std.
RM	**2.14**	**2.55**	**1.94**	**2.01**	**2.14**	**1.70**	2.08	1.22
G-n	**2.00**	**2.69**	**2.07**	**2.16**	**2.43**	**1.84**	**2.20**	1.35
G-t	1.85	**2.87**	**2.06**	1.83	**2.22**	**1.79**	2.10	1.27
GJR-t	1.67	**2.48**	**2.32**	1.83	**2.22**	**1.78**	2.15	1.20
GJR-AL	0.85	0.60	1.03	1.48	0.76	0.45	0.86	0.39
GJR-ALCP	0.43	0.60	1.03	1.32	0.76	0.45	0.77	0.44

Note: Boxes indicate the ratio is closest to 1 in that market, bold indicates that the model is rejected by the unconditional coverage test (at a 5% level), for each market.

Table 15: Chapter 2: After the financial crisis: $\hat{\delta}_\alpha/\delta_\alpha$ for 1-day-ahead ES at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	Mean	Std.
RM	**3.10**	**1.55**	**4.64**	**3.10**	**3.10**	0.00	2.58	2.36
G-n	**6.19**	**1.55**	**4.64**	**3.10**	**3.10**	**3.10**	**3.61**	**3.28**
G-t	1.65	1.63	**4.79**	0.00	**3.36**	0.00	1.91	2.13
GJR-t	**3.23**	1.60	**4.74**	0.00	**3.35**	0.00	2.15	2.32
GJR-AL	0.00	0.00	0.00	0.00	1.59	0.00	0.27	1.03
GJR-ALCP	0.00	0.00	0.00	0.00	1.59	0.00	0.27	1.03
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	Mean	Std.
RM	**2.40**	**1.80**	1.50	1.20	1.50	0.90	1.55	0.79
G-n	**2.10**	**2.10**	**1.80**	1.20	**2.40**	1.50	**1.85**	**1.03**
G-t	1.93	0.96	1.57	0.64	**1.97**	0.63	1.28	069
GJR-t	**2.53**	1.57	**2.48**	0.65	**2.62**	0.94	1.80	1.23
GJR-AL	0.96	0.32	0.96	0.32	0.64	0.64	0.64	0.49
GJR-ALCP	0.96	0.32	0.96	0.32	0.64	0.64	0.64	0.49

Note: Boxes indicate the ratio closest to 1 in that market, bold indicates that the model is rejected by the unconditional coverage test (at a 5% level), for each market.

Table 16 shows the average daily MRC across the forecasting period for each market. The G-n model has smallest average daily capital charge in 3 out of 6 markets, thus it ranks

highest under this back-test. GJR-t model ranks second and has smallest capital charge in 2 markets. LT and GJR-ALCP model are least favoured under this criterion, with largest average daily capital charge in 3 out of 6 markets. These largest capital charges are approximately 2-4% higher than the smallest capital charges in each market. Considering the always conservative performance of GJR-ALCP model, the large capital charges are likely due to large size of VaR forecast from a much heavier ALD tail, rather than more return violations. In fact, the MRC method should be combined with violation rates to assess the performance pf VaR models. A model with considerably high violation rate but smaller MRC is not better than a conservative model with larger MRC. A distribution with less fat tails than the AL distribution, yet heavier than the Student t distribution may produce more adequate capital charges. This will be of interest for future research. However, as the VRate and UC, CC and DQ tests are effective enough, and back-test of capital charge is not applicable to the ES, which is a major focus of this dissertation, the back-test of capital charge will not be considered in later chapters.

Table 16: Chapter 2: Average daily market risk charge for 1-day-ahead VaR at $\alpha = 0.01$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US
ST	12.86	14.54	13.97	19.08	10.26	5.40
LT	**14.13**	16.46	15.18	**21.79**	**11.48**	5.53
CAViaR	12.75	15.29	12.79	19.11	9.78	5.40
RM	12.03	15.63	14.66	20.18	10.28	5.20
G-n	11.90	15.43	13.90	17.83	9.85	5.18
G-t	12.71	14.65	13.77	20.92	9.70	5.33
GJR-t	11.94	14.35	13.78	20.14	9.53	5.22
GJR-AL	13.74	16.61	15.41	20.04	10.35	5.85
GJR-ALCP	14.00	**16.82**	**15.57**	20.29	10.41	**5.91**

Note: Boxes indicate the smallest capital charge in that market, bold indicates the largest capital charges, for each market.

2.4.5.2 Absolute deviation of violating returns

The absoluate deviation (AD) is defined as follows:

$$\mathrm{AD} = |y_t - (\mathrm{VaR}_t)\,| I\left(y_t < \mathrm{VaR}_t\right).$$

As in Chen, Gerlach, Lin and Lee (2011), the AD mean and maximum are calculated and combined with VRate in Table 5 for comparing risk models. The AD mean of violating returns are presented in Table 17, the average of the mean AD for each model is also included. Both GJR-AL and GJR-ALCP model have smallest AD mean in 2 out of 6 markets at 1% risk level. G-t has smallest AD mean in one market, while the CAViaR model has smallest average of AD mean. However, the CAViaR model significantly under-estimates risk levels and has large VRate as shown in Table 5, GJR-AL models are therefore preferred. At 5% level, it is clearly that GJR-AL performs most favourably, with smallest AD mean in two markets, as well as the smallest average of AD mean. GJR-t model ranks second, with smallest AD mean in two markets, followed by ST and GJR-ALCP model. At both risk levels, LT is the least favoured model, generating deviation of violating returns from the VaR threshold as large as three times of that from the GJR-AL model.

Table 18 shows the AD maximum of violating returns. Apparently, the GJR-ALCP model outperforms the other models, with smallest AD maximum in 3 out of 6 markets, as well as smallest average of AD maximum at both 1% and 5% levels. The LT model again significantly under-estimates the risk at both risk levels, hence the size of maximum deviation at violation is around 3 times of that from the GJR-ALCP model. The G-t ,GJR-t and GJR-AL model perform equivalently well at 5% level, ranking behind the GJR-ALCP model.

Overall, the AD method tells a consistent story as VRate, the two GJR-AL specifications are considered the best models, with smallest AD average and maximum when returns exceeding the VaR threshold. Specifically, the LT model significantly under-estimates the risk, and once a violation occurs, the magnitude of loss is massive.

Table 19 shows the AD mean, representing the average size of excessive loss for the returns violating ES threshold. At 1% risk level, GJR-ALCP has smallest AD mean in 5 out of 6 markets, as well as smallest average of AD mean. GJR-AL model ranks second with smallest AD mean in 3 markets. The RM model and GJR-t has largest mean AD in two markets, while G-n and G-t has largest AD mean in one market, and the latter also has largest average AD mean. At 5% level, GJR-AL specification performs most favourably,

Table 17: Chapter 2: AD mean of violating 1-day-ahead returns at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	Mean
ST	0.99	**0.86**	0.94	1.28	0.76	0.37	0.87
LT	**1.33**	0.68	**1.44**	**1.41**	1.68	0.20	**1.12**
CAViaR	0.55	0.61	0.78	0.59	0.90	0.26	0.61
RM	1.07	0.59	0.82	1.11	0.90	0.20	0.78
G-n	0.92	0.59	0.89	1.19	0.88	0.23	0.78
G-t	0.79	0.68	0.79	0.66	1.40	0.13	0.74
GJR-t	0.68	0.65	0.72	0.37	1.42	0.15	0.67
GJR-AL	0.82	0.51	0.56	0.42	**1.70**	**0.38**	0.73
GJR-ALCP	0.76	0.51	0.57	0.30	1.68	0.37	0.70
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	Mean
ST	0.78	0.93	1.09	1.05	0.84	**0.42**	0.85
LT	**1.28**	**1.06**	**1.14**	**1.51**	**1.23**	0.37	**1.10**
CAViaR	0.73	0.98	0.93	1.01	0.94	0.31	0.82
RM	1.04	0.98	1.00	1.35	0.99	0.28	0.94
G-n	0.94	1.00	1.04	1.25	0.95	0.29	0.91
G-t	0.89	0.83	1.02	1.24	0.93	0.28	0.86
GJR-t	0.77	0.81	1.07	1.01	0.99	0.25	0.82
GJR-AL	0.76	0.78	0.92	1.06	0.97	0.29	0.80
GJR-ALCP	0.82	0.78	0.94	1.08	0.96	0.29	0.81

Note: Boxes indicate that the lowest AD in that market without rejection by UC test, bold indicates that the highest AD.

and has smallest AD mean in 3 markets as well as smallest average of AD mean. It is followed by GJR-t and GJR-A. However, each of G-t, GJR-AL and GJR-ALCP has largest AD mean in two markets.

Table 20 presents the maximum AD for ES. The GJR-ALCP outperforms the other models once again, with smallest maximum loss in all markets, as well as smallest average maximum AD, at both 1% and 5% risk levels. RM is least favoured model, with largest maximum AD in 3 out of 6 markets, followed by the G-n model. This result is highly consistent with the maximum AD for VaR in Table 18.

Overall, the two GJR-AL specifications outperform their competitors again, by minimizing the average loss of violating returns, represented by the mean and maximum AD. Furthermore, the result of AD maximum is of less opacity as that of AD mean.

Table 18: Chapter 2: AD maximum of violating for 1-day-ahead returns at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	Mean
ST	4.06	**4.39**	2.73	4.92	5.45	**1.44**	3.83
LT	4.32	**4.39**	**3.45**	**4.99**	**5.63**	0.78	**3.93**
CAViaR	1.41	4.27	2.33	1.65	4.97	0.73	2.56
RM	**4.46**	3.75	2.74	3.05	4.75	0.81	3.26
G-n	4.18	3.76	2.73	2.66	4.98	0.67	3.16
G-t	3.42	2.61	2.51	1.64	4.32	0.51	2.50
GJR-t	2.67	1.94	1.91	0.84	4.33	0.55	2.04
GJR-AL	1.80	1.07	0.88	0.62	3.75	0.38	1.42
GJR-ALCP	1.68	1.10	0.85	0.58	3.73	0.37	1.38
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	Mean
ST	4.46	4.39	3.82	5.03	5.72	**1.52**	4.16
LT	**6.00**	6.13	**6.16**	**8.39**	**6.28**	1.17	**5.69**
CAViaR	3.75	**6.15**	3.76	3.99	5.74	1.07	4.08
RM	5.36	5.35	3.95	5.72	5.77	1.14	4.55
G-n	5.16	5.36	3.64	5.55	5.92	1.04	4.45
G-t	4.90	4.89	3.49	4.55	5.70	0.98	4.08
GJR-t	4.36	4.58	3.72	3.81	5.70	1.00	3.86
GJR-AL	4.22	4.52	3.34	4.13	5.63	1.02	3.81
GJR-ALCP	4.17	4.54	3.31	4.15	5.62	1.02	3.80

Note: Boxes indicate that the lowest AD in that market without rejection by UC test, bold indicates that the highest AD.

Though GJR-ALCP model ranks the first under the criterion of smallest AD, the performance of GJR-AL is very similar to and closely following that of GJR-ALCP. Since the results of AD mean and maximum are largely consistent with the VRate method, as well as the other three formal tests, the AD results is demonstrated in this chapter for example, and will not be shown in later chapters.

2.4.5.3 Loss function

Loss functions are evaluated for VaR and ES forecast at 1% and 5% risk levels. ES forecasts can also be assessed at their approximate quantile levels, whereby δ_α is substituted for α in (2). The best risk forecasts in terms of accuracy should minimise this loss function.

Tables 21 and Table 22 present the loss function for the VaR and ES forecasts from

Table 19: Chapter 2: AD mean for 1-day-ahead ES at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	Mean
RM	**1.10**	0.57	**0.74**	0.68	1.06	0.20	0.73
G-n	0.77	0.66	**0.74**	0.62	1.25	0.19	0.70
G-t	0.91	0.66	0.63	**1.16**	1.18	0.36	**0.81**
GJR-t	0.78	**0.78**	0.61	0.22	1.19	**0.40**	0.66
GJR-AL	0.29	0.00	0.35	0.00	**1.49**	0.00	0.35
GJR-ALCP	0.14	0.00	0.30	0.00	1.47	0.00	0.32
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	Mean
RM	0.79	0.72	**1.05**	1.11	0.96	0.23	0.81
G-n	0.89	0.73	0.97	1.05	0.83	**0.27**	0.79
G-t	0.89	0.67	0.99	**1.17**	0.93	0.26	**0.82**
GJR-t	0.81	0.56	0.81	1.09	0.86	0.25	0.73
GJR-AL	0.66	**0.82**	0.57	0.63	**1.46**	0.22	0.73
GJR-ALCP	**0.95**	**0.82**	0.55	0.64	1.45	0.21	0.77

Note: Boxes indicate that the lowest AD in that market, bold indicates that the highest AD, for each market.

Table 20: Chapter 2: Maximum absolute deviation for 1-day-ahead ES at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	Mean
RM	**4.01**	2.95	**2.32**	**2.17**	4.25	**0.65**	2.73
G-n	3.69	**2.97**	2.31	1.73	**4.51**	0.48	2.61
G-t	3.13	2.03	2.13	1.16	4.01	0.36	2.13
GJR-t	**4.01**	1.44	1.43	0.27	4.03	0.40	1.94
GJR-AL	0.29	0.00	0.35	0.00	2.58	0.00	0.53
GJR-ALCP	0.14	0.00	0.30	0.00	2.56	0.00	0.50
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	Mean
RM	**4.81**	4.37	**3.07**	**3.73**	5.14	**0.94**	3.68
G-n	4.56	**4.38**	3.06	3.51	**5.34**	0.81	3.61
G-t	4.46	4.07	3.01	3.60	5.25	0.78	3.53
GJR-t	**4.81**	3.81	2.72	2.94	5.26	0.81	3.39
GJR-AL	2.72	2.38	1.35	1.94	4.46	0.62	2.25
GJR-ALCP	2.62	2.40	1.33	1.91	4.45	0.62	2.22

Note: Boxes indicate that the lowest AD in that market, bold indicates that the highest AD, for each market.

various models, taken over the six series in the entire forecast period. For VaR at 1% level, the GJR-AL model has the lowest average loss function, while the GJR-t model has lowest loss function in two markets. At 5%, it is clear that GJR-ALCP model is most

favoured under this criteria, with lowest average loss function, as well as ranking first in 5 out of 6 markets. Both sample percentile methods (ST, LT) frequently generate highest loss function at both risk levels, thus is least favoured. Similarly, for the ES, it is obvious that the GJR-ALCP is the best model with lowest loss at both risk levels, followed by G-t model, with lowest average loss function at 1% level. The results of loss function method is consistent with those of other testing methods.

Table 21: Chapter 2: Loss function for 1-day-ahead VaR at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	Mean
ST	**39.17**	**39.90**	**44.28**	49.25	31.32	**15.58**	**36.58**
LT	37.53	34.54	39.90	**49.91**	**33.24**	11.61	34.46
CAViaR	29.10	33.59	30.94	33.38	28.65	10.99	27.77
RM	31.84	32.23	34.30	40.91	29.04	10.72	29.84
G-n	27.43	32.19	34.13	39.87	28.90	11.33	28.97
G-t	29.13	30.62	31.81	37.93	27.50	10.62	27.94
GJR-t	31.84	29.60	31.56	35.78	27.23	10.69	27.78
GJR-AL	27.81	31.79	29.75	35.21	25.58	11.28	26.90
GJR-ALCP	27.89	31.72	29.83	35.48	25.52	11.33	26.96
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	Mean
ST	103.00	116.22	111.84	137.24	84.11	**44.02**	99.41
LT	**110.50**	**125.74**	**120.47**	**157.85**	**97.25**	43.58	**109.23**
CAViaR	99.53	119.32	107.18	133.20	83.58	42.93	97.62
RM	100.82	116.92	111.55	140.19	83.22	40.86	98.93
G-n	97.09	117.13	110.28	138.57	83.45	42.07	98.10
G-t	99.54	114.88	109.95	141.09	83.35	41.43	98.37
GJR-t	100.82	112.18	110.14	134.52	82.89	41.27	96.97
GJR-AL	96.64	111.09	108.79	133.96	81.49	40.76	95.46
GJR-ALCP	96.64	110.97	108.50	132.15	81.13	40.51	94.98

Note: Boxes indicate that the lowest loss function in that market, bold indicates that highest loss function in that market.

2.4.6 Results summary and discussion for one-step-ahead risk forecasts

There is a fairly consistent story across the markets for both risk measures. The GJR-GARCH-AL specifications are the only always conservative methods, in that they slightly but consistently over-estimate risk over the two-year forecast sample period, while maintaining the violation ratios closest to 1, in terms of squared error and significance testing

Table 22: Chapter 2: Loss function for 1-day-ahead ES at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	Mean
RM	**17.17**	**14.93**	16.08	**18.11**	17.09	4.70	**14.68**
G-n	12.85	14.80	**16.51**	17.25	**17.39**	4.93	13.95
G-t	13.73	13.05	13.82	14.53	14.48	4.27	12.31
GJR-t	17.11	13.29	13.49	13.81	14.43	4.34	12.75
GJR-AL	12.30	14.18	13.49	16.02	13.24	5.06	12.38
GJR-ALCP	12.28	14.15	13.48	16.11	13.19	**5.08**	12.38
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	Mean
RM	**51.74**	55.69	**56.63**	**68.60**	44.53	18.91	**49.35**
G-n	48.81	**56.22**	56.28	67.25	**44.95**	**19.73**	48.87
G-t	48.45	52.19	53.87	63.38	41.98	18.33	46.37
GJR-t	50.33	50.96	53.80	60.69	41.93	18.40	46.02
GJR-AL	44.97	50.70	48.93	59.53	39.23	18.01	43.56
GJR-ALCP	45.09	50.59	48.83	59.38	39.16	18.06	43.52

Note: Boxes indicate that the lowest loss function in that market, bold indicates that highest loss function in that market.

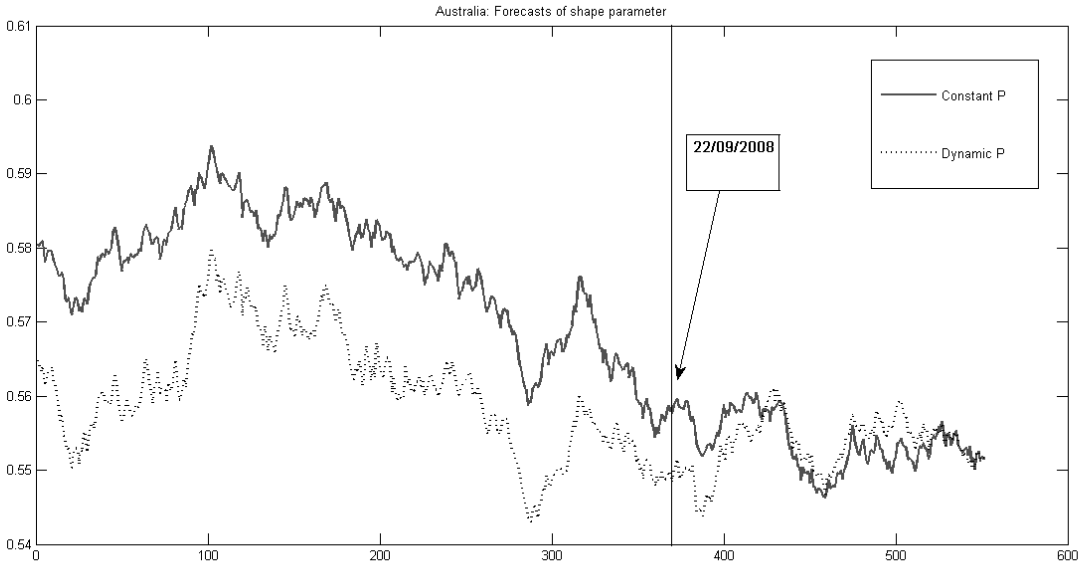
at $\alpha = 0.01$ and often at $\alpha = 0.05$. Further, the dynamics of risk are also best captured by the AL-based models, with the least rejections for the CC and DQ tests across models. The popular RM and GARCH-n models significantly and repeatedly under-estimated risk levels and are rejected in most markets by both the UC and CC tests: these models could not capture either the level or dynamics in the time period studied, for either VaR or ES. However, the use of Student t errors has improved the GARCH models: the GARCH-t and GJR-t models produced VaR violation rates much closer to 1 in average and squared error than the RM and GARCH-n models. The GARCH-t and GJR-t are not rejected in most markets by the UC test; however, they have struggled to capture risk dynamics and are rejected in two, three, or four markets via the CC and DQ tests in each case. These models also under-estimate risk and frequently produce average violation ratios above 1, often significantly so for ES risk levels. The only consistently acceptable models for ES forecasting are the GJR-GARCH-AL models, whose ES forecast performance outperforms the other methods in the two-year forecast sample. Surprisingly, the performance of each model is mostly similar before, during and after the financial crisis. In other words, all the models adapt in a similar way to the financial crisis, where risk levels have increased markedly, in a similar way. The models with poor performance in the pre-crisis period,

end up estimating and forecasting risk with similar poor and over or under estimation performance in the post-crisis period; meanwhile, the models with good performance in the pre-crisis period again provide appropriate risk estimation and forecast in the post-crisis period. The models are further assessed by MRC, AD and loss function methods. The results are combined with the results from VRates, ESRates and formal tests, and tells a consistent story.

The dynamic and fixed-shape GJR-AL models mostly perform in a similar manner across all cases. The dynamic model is marginally favoured in the formal testing results, though the VRates and ESRates are mostly the same across these two models. Figure 13 shows the estimated shape parameter during the forecast period for the Australian market and the AU/US exchange rate using the constant and dynamic shape GJR-AL models. Naturally, since parameters are re-estimated for each observation, using the last 2000 days, for each day in the forecast sample, the 'constant' shape estimates also change. These plots show that the estimates of p change in a highly similar manner whether 'constant' or dynamic. Two sources of difference are highlighted: the dynamic estimate may have a (slightly) different level to the 'constant' estimate: this difference is more marked in the pre-crisis period and diminished during and after the crisis (the difference must tend to 0 towards the end of the data). Second, the dynamic estimate can exhibit intermittent short-term outliers, likely due to either influential return observations on those days or non-converging MC chains. The latter is negated by the consistent estimate results from repeated MC samplings for those forecasting period with outliers. These are the two aspects that have led to the minor differences between these two models in their risk forecasting. We favour the dynamic model for two reasons: (i) the results in Table 6 and Table 13 marginally favour this model, with less rejections overall; and (ii) it is clear, e.g. in Figure 13, that the data favour a changing shape parameter; when a constant p is assumed, the 'constant' estimates change in a highly similar dynamic manner to those in the dynamic model. Thus, to understand how the shape changes dynamically, especially from an in-sample perspective, the dynamic model is required.

Finally, on the issue of whether skewness is required for these data sets, Figure 14 shows the estimated fixed p for each day in the GJR-ALCP model, as well as a 95%

Australian stock market.



The exchange rate of AU dollar to US dollar.

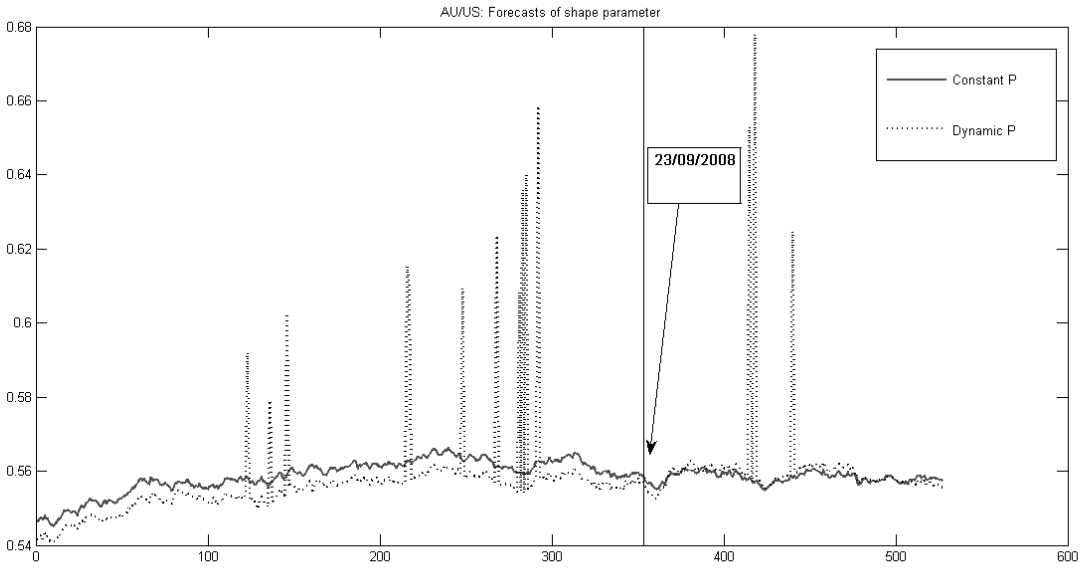


Figure 13: Chapter 2: Forecasts of constant and dynamic shape parameter from the GJR-AL model.

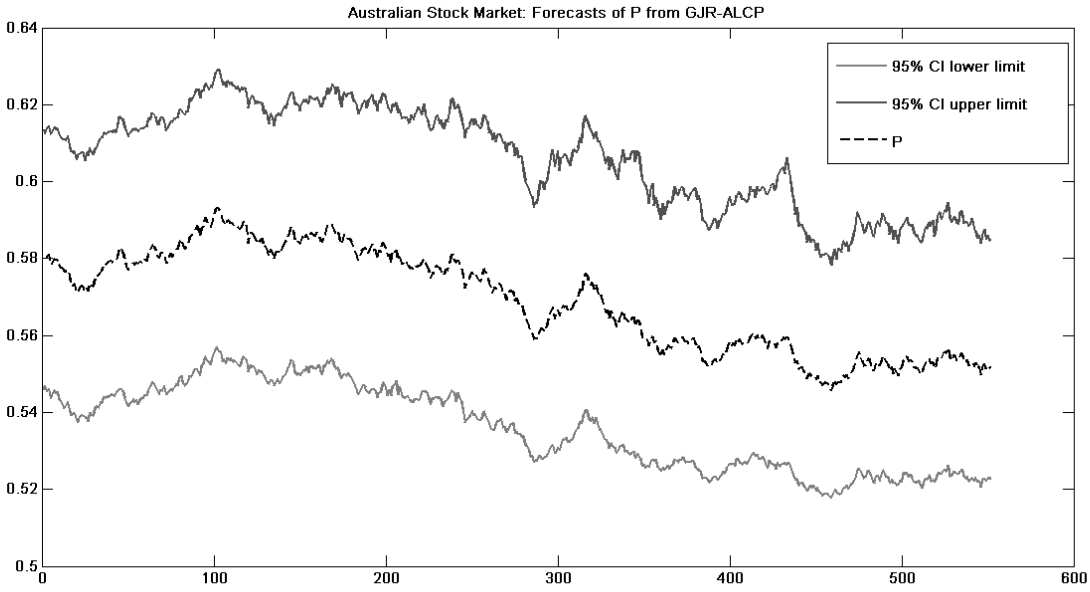


Figure 14: Chapter 2: Dynamic shape estimates from GJR-AL in the Australian market from May 1999 to May 2009, plus 95% credible intervals.

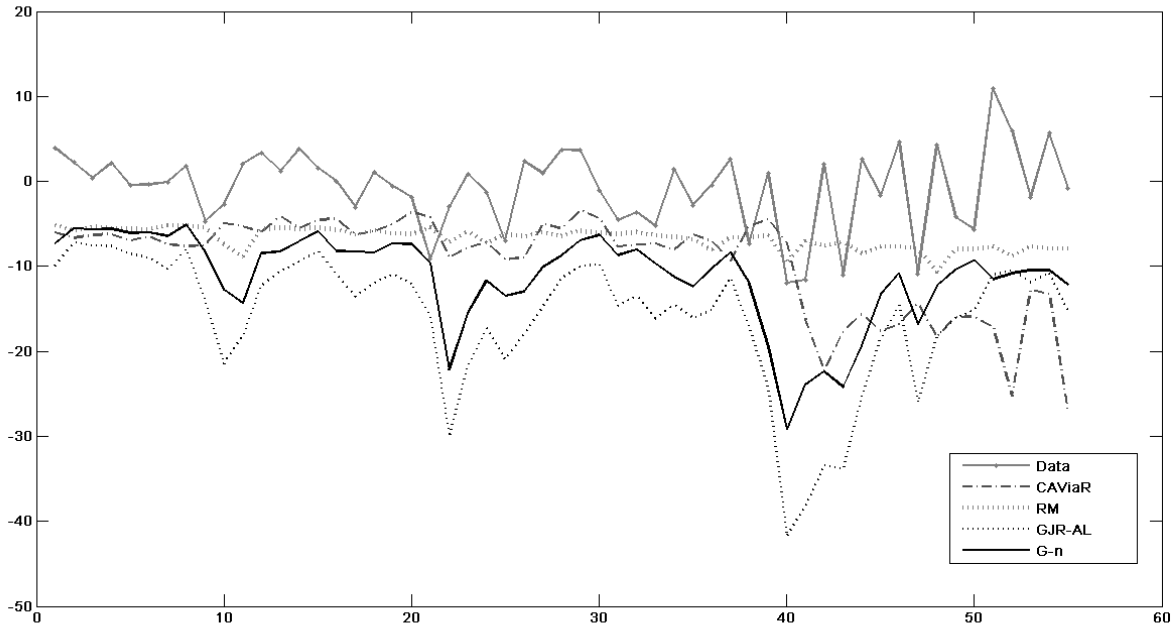
credible interval, for each day in the Australian market forecast sample. Every day's estimate of p is significantly different to $p = 0.5$; thus, skewness seems important in this market. Similar results are found in the other markets, except for HK and EUR/US.

2.4.7 A brief discussion of ten-step-ahead risk forecasts

Figure 15 shows the plots of 1% VaR and ES forecast over 10-day horizon from CAViaR, RM, G-n, G-t and GJR-AL models, for the Australian market. Similar to 1-step-ahead VaR and ES forecast, RM model fails to react to the drastic downward changes of returns, and largely under-estimates risk. The CAViaR responds to the changes more quickly, however fails to forecast adequate size of the risk. On the other hand, the GJR-AL model provide most conservative forecast of risk, G-n and G-t models rapidly respond to the changes and adequately forecast the size of the risk, which is smaller then the GJR-AL yet greater than the RM and CAViaR models.

Table 23 summarizes the ratios of $\hat{\alpha}/\alpha$ for 10-day-ahead VaR for both 1% and 5% risk levels. As there are merely 50-55 observations in the forecasting period for non-overlapping 10-day-ahead forecasts, there should be only 0 or 1 violation at 1% risk level

Australia stock market: 1% 10-day VaR



Australia stock market: 1% 10-day ES.

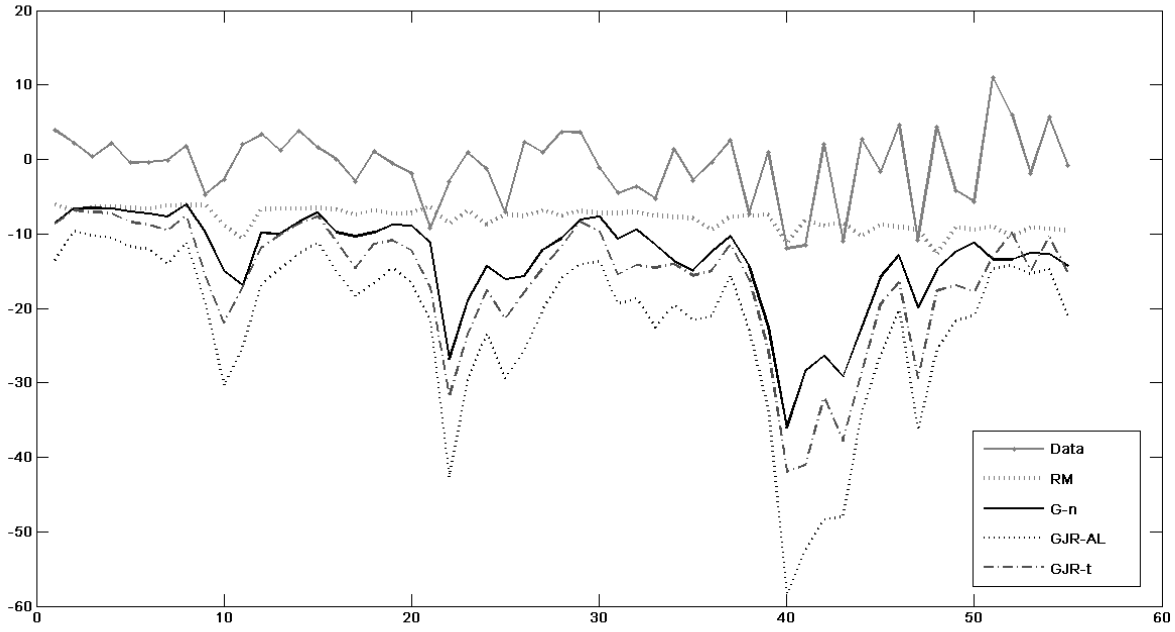


Figure 15: Chapter 2: 1% VaR and ES forecast over 10-day horizon from various in the Australian market.

and 2 to 3 violations at 5% level. Hence, at 1% level, it is acceptable that GJR-AL has no violation in all markets, while G-n, G-t, GJR-t, GJR-AL models have zero violations in 4 to 5 out of 6 markets; and one additional violation may considerably increase the violation ratio. At 1%, the G-n models ranks highest, with average ratio closest to 1. GJR-AL model ranks second, with minimum stand deviation from desired ratio 1. RM model significantly under-estimates 10-day-ahead risk and ranks least favoured, rejected by UC test in 4 out of 6 market, and having average ratio furthest from desired ration 1. The CAViaR model is risky as well, with rejection by UC test in 3 out of 6 markets. The G-t, GJR-t and GJR-ALCP models performed similarly well and rank behind the G-n and GJR-AL models.

At 5% level, G-n model clearly outperforms the other model, with average ratio closest to the desired ratio 1 and minimum standard deviation from 1. G-t closely ranks second best model. GJR-t, GJR-AL and GJR-ALCP models again are equivalently performed, ranking after G-n model. The RM model still significantly under-estimate the risk over 10-day horizon, with worst performance in average ratio and standard deviation. CAViaR model does no better than the RM model in forecasting VaR, rejected by UC test in 3 out of 6 markets.

Overall, G-n is the most favoured model for 10-day-ahead VaR forecast, and outperforms the other models, including the GJR-AL model. This is expected, as the conditional distribution of 10-day returns have much less fat tails than that of daily returns. In this case, G-n model may fairly well capture the volatility and the quantile dynamics in heteroskedastic returns. However, noted that the GJR-t, GJR-AL and GJR-ALCP models are usually more conservative with smaller violation ratio in each market than the G-n and G-t models.

The conditional distribution of 10-day returns is of unknown form, subsequently the actual quantile levels at which the ES locates are also unknown. The ratios of $\hat{\delta_\alpha}/\delta_\alpha$ are therefore unable to be calculated, and the formal tests are not applicable. However, since the conditional distribution of 10-day returns has less fat tails than those of daily returns, the lower bound of possible range for δ_α should not be smaller than the minimum of the values in Table 11 for relevant risk levels. The upper bound of possible range for δ_α is

Table 23: Chapter 2: Ratios of $\hat{\alpha}/\alpha$ for 10-day-ahead VaR at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	Mean	Std.
CAViaR	**5.45**	0.00	3.77	4.00	**5.77**	**7.55**	4.42	4.15
RM	**12.73**	1.89	3.77	**6.00**	**7.69**	**7.55**	**6.60**	**6.57**
G-n	0.00	0.00	0.00	0.00	1.92	3.77	0.95	1.45
G-t	0.00	0.00	0.00	0.00	3.85	3.77	1.27	1.82
GJR-t	0.00	0.00	0.00	0.00	1.92	0.00	0.32	0.99
GJR-AL	0.00	0.00	0.00	0.00	1.92	1.89	0.63	0.97
GJR-ALCP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	Mean	Std.
CAViaR	**2.91**	**2.64**	1.30	2.40	**2.69**	1.51	2.24	1.39
RM	**4.00**	1.13	1.51	2.40	2.31	2.26	**2.27**	**1.56**
G-n	0.73	**0.00**	1.13	1.20	1.92	1.51	1.08	0.61
G-t	0.73	**0.00**	1.13	1.60	1.92	1.51	1.15	0.65
GJR-t	0.36	**0.00**	0.38	0.80	1.92	1.51	0.83	0.70
GJR-AL	0.36	**0.00**	**0.00**	0.80	1.15	1.51	0.64	0.67
GJR-ALCP	0.36	**0.00**	**0.00**	0.80	1.15	1.51	0.64	0.67

Note: Boxes indicate the ratio closest to 1 in that market, bold indicates that the model is rejected by the unconditional coverage test (at a 5% level), for each market.

assumed to be corresponding quantile levels for Gaussian distribution. Therefore, the likely range of nominal quantile levels for ES are $[0.003, 0.004]$ and $[0.018, 0.020]$ for 1% and 5% risk levels respectively.

Figure 24 shows the ESRate for 10-day-ahead ES forecast at 1% and 5% risk levels. Since the sample size is small (50 to 55 observations), and we are considering more extreme quantiles than VaR, it is normal that several models have no violation in most of the markets. The ESRates, which are far outside the likely ranges, are indicated in bold. At 1% level, GJR-AL and GJR-ALCP models are most favoured under this criterion, and the RM model frequently under-estimates the risk. The other models perform equally well, though not better than the models with AL errors. At 5% level, GJR-AL is always conservative, however, G-n and G-t models have most adequate ESRate within the possible range in 3 out of 6 markets. RM still fails to provide enough risk coverage and incurs excessive violations.

In this case, there is not as clear conclusion as for one-step-ahead risk forecast. However, this can be improved with larger samples and more evaluation methods, such as root

mean square error (RMSE) and mean absolute deviation (MAD), etc., as an interesting topic for future research. In the following chapters, one-step-ahead risk forecast for daily returns is the focus, and 10-step-ahead VaR and ES forecasts will not be included.

Table 24: Chapter 2: ESRates for 10-day-ahead ES at $\alpha = 0.05$.

$ESRate[0.003, 0.004]$	Aust	US	UK	HK	AU/US	EUR/US
RM	**0.091**	**0.019**	**0.019**	**0.060**	**0.058**	**0.038**
G-n	0.000	0.000	0.000	0.000	**0.019**	0.000
G-t	0.000	0.000	0.000	0.000	**0.019**	0.000
GJR-t	0.000	0.000	0.000	0.000	**0.019**	0.000
GJR-AL	0.000	0.000	0.000	0.000	0.000	0.000
GJR-ALCP	0.000	0.000	0.000	0.000	0.000	0.000
$ESRate[0.018, 0.020]$	Aust	US	UK	HK	AU/US	EUR/US
RM	**0.145**	0.038	**0.075**	**0.080**	**0.096**	**0.113**
G-n	0.018	0.000	0.019	0.020	0.058	0.057
G-t	0.018	0.000	0.019	0.020	0.058	0.038
GJR-t	0.000	0.000	0.000	0.000	0.038	0.057
GJR-AL	0.000	0.000	0.000	0.000	0.038	0.038
GJR-ALCP	0.000	0.000	0.000	0.000	0.058	0.038

Note: Bold indicate that the ESRate is far beyond the possible ranges in that market.

2.5 Conclusion

The recent GFC challenged market participants' abilities to provide reasonable coverage of risk. As a coherent risk measurement method, ES is able to measure the size of loss in extreme cases. Despite the benefit of this method, ES is absent in regulations such as Basel II, perhaps because the backtesting of ES is less straightforward than that of VaR.

A VaR and ES forecast model is proposed with the assumption of a conditional asymmetric Laplace distribution, to take into account the potential skewness and heavy tails in financial return series. Dynamics in higher moments are captured by allowing the shape parameter in this distribution to be time-varying. It is proven in the empirical study section that a skewness is necessary in parametric VaR and ES models.

A Bayesian computational method is designed. The mixture of Gaussian proposal densities in the MH algorithm shows improved inferential properties over a standard

Gaussian proposal or a Student t proposal. The proposed method and model are applied to forecast VaR and ES one-day-ahead for four market indices and two exchange rate series. Formal and informal tests are employed; specifically, the proposed ES backtesting method can be applied in a similar way as that for the VaR. The test results illustrate that the proposed model outperforms, or is at least highly competitive with, several popular alternatives and is the only consistently conservative risk model throughout the period studied.

3 The Two-sided Weibull Distribution and Forecasting Financial Tail Risk

The need for more general GARCH type models and more flexible non-Gaussian conditional return distributions on modelling highly volatile financial series has increased in the past decades, as indicated in Mittnik and Paolella (2000). This chapter expands the scope of the study in Chapter 2.

The aims of this chapter are to:

1. Develop a more general distribution to capture the potential skewness and heavy tails for the conditional return distribution.

2. Develop a more flexible model to accurately forecast the financial risk levels, and provide adequate risk coverage, but slightly less conservative than the GJR-GARCH-AL model proposed in Chapter 2.

3. Examine the impacts of asymmetric volatility specification and asymmetric conditional distribution on the VaR and ES forecasting.

Hence, a two-sided Weibull distribution is proposed to model the conditional return distribution. Gaussian errors, Student t errors, skewed Student t (Hansen, 1994) errors, AL errors, and two-sided Weibull errors are combined with four GARCH-type volatility specifications to forecast the tail risk in real financial return series. For comparison, the ES and corresponding quantile level function for ES of the skewed Student t are also derived. The four volatility specifications include a symmetric GARCH model, and three non-linear GARCH models. The parameter estimation and forecasting is via the adaptive MCMC scheme in Chapter 2. These models are applied to seven return series and produce one-day-ahead VaR and ES forecasts for a four-year forecasting period, covering the recent global financial crisis. The forecasting performance is evaluated by formal and informal backtesting methods. The proposed models with two-sided Weibull errors outperforms all the models with other error distributions, including the AL distribution, under various criteria. The models with two-sided Weibull conditional return distribution provide the

most accurate risk assessment, both before, during and after the recent global financial crisis.

The contributions in this chapter lie in both the proposed distribution and the findings in the empirical study. The first four moments, cdf, inverse cdf, formula for VaR and ES as well as the quantile level function for ES of the TW distribution are derived. On the other hand, an essential finding in the empirical studies is that, the forecast of risk levels is influenced more by the choice of conditional return distribution than the choice of GARCH specifications.

A shortened version of the material in this chapter has been submitted for publication at the Annals of Applied Statistics, for details see "The Two-sided Weibull Distribution and Forecasting Financial Tail Risk" Chen and Gerlach (2011) at `http://sydney.edu.au/business/ome/research/working_papers`.

3.1 A Two-sided Weibull Distribution

The Weibull distribution, introduced by Weibull (1951), is a special case of an extreme value distribution and of the generalised gamma distribution. It is widely applied in the fields of material science, engineering, and also in finance, due to its versatility. Mittnik and Ratchev (1989) found it to be the most accurate for the unconditional return distribution for the S&P500 index when applied separately to positive and negative returns; while various authors have employed it as an error distribution in range data modelling (see Chen et al., 2008) and trading duration (ACD) models (see e.g. Engle and Russell, 1998).

Sornette et al. (2000) proposed and used a symmetric modified (two-sided) Weibull distribution as an unconditional return distribution, later combined with a Gaussian copula, to choose efficient portfolios in Malevergne and Sornette (2004); they also mentioned a two-sided Weibull but did not explore its properties. We introduce a similar, though more flexible, transformed Weibull, called the two-sided Weibull distribution (TW). The motivation for this is to capture empirical traits in conditional return distributions such as fat-tails and skewness for the purposes of risk measure forecasting: thus, the tails are the

most important regions to be accurately modelled. The idea, as in Mittnik and Ratchev (1989) and Malevergne and Sornette (2004), is to allow a different Weibull distribution for positive and negative returns. This also sets up a flexible extension of the AL distribution in Chen et al (2011) and in Chapter 2, where a different exponential was allowed for positive and negative returns; i.e. if $X \sim \text{Exp}(\lambda)$ then $X^k \sim \text{Weibull}(\lambda, k)$.

Since a conditional error in a GARCH-type model needs to have mean 0 and variance 1, we further develop the standardised two-sided Weibull distribtion (STW). We subsequently derive the pdf, cdf, quantile function, and the conditional expectation functions required to calculate the likelihood as well as VaR and ES measures for the STW distribution.

The TW's shape and scale can be tuned by four Weibull parameters. The definition of a TW is, $Y \sim TW(\lambda_1, k_1, \lambda_2, k_2)$ if:

$$\begin{cases} -Y \sim \text{Weibull}(\lambda_1, k_1) \ ; \ Y < 0 \\ Y \sim \text{Weibull}(\lambda_2, k_2) \ ; \ Y \geq 0 \end{cases}$$

where the shape parameters satisfy $k_1, k_2 > 0$ and scale parameters $\lambda_1, \lambda_2 > 0$.

3.1.1 Standardised Two-sided Weibull distribution

Error distributions in volatility models should be standardised. A standardised TW distribution is equivalent to $\frac{Y}{\sqrt{\text{Var}(Y)}}$. For a TW, it can be shown that:

$$\text{Var}(Y) = b_p^2 = \frac{\lambda_1^3}{k_1}\Gamma\left(1+\frac{2}{k_1}\right) + \frac{\lambda_2^3}{k_2}\Gamma\left(1+\frac{2}{k_2}\right) - \left[-\frac{\lambda_1^2}{k_1}\Gamma\left(1+\frac{1}{k_1}\right) + \frac{\lambda_2^2}{k_2}\Gamma\left(1+\frac{1}{k_2}\right)\right]^2.$$

The pdf for an STW random variable $X = \frac{Y}{\sqrt{\text{Var}(Y)}}$, where $Y \sim TW(\lambda_1, k_1, \lambda_2, k_2)$, is:

$$f(x|\lambda_1, k_1, k_2) = \begin{cases} b_p \left(\frac{-b_p x}{\lambda_1}\right)^{k_1-1} \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right] \ ; \ x < 0 \\ b_p \left(\frac{b_p x}{\lambda_2}\right)^{k_2-1} \exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right] \ ; \ x \geq 0. \end{cases} \tag{22}$$

To ensure the pdf integrates to 1:

$$\frac{\lambda_1}{k_1} + \frac{\lambda_2}{k_2} = 1. \tag{23}$$

Thus, in this formulation there are only three free parameters, and we write $X \sim STW(\lambda_1, k_1, k_2)$ where λ_2 is fixed by (23). In this parameterization $Pr(X < 0) = \frac{\lambda_1}{k_1}$;

thus, if $\frac{\lambda_1}{k_1} < 0.5$, the density is positively skewed to the right, while negative or left skewness occurs when $\frac{\lambda_1}{k_1} > 0.5$. The $STW(\lambda_1, k_1, \lambda_2)$ has cdf, obtained by direct integration,

$$F(x|\lambda_1, k_1, k_2) = \begin{cases} \frac{\lambda_1}{k_1} \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right] \ ; \ x < 0 \\ 1 - \frac{\lambda_2}{k_2} \exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right] \ ; \ x \geq 0. \end{cases} \tag{24}$$

The inverse cdf or quantile function of an STW is:

$$F^{-1}(\alpha|\lambda_1, k_1, k_2) = \begin{cases} -\frac{\lambda_1}{b_p}\left[-\ln\left(\frac{k_1}{\lambda_1}\alpha\right)\right]^{\frac{1}{k_1}} \ ; \ 0 \leq \alpha < \frac{\lambda_1}{k_1} \\ \frac{\lambda_2}{b_p}\left[-\ln\left(\frac{k_2}{\lambda_2}(1-\alpha)\right)\right]^{\frac{1}{k_2}} \ ; \ \frac{\lambda_1}{k_1} \leq \alpha < 1. \end{cases} \tag{25}$$

The derivation of the first four moments of an STW, including the mean μ_X, are given in Appendix 2. Thus $Z = X - \mu_X$ has a shifted $STW(\lambda_1, k_1, k_2)$ distribution with mean 0 and variance 1.

For the purposes of parsimony and simplification, and since we notice that real return data supports this choice, we consider only the case $k_1 = k_2$. Setting $k_1 = k_2$ means we can write simply $STW(\lambda_1, k_1)$ with only two parameters to estimate. As $Pr(X < 0) = \frac{\lambda_1}{k_1}$, thus $0 < \lambda_1 \leq k_1$, and $\lambda_2 = k_1 - \lambda_1$. Chen et al. (2011) and Chapter 2 considered the AL distribution, whose skewness ranged from $[-2, 2]$ and kurtosis ranged from $[6, 9]$. When $k_1 = k_2$, the range of skewness in the STW is $[-2.4, 2.4]$ and kurtosis is $[2.5, 11.5]$, as shown in Figure 16. If we remove the constraints that $k_1 = k_2$, even wider ranges for skewness and kurtosis can be obtained. Since $[-2.4, 2.4]$ and $[2.5, 11.5]$ are flexible enough for the potential skewness and kurtosis of the real return series, we stay withe setting $k_1 = k_2$, which illustrates the increased flexibility.

Malevergne and Sornette (2004) considered only the case $k_1 \leq 1$, which preserves a single mode of the density. However, the tails of the STW density become fatter as $k_1 < 1$ compared to $k = 1$, which is the AL distribution that Chen et al. (2011) and Chapter 2 found already too fat-tailed during the GFC period. As such, we do not restrict k_1 in estimation. This will allow the conditional distribution, potentially, to be multi-modal, which may not be a good fit to the data in the centre of the distribution; however, it will potentially allow the tails, and thus VaR and ES, to be estimated more accurately. This result is confirmed in the empirical section to come.

Figure 17 shows some STW densities and log-densities for the range of parameter

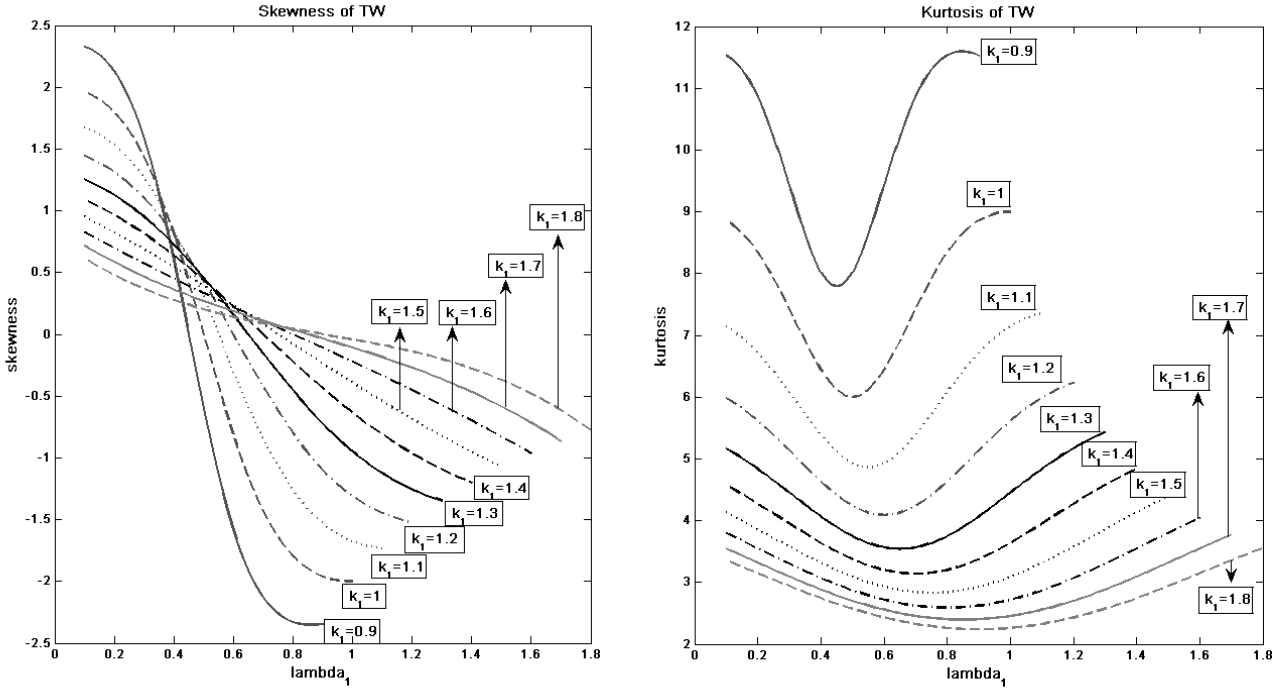
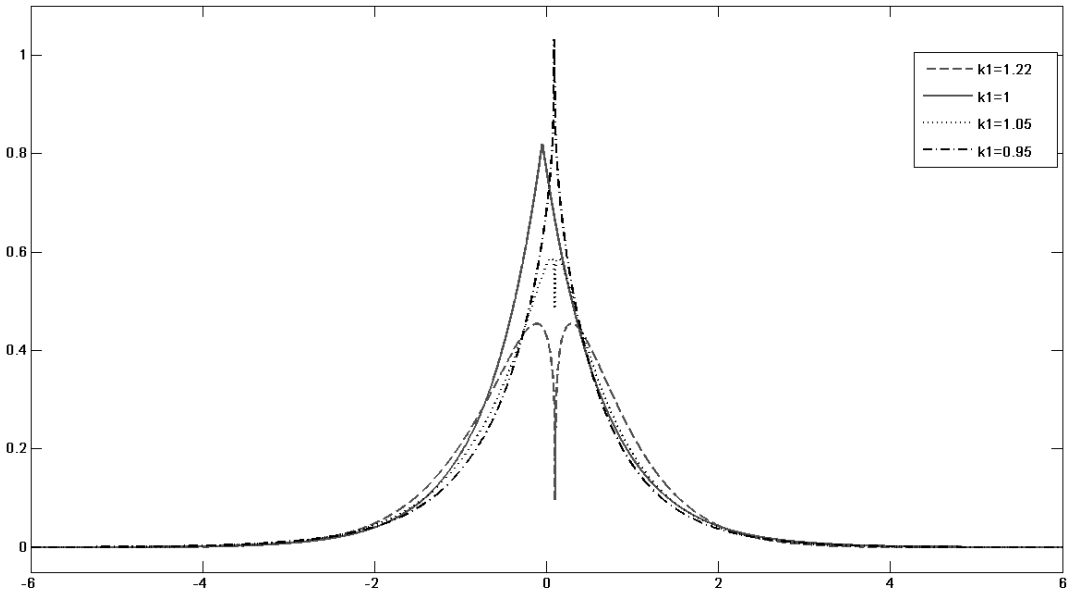


Figure 16: Chapter 3: Ranges for skewness and kurtosis from TW distribution with $k_1 = k_2$.

STW densities



STW log-densities

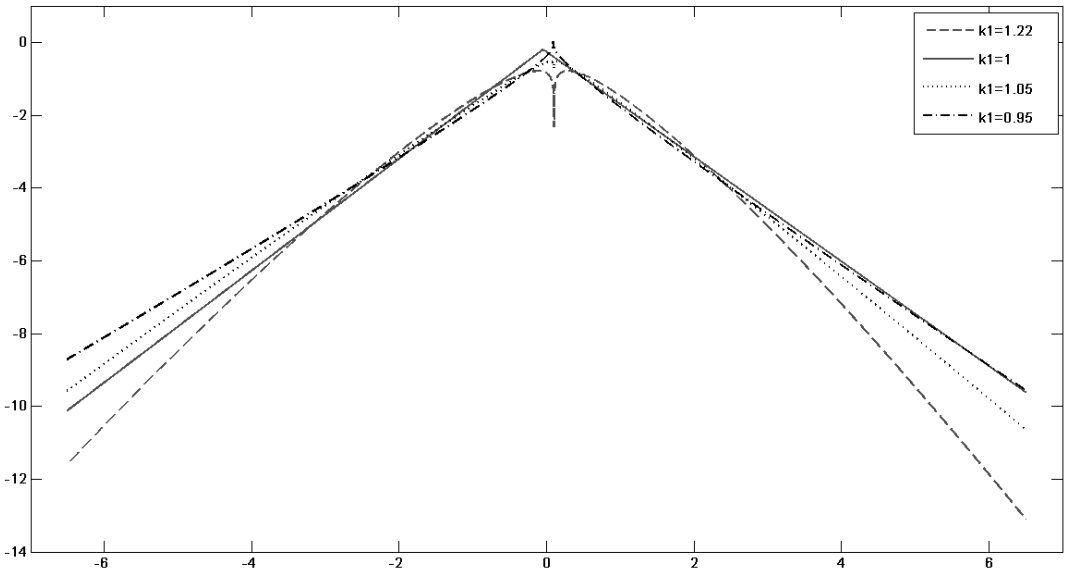


Figure 17: Chapter 3: Some standardised two-sided Weibull densities.

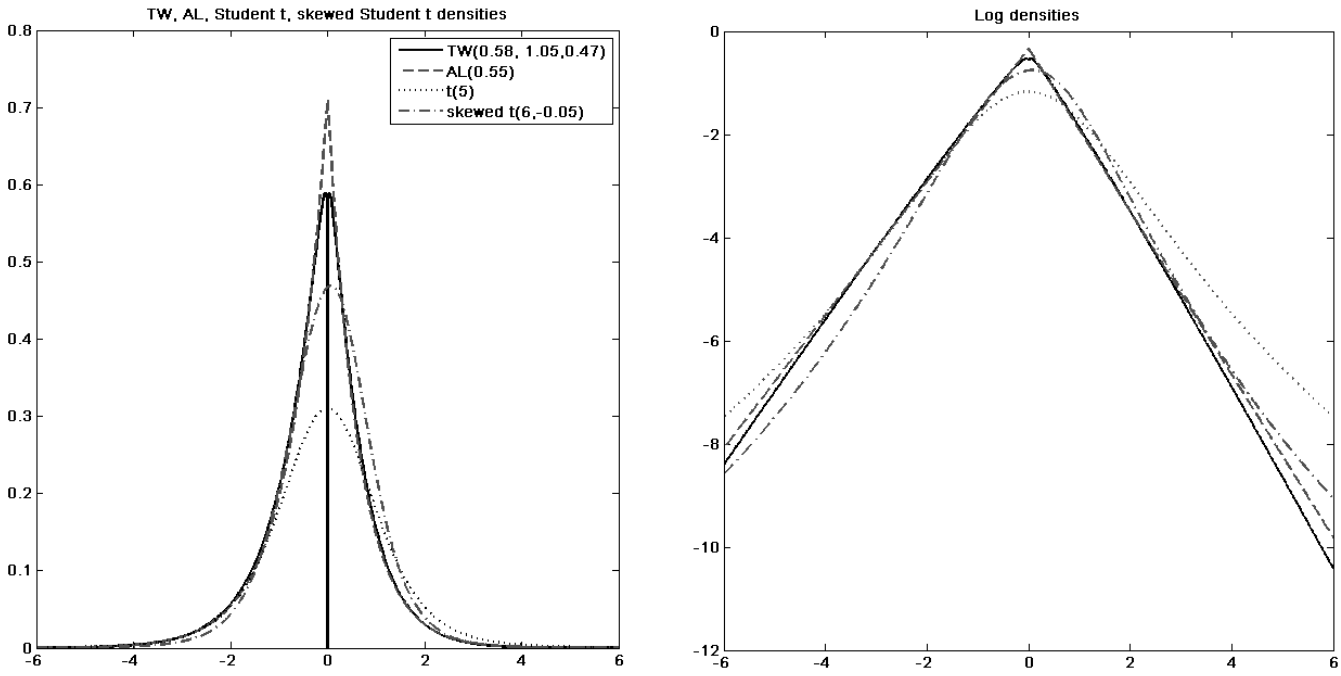


Figure 18: Chapter 3: STW, AL, skewed Student t and Student t densities and log densities.

estimates found for k_1 in the real return series we analyse (i.e. $k_1 \in (1, 1.22)$), as well as $k_1 = 0.95$; the skewness is kept constant in each density; the STW distribution's flexibility is demonstrated, as is the slight thinning of the tails as $k_1 > 1$.

Figure 18 demonstrates the STW, AL, and skewed Student t densities with similar skewness; the Student t with 6 degrees of freedom is also shown.

3.1.2 VaR and tail conditional expectations for the two-sided Weibull

The VaR and ES are formally defined by (3) and (4) in Chapter 2. The VaR is thus simply the quantile given in (25) for the STW distribution.

In practice, the return series is usually lightly skewed, therefore, the estimated values for $\frac{\lambda_1}{k_1}$ is usually close to 0.5. Since risk management focuses on only the extreme tails of returns, particularly the cases of $\alpha \leq 0.05$ for the long position, thus only the case $\alpha < \frac{\lambda_1}{k_1}$ in (25) is relevant here. In this context, the tail expectation of an STW for the long position is:

$$\begin{aligned}
\mathrm{ES}_\alpha &= \frac{-\lambda_1^2}{\alpha b_p k_1}\int_{\left(\frac{-b_p VaR_\alpha}{\lambda_1}\right)^{k_1}}^{\infty}\left[\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right]^{\frac{1}{k_1}+1-1}\exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right]d\left(\frac{-b_p x}{\lambda_1}\right)^{k_1} \\
&= \frac{-\lambda_1^2}{\alpha b_p k_1}\Gamma\left(1+\frac{1}{k_1},\left(\frac{-b_p VaR_\alpha}{\lambda_1}\right)^{k_1}\right)\ ;\ 0\le\alpha<\frac{\lambda_1}{k_1}. \qquad (26)
\end{aligned}$$

Similarly, the tail expection of an STW for the short position is:

$$\begin{aligned}
\mathrm{ES}_\alpha &= \frac{\lambda_2^2}{(1-\alpha)\,b_p k_1}\int_{\left(\frac{b_p VaR_{1-\alpha}}{\lambda_2}\right)^{k_2}}^{\infty}\left[\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right]^{\frac{1}{k_2}+1-1}\exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right]d\left(\frac{b_p x}{\lambda_2}\right)^{k_2} \\
&= \frac{\lambda_2^2}{(1-\alpha)\,b_p k_2}\Gamma\left(1+\frac{1}{k_2},\left(\frac{b_p VaR_{1-\alpha}}{\lambda_2}\right)^{k_2}\right)\ ;\ \frac{\lambda_1}{k_1}\le\alpha<1. \qquad (27)
\end{aligned}$$

where $\Gamma(s,x)=\int_x^\infty t^{s-1}e^{-t}dt$ is the upper incomplete gamma function.

This work just considers the case of the long position. The function for actual quantile level of ES_α is

$$\begin{aligned}
\delta_\alpha &= \frac{\lambda_1}{k_1}\exp\left\{-\left[\frac{\lambda_1}{\alpha k_1}\Gamma\left(1+\frac{1}{k_1},\left(\frac{-b_p VaR_\alpha}{\lambda_1}\right)^{k_1}\right)\right]^{k_1}\right\}\ ; \qquad (28) \\
&= \frac{\lambda_1}{k_1}\exp\left[-\left(\frac{-b_p}{\lambda_1}\mathrm{ES}_\alpha\right)^{k_1}\right].
\end{aligned}$$

3.2 Model Specification

This section discusses the general forms for the financial return series' models considered in the empirical section. We follow the common assumption that the mean of a return series is (well-approximated as) zero. The generalized model for a financial return series $\mathbf{y}$ is:

$$y_t=(\epsilon_t-\mu_\epsilon)\sqrt{h_t}\quad,\ \epsilon_t\overset{\text{i.i.d.}}{\sim}D(1), \qquad (29)$$

where $\mathrm{Var}(y_t|\Omega_t)=h_t$ is the conditional variance, and D is the conditional distribution and has variance 1 and mean μ_ϵ (often 0). The VaR and ES in this model are:

$$\mathrm{VaR}_{t+1}=D_\alpha^{-1}\sqrt{h_{t+1}}\ ;\ \mathrm{ES}_{t+1}=\mathrm{ES}_\alpha^D\sqrt{h_{t+1}}. \qquad (30)$$

where D_α^{-1} is the inverse cdf of D, and ES_α^D is the expected shortfall of D, at the $\alpha \times 100\%$ level. The Gaussian, Student t, Skewed t of Hansen (1994), the AL of Chen et al. (2011) and in Chapter 2, and the STW distribution are considered. The latter two have non-zero means that are subtracted in (29). Expressions for ES_α^D in the Gaussian and Student-t cases can be found in McNeil, Frey, and Embrechts (2005, pg. 45, 46), while for the AL see Chen et al. (2011) and Chapter 2. Appendix 4 contains a derivation of ES_α^D for the Skewed t distribution.

3.2.1 Volatility models

The most general volatility model is a two-regime smooth-transition GARCH model (denoted as ST-GARCH), with a logistic smooth-transition function as in Gerlach and Chen (2008). As the data considered are observed daily, such a smooth change between regimes is potentially more reasonable than a sharp regime transition, as in a T-GARCH, though both will be considered and compared.

The specified ST-GARCH model has volatility dynamics:

$$\begin{aligned} h_t &= h_t^{[1]} + G(x_{t-1}; \iota, r) h_t^{[2]}, \\ h_t^{[i]} &= \alpha_0^{[i]} + \alpha_1^{[i]} y_{t-1}^2 + \beta_1^{[i]} h_{t-1}, i = 1, 2 \ , \end{aligned} \tag{31}$$

and thus represents a continuous mixture of two regimes: where $h_t^{[2]}$ is the difference of conditional variances between the two regimes. $G(x_{t-1}; \gamma, r)$ is a function defined on $[0, 1]$: we take a logistic as standard:

$$G(x_{t-1}; \gamma, r) = \frac{1}{1 + \exp\left\{-\gamma\left(\frac{x_{t-1}-r}{s_x}\right)\right\}},$$

where γ is the smoothness or speed of transition parameter, assumed positive for identification; s_x is the sample standard deviation of the observed threshold variable $\mathbf{x}$, allowing γ to be independent of the scale of $\mathbf{x}$. In this study, we choose the one-period lag returns as the threshold variable X.

The T-GARCH model is a special case of (31), where $\gamma \to \infty$. Further, the GJR-GARCH is then a special case of the T-GARCH, where $x_{t-1} = y_{t-1}, r = 0$, and $\alpha_0^{[2]} =$

$\beta_1^{[2]} = 0$, and $G(y_{t-1}|\gamma = \infty, r = 0) = 1$ when $y_{t-1} < 0$ and 0 otherwise. The symmetric GARCH model has constant $G(y_{t-1}|\gamma, r = -\infty) = 0$, so there is only one regime.

The standard sufficient second order stationary and positivity constraints are:

$$\begin{aligned}
\alpha_0^{[1]} > 0 \ ; \ 0 \leq \alpha_1^{[1]} + \beta_1^{[1]} < 1 \ ; \alpha_1^{[1]}, \beta_1^{[1]} \geq 0; \\
0 \leq \alpha_1^{[1]} + 0.5\alpha_1^{[2]} + \beta_1^{[1]} + 0.5\beta_1^{[2]} < 1; \\
\alpha_0^{[1]} + \alpha_0^{[2]} > 0 \ ; \ \alpha_1^{[1]} + \alpha_1^{[2]}, \beta_1^{[1]} + \beta_1^{[2]} \geq 0; \\
0 \leq \alpha_1^{[1]} + \alpha_1^{[2]} + \beta_1^{[1]} + \beta_1^{[2]} < 1,
\end{aligned} \tag{32}$$

which apply whenever $\exp(G(\cdot)) = 0.5$ and D is symmetric. Chen et al (2011) and Appendix 1 derived expressions for the second-order stationarity condition c_p, which replaces 0.5 in these expressions in the case of the GJR-GARCH model for the AL distribution. Appendix 3 contains derivations of the extensions of these expressions to the case of the GJR-GARCH model with STW errors. Expressions are not known for the T-GARCH or ST-GARCH models in general. However, we note that, practically in daily financial returns, the values of c_p are very close to 0.5 from the right side, indicating that (32) is approximately sufficient for stationarity most of the time.

Tong, Xia, and Li (2007) use a semi-parametric method to determine the threshold variable in a TAR model. In this study, to allow maximum uncertainty of the parameter, the threshold variable is estimated and inferenced by Bayesian methods as are other parameters in the model.

3.3 Estimation and Forecasting Methodology

This section specifies the Bayesian methods and MCMC procedures for estimating parameters and generating forecasts.

3.3.1 Bayesian estimation methods

In a Bayesian analysis, a likelihood function and a prior are usually required. The required likelihood follows from the choice of error distribution D and equation (29) together with a volatility equation (31). We consider the priors for the most general ST-GARCH model

with STW errors.

The ST-GARCH parameters in each regime are grouped and denoted $\boldsymbol{\theta}_{[\mathbf{1}]}$, $\boldsymbol{\theta}_{[\mathbf{2}]}$ and each group is generated separately in the MCMC scheme. Let $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_{[\mathbf{1}]}, \boldsymbol{\theta}_{[\mathbf{2}]}\right)$, the prior is

$$\begin{aligned}
\pi(\boldsymbol{\theta}_{[\mathbf{1}]}) &\propto I\left(0 < \alpha_0^{[1]} < s_y^2, \alpha_1^{[1]} + \beta_1^{[1]} < 1, \alpha_1^{[1]}, \beta_1^{[1]} \geq 0\right) ; \\
\pi(\boldsymbol{\theta}_{[\mathbf{2}]}|\boldsymbol{\theta}_{[\mathbf{1}]}) &\propto I\left(\begin{array}{l} -\alpha_0^{[1]} < \alpha_0^{[2]} < s_y^2 - \alpha_0^{[1]}, 0.5(\alpha_1^{[2]} + \beta_1^{[2]}) < 1 - (\alpha_1^{[1]} + \beta_1^{[1]}), \\ \alpha_0^{[2]} \geq -\alpha_1^{[2]}, \beta_1^{[2]} \geq -\beta_1^{[1]}, -(\alpha_1^{[1]} + \beta_1^{[1]}) \leq \alpha_1^{[2]} + \beta_1^{[2]} < 1 - (\alpha_1^{[2]} + \beta_1^{[2]}) \end{array}\right),
\end{aligned}$$

where s_y^2 is the sample variance of the return data. This prior ensures that (32) are satisfied and that the volatility intercepts are suitably bounded.

For the threshold value r a constrained uniform is applied, as standard, i.e. $\pi(r) \propto I\left(ll \leq r \leq ul\right)$; where ll and ul are the 10th and 90th percentiles of the return series, which is the threshold variable in this study, to ensure sufficient observations for identification and inference in each regime. The prior for the speed of transition parameter γ is:

$$\pi(\gamma) \propto I\left(\xi_1 < \gamma < \xi_2\right);$$

where values for ξ_1, ξ_2 can be user-defined. Here, ξ_1, ξ_2 are chosen as finite, and not close to 0 or ∞ to ensure proper posterior. The limits on the transition function is:

$$\begin{aligned}
& 0.01 \leq G\left(x_{t-1}; \gamma, r\right) \leq 0.99; \\
\Rightarrow \quad & -\log(99) \leq \gamma \tfrac{x_{t-1}-r}{s_y} \leq \log(99); \\
\Rightarrow \quad & -\log(99) \leq \gamma \tfrac{\min(x_{t-1})-r}{s_y} \leq \gamma \tfrac{x_{t-1}-r}{s_y} \leq \log(99); \\
\Rightarrow \quad & -\tfrac{s_y \log(99)}{\min(\boldsymbol{x})-r} \leq \gamma \leq \tfrac{s_y \log(99)}{\min(\boldsymbol{x})-r}.
\end{aligned}$$

Since when $\gamma > 20$, $G(\cdot)$ is very close to 1, that is, the ST-GARCH becomes a T-GARCH, we have

$$\pi(\gamma) \propto I\left(-\frac{s_y \log(99)}{\min(\boldsymbol{x}) - r} \leq \gamma \leq 20\right) ;$$

similar to that suggested in Chen, Gerlach, Choy, and Lin (2010), which together with the bounded prior on r ensures that the parameter γ does not get too close to 0, in which case $\boldsymbol{\theta}_{[\mathbf{1}]}$ and $\boldsymbol{\theta}_{[\mathbf{2}]}$ are not identified, since $G = 0.5$ is constant in that case, while also not

allowing $\gamma \to \infty$. The prior effectively ensures that the function G is below 0.01 at the minimum value of the threshold $\boldsymbol{x}$ and thus not constant over the range.

For the STW distribution the parameters λ_1 and k_1 have a flat prior:

$$\pi(\lambda_1) \propto I\left(0 < \lambda_1 < k_1\right) \ ;$$

The AL distribution has $k_1 = 1$ and the same prior on $\lambda_1 = p$. For the Skewed t distribution the degrees of freedom and shape parameters, respectively, ν and ζ, have:

$$\pi(\nu) \propto I\left(4 < \nu < 30\right) \ ; \pi(\zeta) \propto I\left(-1 < \zeta < 1\right) \ ;$$

which are the ranges for their likely values. Although Hansen (1994) suggested $2 < \nu < \infty$, as the Student t distribution becomes very close to normal distribution when $\nu > 30$, thus the upper limit for the likely range for ν is set to be 30. On the other hand, when $\nu = 4$, the skewed Student t distribution is fat tailed enough, and ν is ensured well above 2 to ensure a proper posterior.

None of the parameter groupings have a standard recognisable conditional posterior density and as such Metropolis and Metropolis-Hastings methods are required. Gerlach and Chen (2008) illustrated the efficiency and speed of mixing gains from employing an adaptive scheme where iterates in the burn-in period, simulated from standard random-walk Metropolis methods with tuning to achieve desired acceptance rates, are used to build a Gaussian proposal density for use in the sampling period. Chen et al. (2011) and Chapter 2 extended this method to cover a mixture of Gaussian proposals, both in the burn-in and sampling periods. This method is adapted to estimate the model parameters in this chapter. This mixture of Gaussian method is a special simplified case of the more general and flexible "AdMit" mixture of the Student t proposal procedure proposed by Hoogerheide, Kaashoek, and van Dijk (2007).

Convergence is examined for by running the MCMC scheme from multiple and wide-ranging starting points and checking trace plots of iterates for convergence to the same posterior.

3.3.1.1 Simulation Results

We demonstrate the above parameter estimation methods and sampling schemes via simulation study of the TGARCH-STW model.

Sample series of size $n = 2000$, the same size as the real data we consider in this work, are simulated from the following model:

$$\begin{aligned} \epsilon_t &\overset{\text{i.i.d.}}{\sim} TW(0.53, 1.05, 1.05), \\ \sigma_t^2 &= \begin{cases} 0.05 + 0.05a_{t-1}^2 + 0.85\sigma_{t-1}^2 \ ; \ x_{t-1} < 0 \\ 0.01 + 0.02a_{t-1}^2 + 0.7\sigma_{t-1}^2 \ ; \ 0 \leq x_{t-1} \ , \end{cases} \\ \textit{and} & \\ y_t &= (\epsilon_t - \mu_{\epsilon_t})\sigma_t. \end{aligned}$$

The simulation results are shown in Table 25. The estimates for the parameters are very accurate, with an average close to the true values and high coverage rates. Figure 19 shows the MCMC plots in the burn-in period with six different starting positions for parameter λ_1 in the TGARCH-STW model. It is clear that the adaptive MCMC sampling scheme converges quickly. For TGARCH-STW, a typical MCMC running time for $n = 2000$ for one sample in the Matlab is 3 minutes; the typical MH acceptance rates are between 29% and 37% in the burn-in period, and between 26% and 54% in the sampling period.

3.3.2 VaR and ES forecasts

One-step-ahead forecasting is considered. This chapter's focus is the impact of the asymmetry in the volatility and in the shape of the conditional return distributions on the risk level forecasting, thus only the parametric models is considered, and semi-parametric CAViaR model and non-parametric historical model are dropped. The GARCH family in (31) provides one-step-ahead forecasts of volatility based on known parameter values. In MCMC methods, at each stage the entire parameter vector, denoted $\boldsymbol{\theta}$, has values simulated for it from the posterior, combining to give a Monte Carlo sample $\boldsymbol{\theta}^{[1]}, \ldots, \boldsymbol{\theta}^{[N]}$,

Table 25: Summary statistics for parameter estimates from 100 simulated data sets from the TGARCH-TW model.

par.	True	Mean	Standard error	Coverage
λ_1	0.53	0.532	0.0115	100
k_1	1.05	1.061	0.012	100
r	0	0.000	0.0113	96
α_0^1	0.05	0.061	0.0126	92
α_1^1	0.05	0.068	0.0299	94
α_2^1	0.85	0.780	0.0845	96
α_0^2	0.01	0.016	0.0075	93
α_1^2	0.02	0.032	0.016	95
α_2^2	0.7	0.636	0.0685	90

where N is the MC sample size. Each of these iterates provides a one-step-ahead forecast of h_t, which can be combined with (30) via e.g. (25) and (26) for STW errors, to give MC iterate forecasts of VaR and ES, i.e. $\text{VaR}^{[i]}$, $\text{ES}^{[i]}$ for $i = 1, \ldots, N$, for each model. These are simply averaged over the iterates in the sampling period of the MCMC scheme, to give a one-step-ahead forecast of VaR and ES for each model.

All the model's estimation and forecasting are realized in Matlab. I wrote the code for all the models except GARCH-n, GARCH-t, GJR-GARCH-n, and GJR-GARCH-t, which are estimated via the GARCH Toolbox in Matlab.

In this chapter, we again employ the backtesting methods in Chapter 2 for evaluating the VaR and ES models. Chen et al. (2011) and Chapter 2 exploited the ES quantile levels for the Gaussian and AL distributions, available in Table 10 in Chapter 2, to employ the standard VaR backtesting methods to back-test ES models. For the Student t and Skewed t, however, the quantile level of ES depends on α, plus the degrees of freedom ν and λ for the Skewed t. Similarly, for the STW, the ES quantile level, derived in (28), depends on the parameters λ_1, k_1. To backtest ES models with these distributions, we approximate by considering the relevant ES level for the average estimated parameters during the forecast sample. This works well, since these parameters do not change very much during the forecast period.

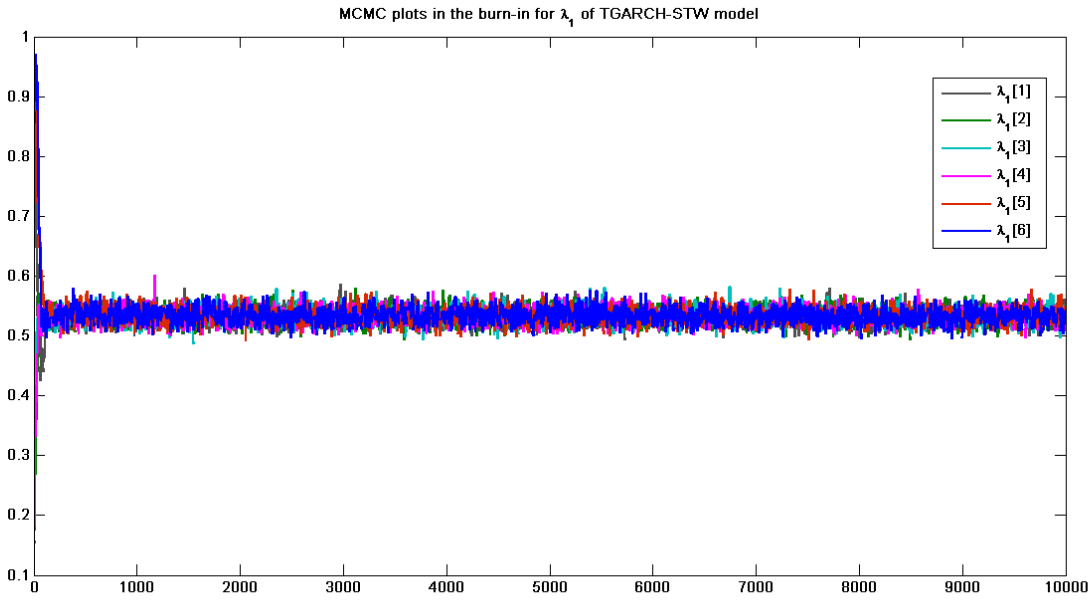


Figure 19: Chapter 3: MCMC plots in the burn-in period with six different starting positions for λ_1 in the TGARCH-TW model.

3.4 Empirical Study

3.4.1 Data

The model is illustrated by applying it to daily return series from four international stock-market indices: the S&P 500 (US); FTSE 100 (UK); AORD All Ordinaries Index (Australia); the HANG SENG Index (Hong Kong); plus two exchange rate series: the AU dollar to the US dollar and the Euro to the US dollar; as well as one single asset series: IBM. The data are obtained from Yahoo! Finance, covering twelve years, January 1998 to January 2010, except the exchange rate of the Euro to US dollar, which starts from January 1999. The daily percentage log return series is $y_t = (ln(P_t) - ln(P_{t-1})) \times 100$, where P_t is the closing price/value on day t.

The sample is initially divided into two periods: the period from January 1998 to December 2005, roughly the first 2000 returns, is used as an initial learning period. The data from January 2006 to January 2010 are used as the forecasting period. The forecast sample sizes vary from 770 to 1050 days, due to different trading day holidays, etc., and

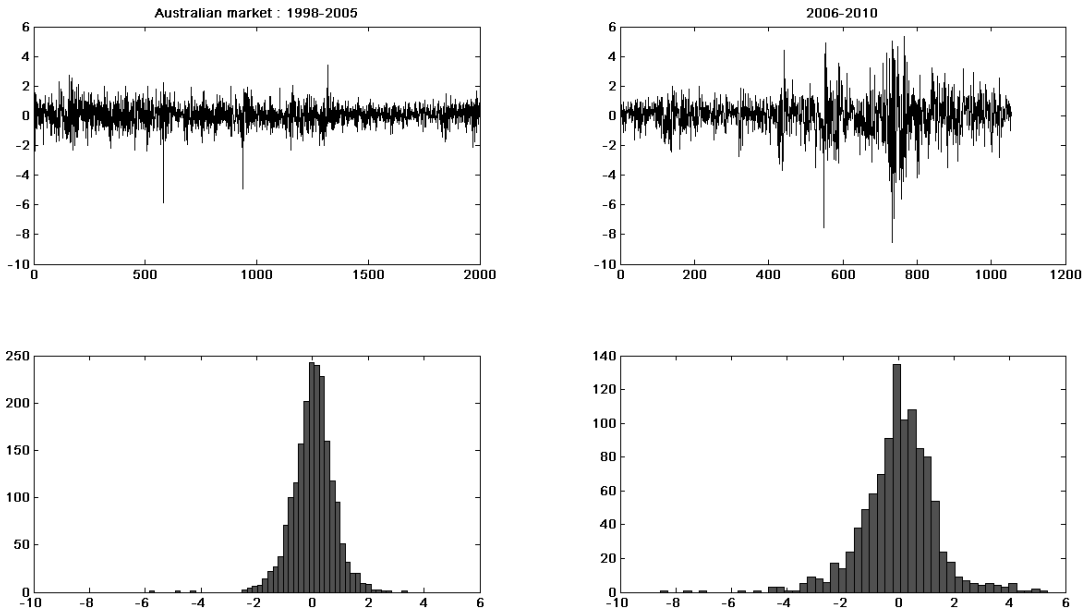


Figure 20: Chapter 3: Plots and histograms of the Australian market for both the initial learning period and the forecasting period.

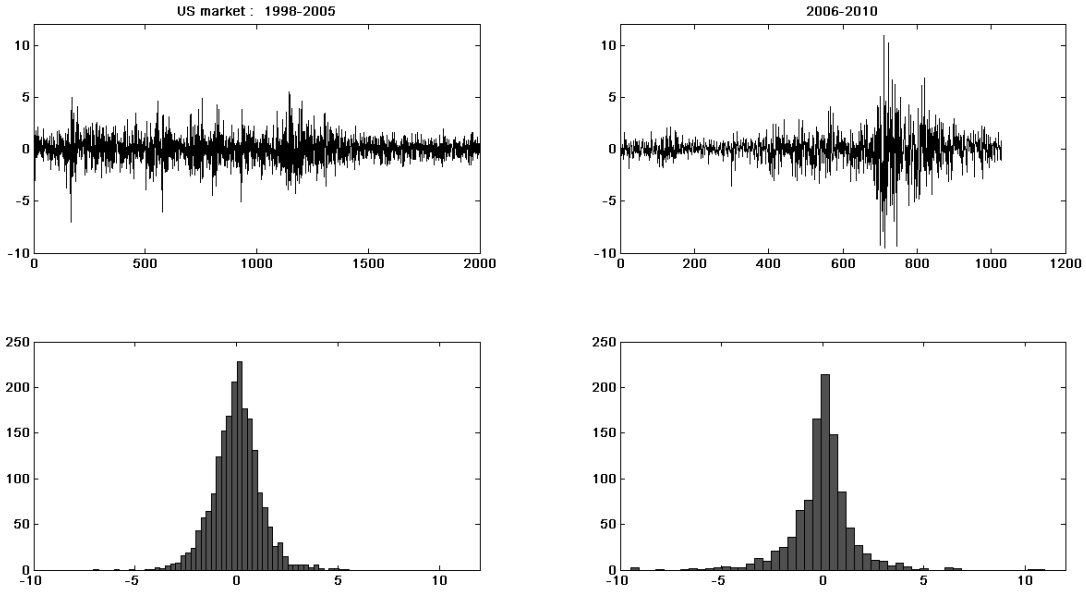


Figure 21: Chapter 3: Plots and histograms of the US market for both the initial learning period and the forecasting period.

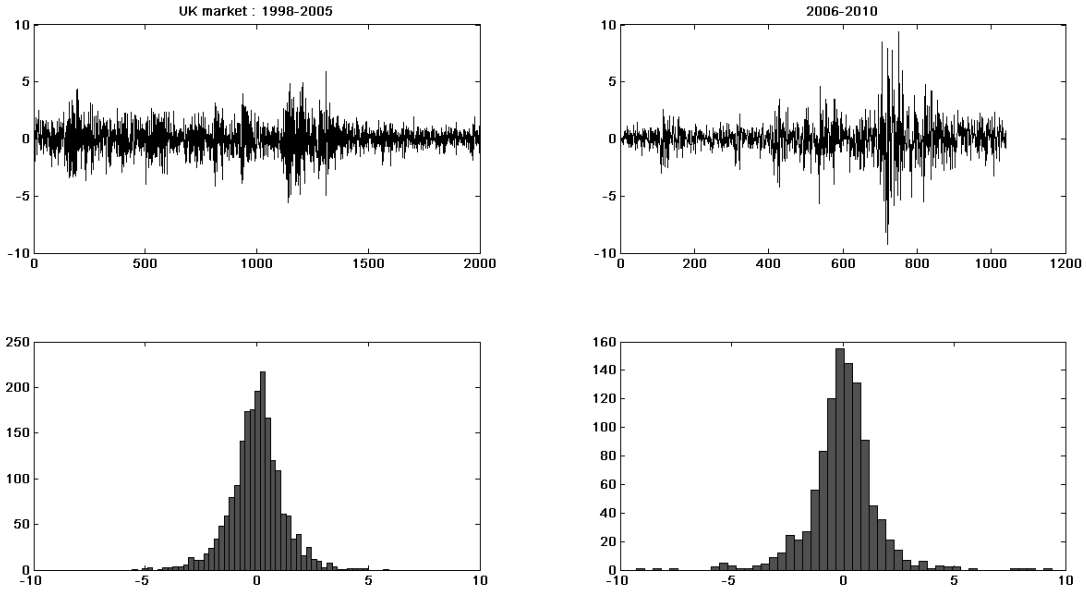


Figure 22: Chapter 3: Plots and histograms of the UK market for both the initial learning period and the forecasting period.

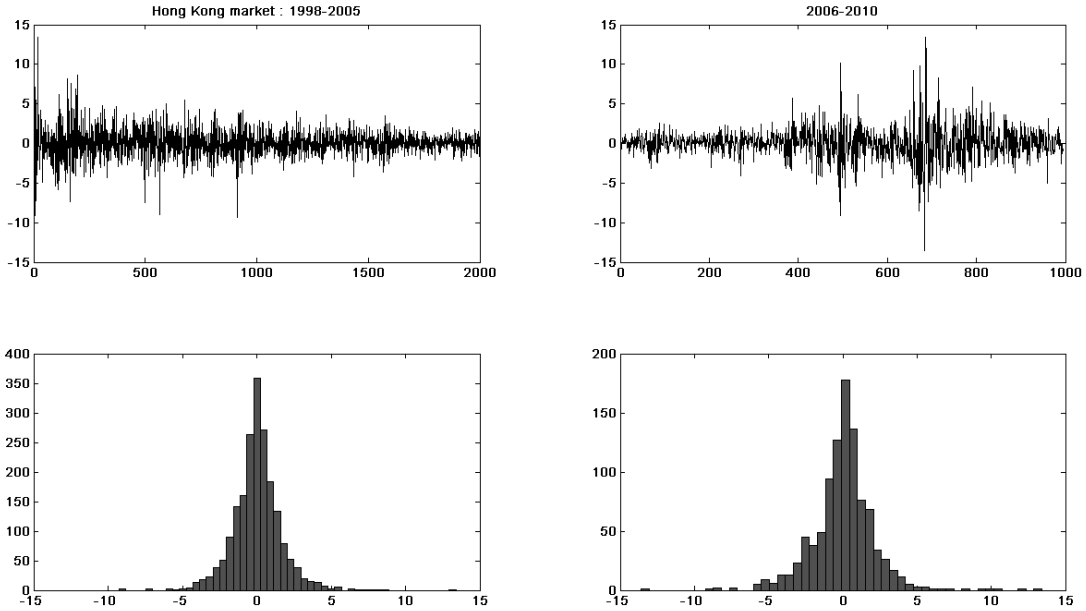


Figure 23: Chapter 3: Plots and histograms of the HK market for both the initial learning period and the forecasting period.

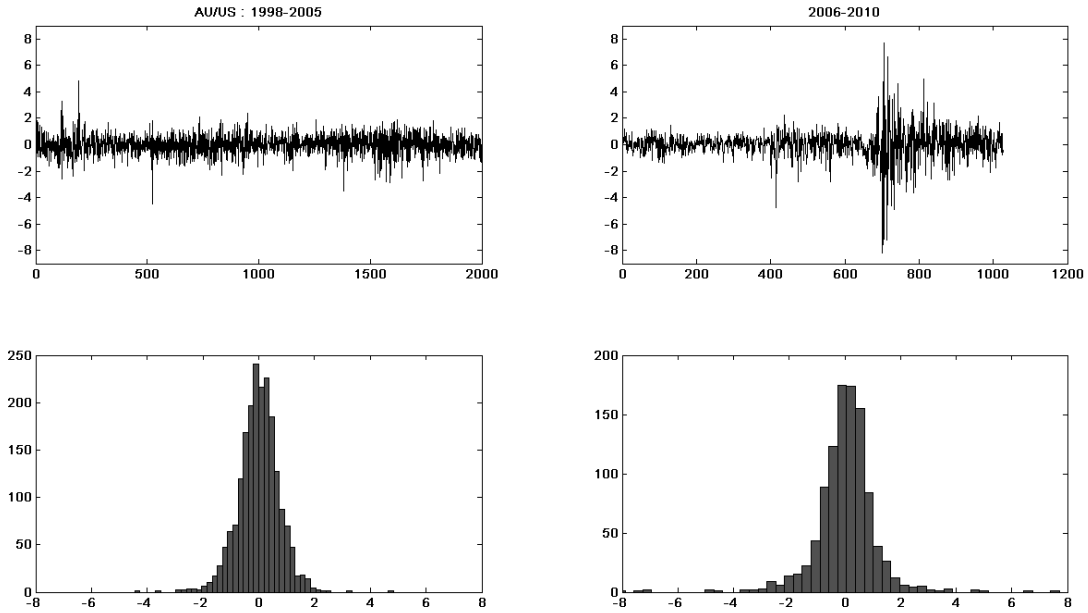


Figure 24: Chapter 3: Plots and histograms of AU/US for both the initial learning period and the forecasting period.

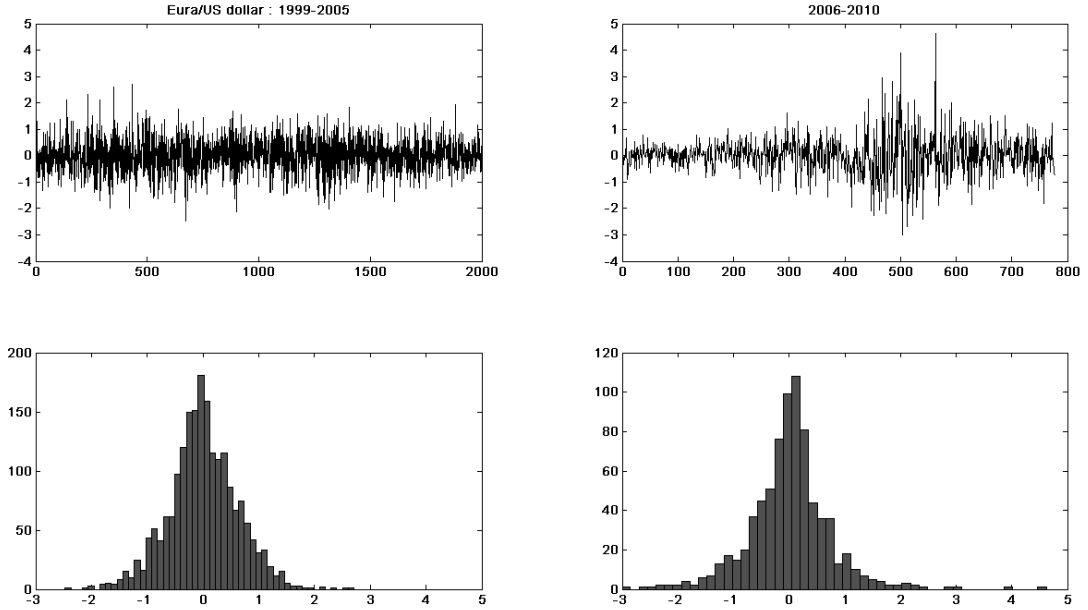


Figure 25: Chapter 3: Plots and histograms of Euro/US for both the initial learning period and the forecasting period.

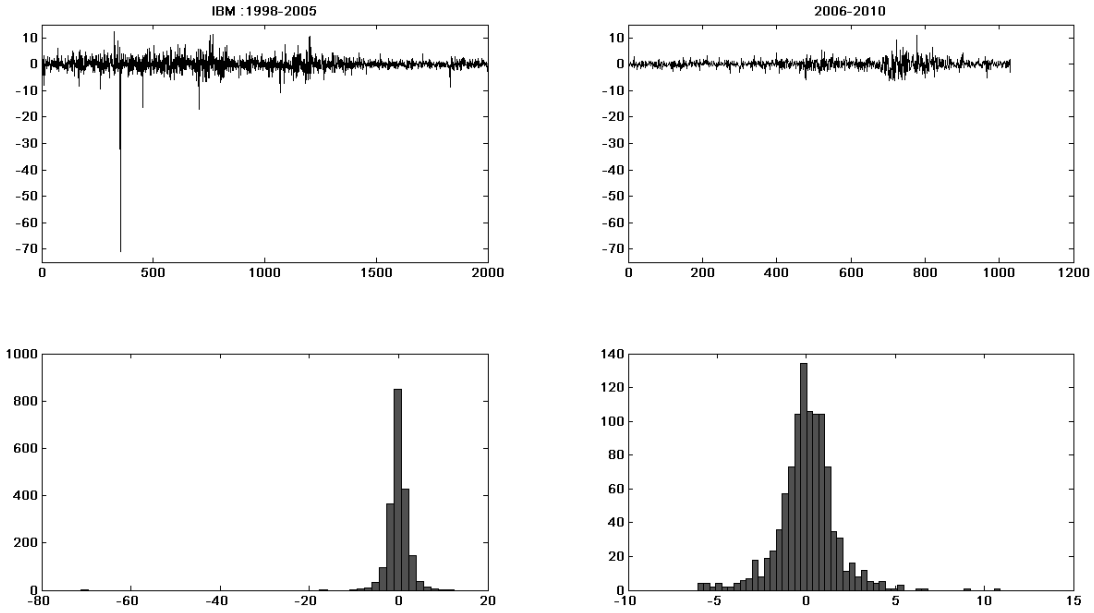


Figure 26: Chapter 3: Plots and histograms of IBM for both the initial learning period and the forecasting period.

this period completely contains the GFC. Table 26 shows summary statistics for the seven return series in the learning and forecast samples. Clearly, the forecast period is mostly more volatile and more fat-tailed (higher kurtosis), except notably for IBM, as also shown in Figures 20 to 26. The estimation results in each series are mostly as expected and well-documented in the literature: high volatility persistence ($\alpha_1 + \beta_1$); fat-tailed (e.g. $\nu < 10$ in Student t and Skewed t error models) and mildly negatively skewed (e.g. $\lambda_1/k_1 > 0.5$ in STW, $p > 0.5$ in AL and $\lambda < 0$ in Skewed t) conditional distributions.

The initial learning period parameter estimates and standard errors (in brackets) from the TGARCH-STW model are given in Table 27.

Figure 27 shows the estimated shape parameters, that is $\frac{\lambda_1}{k_1}$ and p for TGARCH-TW and TGARCH-AL, respectively, in the forecasting period.

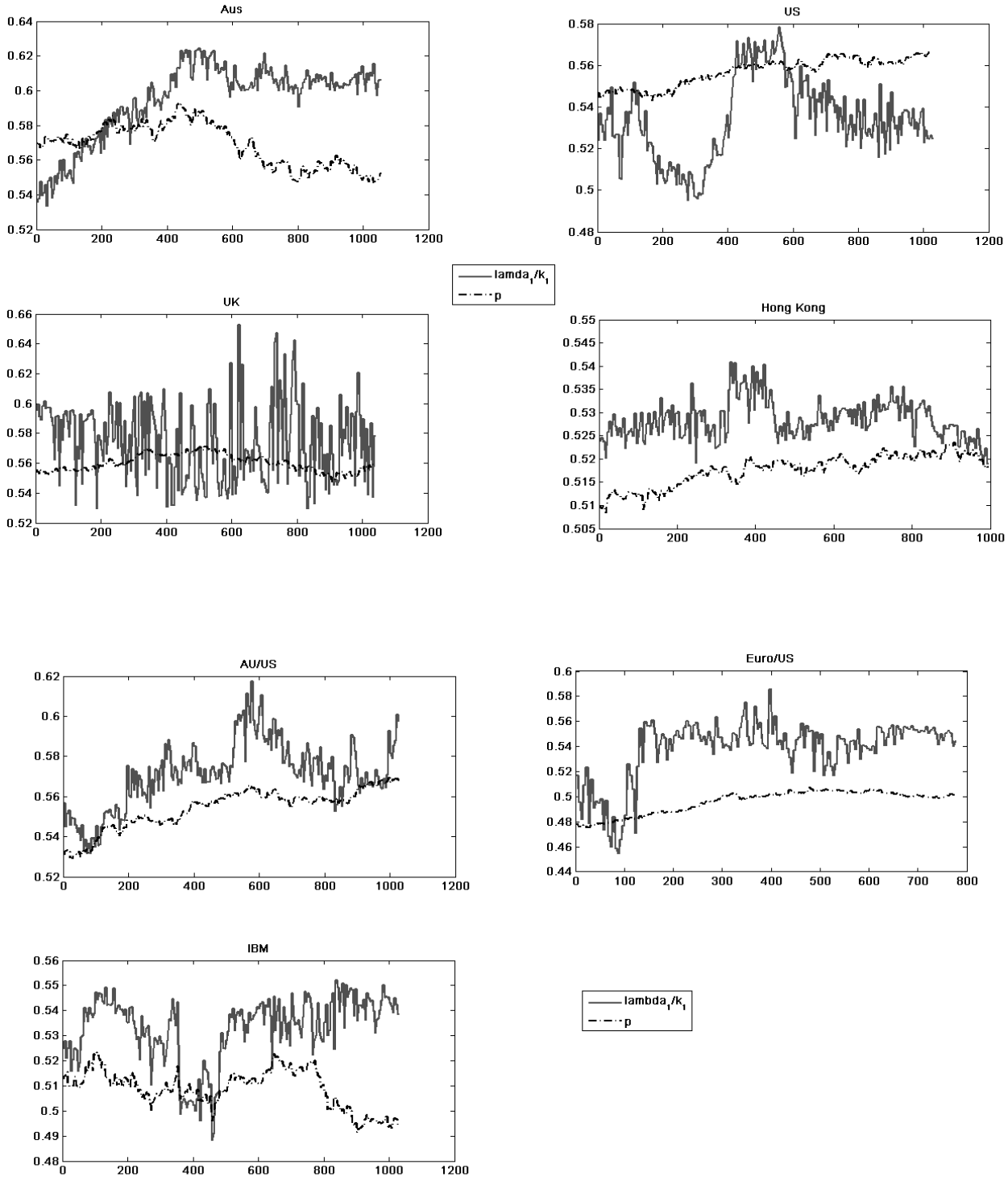


Figure 27: Chapter 3: Estimated $\frac{\lambda_1}{k_1}$ and p for TGARCH-TW and TGARCH-AL from January 2006 to January 2010.

Table 26: Chapter 3: Summary statistics of seven return series.

Index	Period	Mean	Std	Skewness	Kurtosis	Min	Max
Aus	98-05	0.028	0.73	-0.53	7.12	-5.85	3.39
	06-10	0.007	1.36	-0.54	7.23	-8.55	5.36
US	98-05	0.013	1.20	0.00	5.36	-7.04	5.57
	06-10	-0.011	1.65	-0.22	11.44	-9.47	10.96
UK	98-05	0.003	1.20	-0.11	5.19	-5.59	5.90
	06-10	-0.001	1.53	-0.10	10.00	-9.26	9.38
HK	98-05	0.019	1.64	0.20	8.64	-9.29	13.40
	06-10	0.033	2.13	0.09	9.26	-13.58	13.41
AU/US	98-05	0.006	0.72	-0.19	5.53	-4.45	4.82
	06-10	0.022	1.13	-0.72	15.10	-8.21	7.70
EUR/US	99-05	0.006	0.61	0.02	3.72	-2.47	2.71
	06-10	0.010	0.74	0.39	7.36	-3.00	4.62
IBM	98-05	-0.012	2.677	-9.449	253.415	-71.130	12.364
	06-10	0.044	1.623	0.177	7.681	-6.102	10.899

Table 27: Chapter 3: TGARCH-TW parameter estimates and standard errors (bracketed).

	Aus	US	UK	HK	AU/US	EUR/US	IBM
λ_1	0.662	0.642	0.721	0.580	0.658	0.604	0.581
	(0.034)	(0.041)	(0.054)	(0.033)	(0.039)	(0.046)	(0.037)
k_1	1.228	1.231	1.241	1.121	1.196	1.182	1.102
	(0.032)	(0.039)	(0.031)	(0.038)	(0.040)	(0.036)	(0.030)
r	-0.355	-1.117	-0.939	-1.550	-0.059	-0.546	-2.102
	(0.151)	(0.238)	(0.343)	(0.133)	(0.578)	(0.315)	(0.051)
$\alpha_0^{[1]}$	0.065	0.169	0.122	0.147	0.033	0.040	2.370
	(0.025)	(0.098)	(0.077)	(0.098)	(0.034)	(0.034)	(0.452)
$\alpha_1^{[1]}$	0.095	0.131	0.116	0.093	0.048	0.064	0.001
	(0.026)	(0.037)	(0.026)	(0.026)	(0.016)	(0.022)	(0.001)
$\beta_1^{[1]}$	0.882	0.846	0.861	0.873	0.916	0.775	0.928
	(0.034)	(0.046)	(0.042)	(0.048)	(0.046)	(0.092)	(0.048)
$\alpha_0^{[2]}$	0.005	0.025	0.011	0.013	0.019	0.006	0.054
	(0.003)	(0.010)	(0.005)	(0.006)	(0.013)	(0.010)	(0.018)
$\alpha_1^{[2]}$	0.013	0.006	0.024	0.017	0.025	0.018	0.077
	(0.009)	(0.006)	(0.015)	(0.009)	(0.014)	(0.009)	(0.021)
$\beta_1^{[2]}$	0.920	0.927	0.931	0.957	0.932	0.967	0.892
	(0.015)	(0.013)	(0.015)	(0.011)	(0.031)	(0.024)	(0.019)

3.4.2 VaR forecast comparison

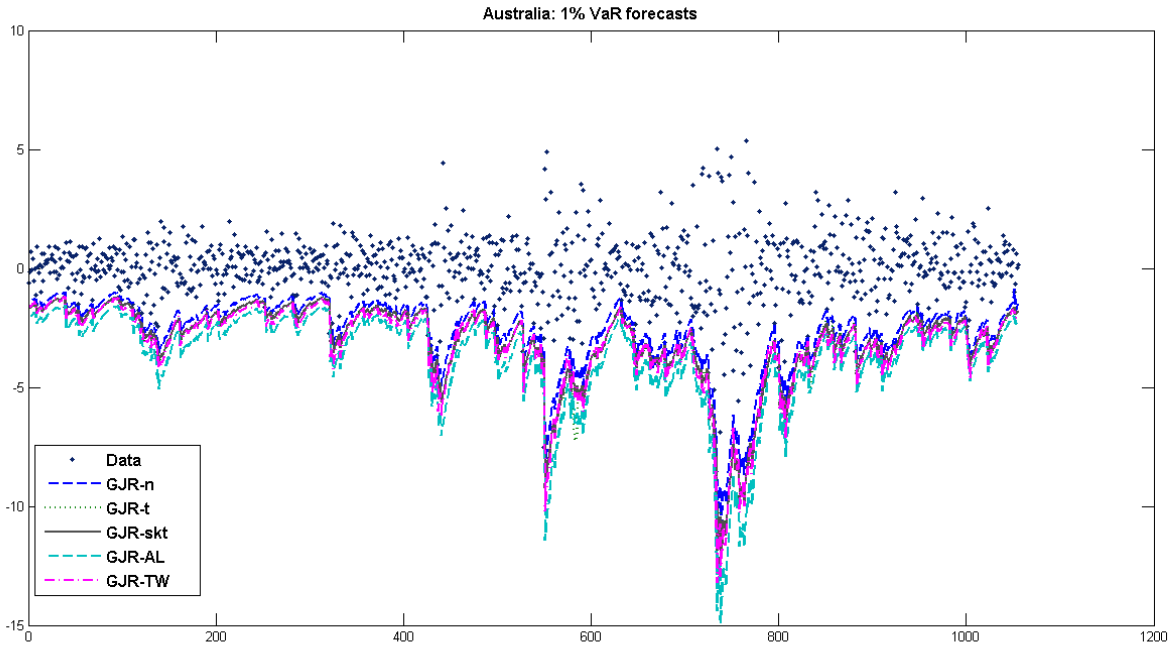
The VaR forecasts from several parametric models, for the returns on the Australian stock market and the AU to US dollar exchange rate, are shown in Figure 28.

The plots indicate a clear ordering in ES levels across distributions: the Gaussian is least extreme, followed by the Student t, Skewed t, STW, while the AL distribution gives the most extreme ES forecasts. This pattern occurred consistently across the seven series, holding the volatility model constant.

Table 28 shows the ratios, and their summaries, of observed VRates to the true nominal levels $\alpha = 0.01, 0.05$ across all series; summaries shown are average ('Mean') and deviation ('Std') for each model and series. 'Std' is the square root of the average squared distance of the observed ratio away from the expected ratio of 1. For each series, the ratio closest to 1 is boxed, while the mean ratio and deviation closest to 1 over the models, for each series, is also boxed. Violation ratios that are significantly different from 1, at a 5% significance level by the UC test, are in bold.

First, it is clear that the differences between models are dominated by the choice of error distribution: models with the same distribution but different volatility equation are much closer in terms of VRate, to each other, than they are to models with same volatility models but a different distribution. Thus, models with the same error distributions appear together in the table. As such, discussion centres on the different distributions first. At $\alpha = 0.01$, it is clear that models with Gaussian errors are always anti-conservative and under-predict risk levels in all series: on average VRates are double or more the nominal 1%. Alternatively, models with AL errors over-predict risk levels: on average VRates are half the nominal 1%, and are thus conservative; this agrees with results in Chen et al. (2011) and Chapter 2. Models with Skewed t errors tend to under-predict risk, but less so than Gaussian models, with average VRates about 20–30% too high. Models with Student t and STW errors are clearly the best performers and most favoured, with VRates close to nominal on average. The GJR-t model ranks highest with the average VRate closest to 1 (1.02), closely followed by the GARCH-STW with 1.03, which also has the minimum deviation from 1 (0.3), equal best with the ST-GARCH-STW model.

Australian market.



AU/US.

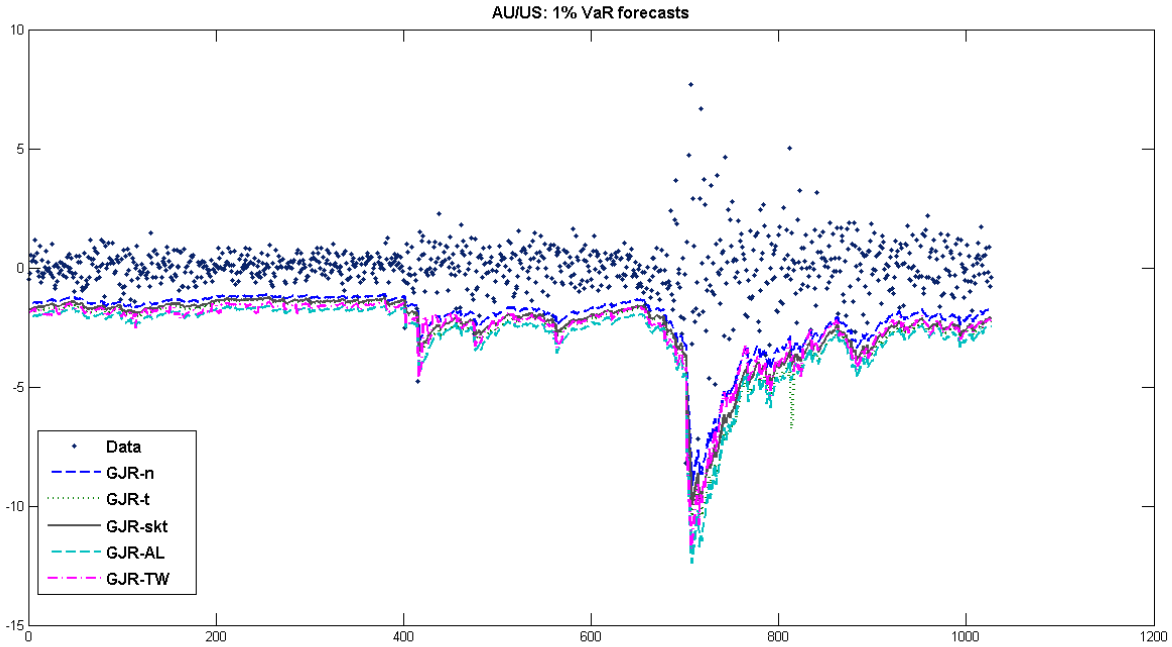


Figure 28: Chapter 3: 1% VaR forecasts from GJR-n, GJR-t, GJR-skt, GJR-ALCP, and GJR-TW.

All models with STW errors have VRate ratio deviations equal to or lower than all other models. Informally, then, models with STW errors have done best in forecasting risk levels at $\alpha = 0.01$, very marginally ahead of models with Student t errors.

Similar results hold for $\alpha = 0.05$, with Gaussian and Skewed t error models frequently under-predicting risk, while models with AL, Student t and STW errors have VRate ratios quite close to 1 across the seven series.

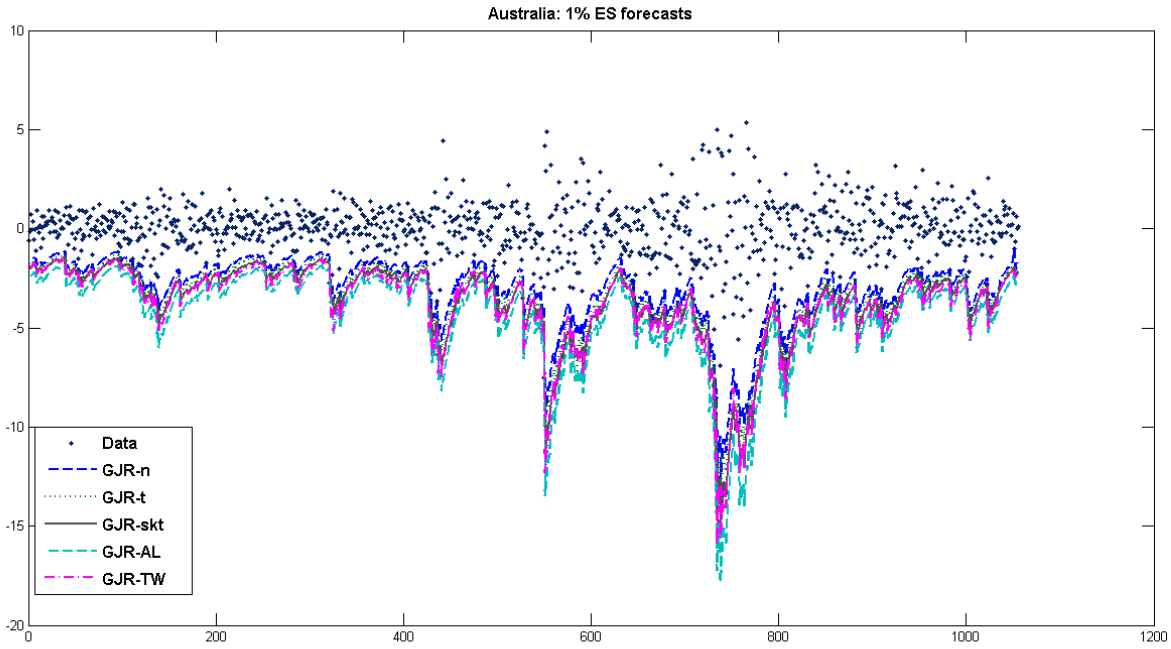
Table 29 shows counts of the number of rejections for each model, at a 5% significance level, across the seven series, under the three formal backtests: the UC, the CC, and the DQ test. Following Engle and Manganelli (2004), we chose a lag of 4 for the DQ test, while using the extended CC test in Chen et al. (2011) and in Chapter 2, also with a lag of 4. At $\alpha = 0.01$ the Gaussian error models are rejected in all or most series, while the models with Student t errors are rejected on average more than the other models. The three best models are rejected only in one series: the GJR-GARCH-STW, and the ST-GARCH and GJR-GARCH, both with Skewed t errors. Models with AL, Skewed t and STW errors are quite comparable and do the best on these tests across the seven series. At $\alpha = 0.05$, models with Gaussian errors are again rejected in most series. The other models are quite comparable, except for the GJR-GARCH-STW and T-GARCH-STW models, which are only rejected in one series each.

In summary, models with STW and Student t errors tend to have average VRates closest to nominal at both $\alpha = 0.01, 0.05$. In terms of deviation in VRate ratios from 1, again models with STW errors do best overall, though models with AL errors do very well at $\alpha = 0.05$. In terms of the tests, for both $\alpha = 0.01, 0.05$ a model with STW had the minimum number of rejections: one in seven series. Models with Gaussian errors significantly under-predicted risk in most series at $\alpha = 0.01, 0.05$ by over 100% at $\alpha = 0.01$; models with Skewed t errors, while doing reasonably well in the formal tests, under-predict risk levels by 10–30% on average.

3.4.3 Expected Shortfall forecast comparison

The ES forecasts from several parametric models, for the returns on the Australian stock market and the AU to US dollar exchange rate, are shown in Figure 29.

Australian market.



AU/US.

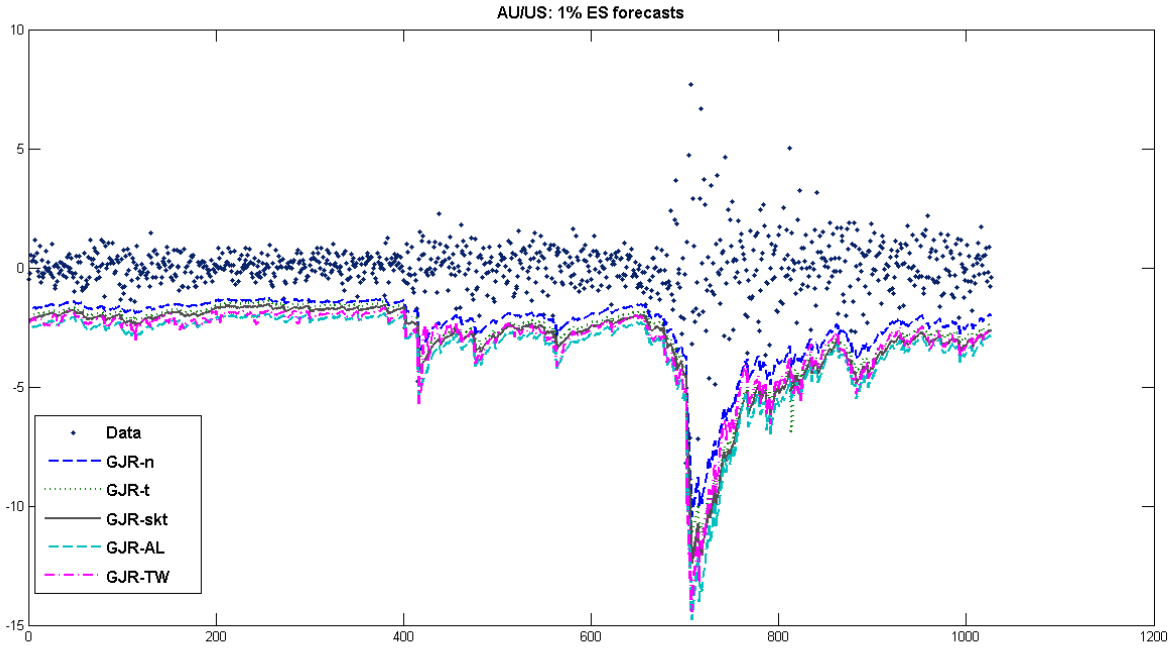


Figure 29: Chapter 3: 1% ES forecasts from GJR-n, GJR-t, GJR-skt, GJR-ALCP, and GJR-TW.

Again, similar to that of VaR level, there is a clear ordering in ES levels across distributions: the Gaussian is least extreme, followed by the Student t, Skewed t, STW, while the AL distribution gives the most extreme ES forecasts. This pattern occurred frequently across the seven series, holding the volatility model constant.

The quantile levels that ES occurs at, for various VaR quantile levels α, are well known and calculable in standard software for the Gaussian and Student t distributions, using their cdf functions; the ES quantile levels, constant for fixed α, for the AL distribution were derived by Chen et al. (2011) and are given in (13) in Chapter 2 and Table 10. The closed forms for the ES and the relation between ES and VaR for the Skewed t are derived and given in Appendix 3, while for the STW this is given by (26) and (28) and allows evaluation of the ES quantile level for a STW at VaR level α. Table 30 shows the approximate quantile levels for ES from the Student t, Skewed t, and STW models, obtained using the average of the estimates of each distribution's parameters over the forecast period in each series.

Using these ES quantile levels δ_α, the ES violation rate, ESRate, is defined as:

$$\text{ESRate} = \frac{1}{m} \sum_{t=n+1}^{n+m} I(y_t < \text{ES}_t),$$

and a good model should have the ESRate very close to the nominal δ_α.

Table 31 contains the ratios of $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$ across all models and the seven series in the forecast period. Again the best risk ratio, closest to 1, is boxed and ESRates significantly different to nominal by the UC test are in bold. At $\alpha = 0.01$, it is clear that models with Gaussian errors are always anti-conservative and significantly under-predict risk levels in all series: on average ESRates are close to 3 times or more the nominal 1%. Further, models with Student t errors also under-predict risk, sometimes significantly; on average, their ES violation rates are 55–84% above nominal. Alternatively, models with AL errors again over-predict risk levels, but not significantly; on average, ESRates are half the nominal 1%, and are thus conservative, agreeing with Chen et al. (2011) and results in Chapter 2. Models with Skewed t errors tend to under-predict risk, not significantly, with ESRates 16–39% too high on average. However, the 3rd and 4th ranked models, by average ESRate ratio, with 1.16 and 1.17 respectively, are the GARCH and

GJR-GARCH with Skewed t errors. The top two ranked models by average ESRate ratio, with 1.02 and 1.05, are the GARCH and T-GARCH with STW errors. The GJR-GARCH and ST-GARCH with STW rank 5th and 6th, respectively, on this measure. Further, by minimum deviation of ratios from 1, the models with STW errors rank 1st, 2nd, 3rd, with the ST-GARCH-STW ranking 5th best. The 4th ranked model is the GJR-GARCH with Skewed t errors. Under these criteria, it is clear that models with STW errors have performed most favourably, followed by the GARCH and GJR-GARCH with Skewed t errors.

At $\alpha = 0.05$, a similar story now holds. Models with Gaussian errors are signifcantly anti-conservative, but now it shows an increase of approximately 50–70% on average, and Student t error models perform similarly and are mostly rejected in 3 of the 7 series by the UC test. Models with AL errors now only marginally over-predict risk levels, with ESRates on average 15–20% below nominal, while Skewed t error models under-predict risk levels again by 17–30% on average. Here, the top-four ranked models, with ESRates clearly closest to nominal on average, are the four STW error models. Three of these, excluding the GJR-GARCH-STW, occupy the top ranked positions by minimum deviation in ratios from 1.

Table 32 shows counts of the number of rejections for each ES forecast model, at a 5% significance level, across the seven series, under the three formal backtests: the UC, the CC, and the DQ test using the ES quantile levels discussed above. At $\alpha = 0.01$ and 0.05, the Gaussian error models are again rejected in all or most series by all tests, while the models with Student t errors are again rejected on average more than the other models. At $\alpha = 0.05$, Student t error models are rejected in all or most series for ES forecasting. The two best models could not be rejected in any series: the T-GARCH-AL and the GJR-GARCH-STW. Models with AL, STW, and Skewed t errors were generally rejected in 1 series only at $\alpha = 0.01$, and thus are quite comparable on these tests across the seven series. At $\alpha = 0.05$, only the GJR-GARCH with STW errors could not be rejected in any series; all other models were rejected at least twice.

Overall, for forecasting ES during this forecast period, models with STW errors have performed more favourably than all other models and error distributions considered, with

ESRates generally closest to nominal in both average and squared deviation and ES forecasts mostly not rejected by the formal tests, across the seven return series. Under each criteria, a model with STW errors ranks first. The models with AL errors may also be attractive for regulatory purposes, since they have very small violation ratios; basically, half the amount of violations expected. However, these smaller violation ratios do signal over-estimation of risk and excessive allocation of capital, which may not be ideal. In this sense, models with STW errors semi-parametric adequate and accurate risk coverage.

3.4.4 Pre-financial-crisis and post-financial-crisis forecast performance

As in Chapter 2, we again consider pre-crisis and post-crisis of the model performance in detail, since model forecasting performance may be different in each period and/or suffer during the crisis. The specific date for the start of the crisis has been chosen for each series in Chapter 2; in this Chapter there are approximately 700 days in the pre-crisis period and approximately 350 days in the post-crisis sample.

Tables 33 and 34 present the ratios of VRate/α and ESRate/δ_α at $\alpha = 0.01, 0.05$ for the pre-crisis period for the VaR and ES forecast models, for the post-crisis period, the ratios are presented in Tables 35 and 36. Again, the boxes indicate the models with ratios closest to one, the bold indicates the models are rejected by UC tests, in each market. These ratios are shown in Figures 30 and 31 across error distributions. Circles indicate GARCH; squares indicate GJR; crosses indicate TGARCH; diamonds indicate STGARCH; large triangles indicate the mean of the VRates for each distribution.

The results for the pre-crisis sample are highly consistent with those for the whole forecast sample, no doubt influenced by the larger overlapping sample size: models with STW and Student t errors forecast VaR most accurately at the 1% and 5% risk levels, with the VRate averaging close to 1, though STW error models have VRate ratios with slightly lower variation around 1. Further, only models with STW errors have ESRate ratios consistently, and averaging, close to 1. Models with AL errors are again the only always conservative risk forecasters for both VaR and ES.

Results for the post-crisis period tell a slightly different story. For VaR forecasting, models with Student t, Skewed t, and STW errors perform well at $\alpha = 0.01$, all with

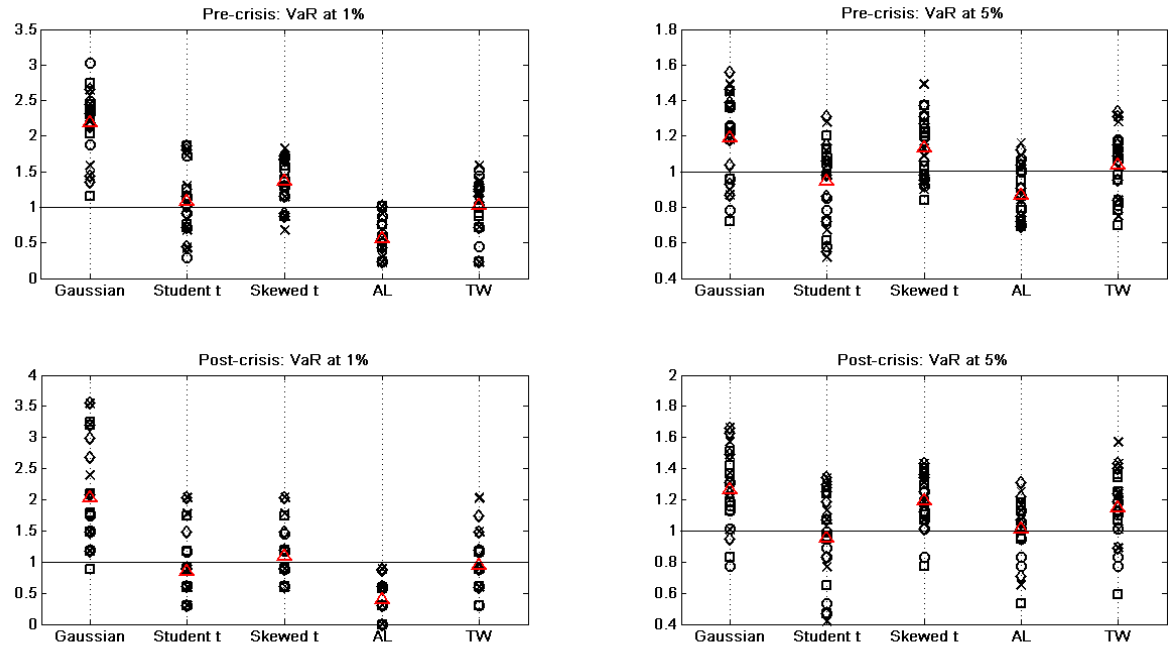


Figure 30: Chapter 3: The ratios of VRate/α at $\alpha = 0.01, 0.05$ for the pre-crisis and post-crisis periods.

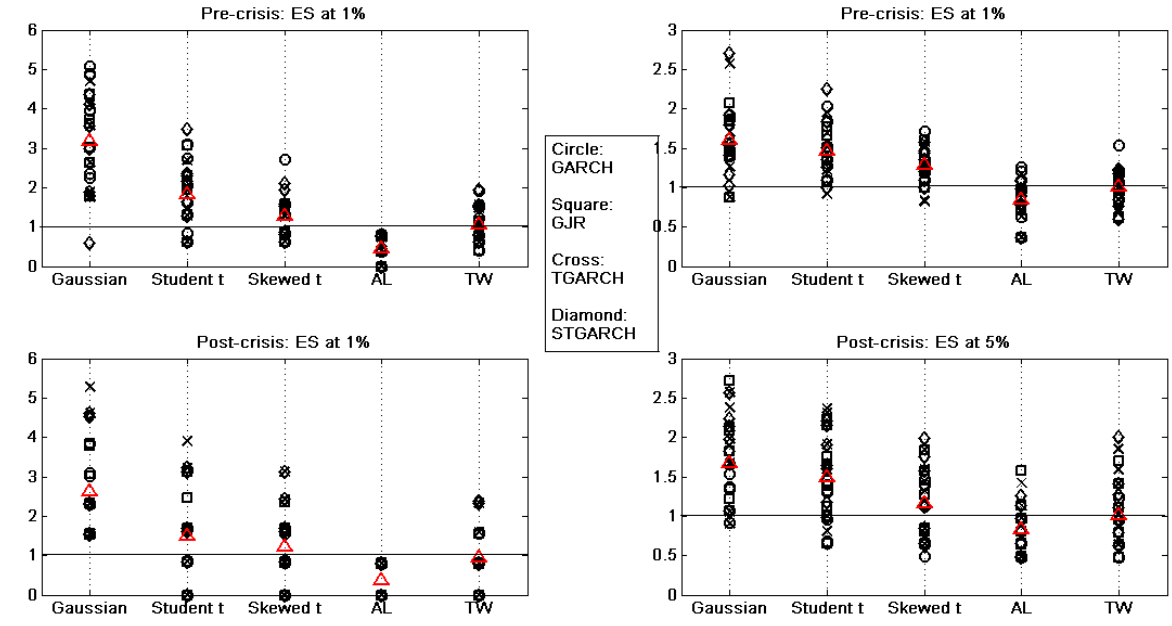


Figure 31: Chapter 3: The ratios of ESRate/δ_α at $\alpha = 0.01, 0.05$ for the pre-crisis and post-crisis periods.

average ratios close to 1 and similar deviations about 1, across the seven series. For ES forecasting, the TW is clearly the best model post-crisis, with its average ratio closest to 1 and smallest deviation about 1. At the 5% risk level however, the models with AL and Student t errors perform best for VaR forecasting, with STW models slightly under-predicting risk levels, on average. For ES forecasting at $\alpha = 0.05$, the TW has the closest average ratio to 1 post-crisis, but the AL also does well and has the smallest deviation in ratios from 1.

3.4.5 Loss function

Loss function is again used to evaluate the performance of risk forecasting models as in Chapter 2. Tables 37 and Table 38 present the loss function for the VaR and ES forecasts via various models, taken over the seven series in the entire forecast period. Figure 32 shows the mean of the loss function of each model. Two things are apparent: the GJR model (shown as squares) usually has the lowest average loss for each error distribution; for VaR forecasting at $\alpha = 0.01$, models with Student t, Skewed t, and STW errors have the lowest, and comparable, average losses. For VaR forecasting at $\alpha = 0.05$, however, the Skewed t, AL, and STW-error models have comparable and the lowest average loss. For ES, losses among all distributions except the Gaussian, which has the largest average loss in each case, seem quite close and comparable.

Overall, the STW model is the most favourably performed risk forecaster for this forecast data period across the seven series over both VaR and ES forecasting at both $\alpha = 0.01, 0.05$ levels. By almost all criteria, models with STW errors ranks best or equal best, with violation rates closest to 1 by average and squared deviation, and minimum number of model rejections by formal tests, both in the entire period and in the pre- and post-GFC periods. Models with Student t errors did well at VaR forecasting for $\alpha = 0.01$, while models with AL errors were always conservative and exhibited violation rates usually below nominal, with comparatively small variation in violation rate ratios.

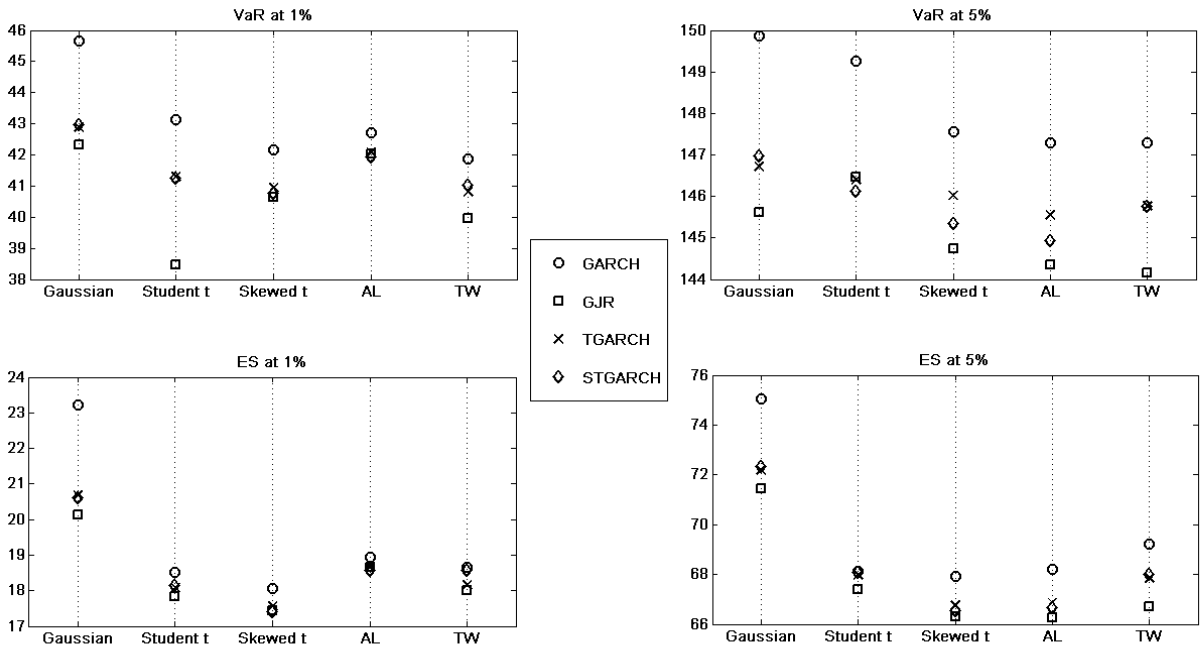


Figure 32: Chapter 3: Average loss functions of VaR and ES forecasts from various distributions across various volatility models.

3.5 Conclusion

Calculating a benchmark for allocating regulatory capital and thus protecting the financial institutions from the risk during extreme market movements is the ultimate goal of VaR and ES models. However, as another essential function of these financial entities is to make profit, the allocation of capital is a matter of great account .

In this chapter, we argue that other than using an extremely conservative model, a more appropriate approach should be able to relieve the burden of over-allocation of regulatory capital, and protect against the risky under-allocation of capital, by more accurately forecasting dynamic risk levels, thus carefully and properly increasing the investment opportunities in more profitable assets.

For this purpose, we proposed the use of a two-sided Weibull conditional return distribution, coupled with a volatility model. Properties of this distribution were developed and presented, including the VaR and expected shortfall functions. This distribution is more flexible than the AL distribution, with a wider range of skewness and kurtosis. It

has heavier tails than the Gaussian distribution and the Student t family, but less fat than the AL tails.

An adaptive MCMC method was employed for estimation and forecasting. An empirical study of seven asset return series found that models with conditional two-sided Weibull errors were highly accurate at forecasting both VaR and ES levels and could not be consistently rejected or bettered across several criteria, compared to the Gaussian, Student t, Skewed t, and asymmetric Laplace conditional return distributions. This accurate performance was found to hold both before the GFC hit markets as well as during and after the GFC period.

Hopefully, the model introduced in this chapter offers both the regulators and the financial institutions a new option or compromise between suffering from excess violations or from unnecessarily reduced profit. It is clear that the two-sided Weibull has improved the modelling of the tails of the conditional return distribution.

Table 28: Chapter 3: Ratios of $\hat{\alpha}/\alpha$ at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	IBM	Mean	Std.
G-n	**2.09**	**2.43**	**2.02**	**2.11**	**2.05**	**1.93**	**1.84**	2.07	1.08
GJR-n	**2.65**	**2.53**	**2.50**	**1.81**	**2.14**	1.54	1.26	2.06	1.17
TG-n	**2.94**	**2.43**	**2.60**	**1.91**	**2.24**	1.54	1.46	2.16	1.27
ST-n	**2.94**	**2.63**	**2.69**	**1.81**	**2.24**	**1.93**	1.46	2.24	1.34
G-t	1.14	0.88	1.54	0.70	0.68	0.90	**0.29**	0.88	0.38
GJR-t	1.23	1.07	**1.83**	0.60	0.68	1.03	0.68	1.02	0.40
TG-t	**1.80**	1.07	**1.83**	0.91	0.68	0.77	**0.39**	1.06	0.52
ST-t	**1.71**	1.07	**1.92**	0.70	0.68	0.64	0.78	1.07	0.50
G-SKT	1.33	1.26	1.63	1.31	1.07	1.29	1.17	1.29	0.34
GJR-SKT	1.33	1.17	1.63	1.31	0.78	1.16	1.17	1.22	0.32
TG-SKT	**1.71**	1.17	**1.83**	1.51	0.78	0.77	1.17	1.28	0.48
ST-SKT	1.61	1.17	**1.83**	1.41	0.78	1.03	1.17	1.28	0.44
G-AL	0.47	0.49	0.58	0.60	0.58	**0.13**	0.68	0.50	0.52
GJR-AL	0.47	**0.39**	0.48	0.60	**0.39**	**0.13**	0.87	0.48	0.56
TG-AL	0.66	**0.29**	0.48	0.70	0.49	**0.13**	0.78	0.50	0.54
ST-AL	0.66	**0.39**	0.58	0.60	0.49	**0.13**	0.78	0.52	0.52
G-TW	1.23	1.07	1.35	1.31	0.68	0.51	1.07	1.03	0.30
GJR-TW	1.04	0.88	1.25	0.70	0.68	**0.26**	0.87	0.81	0.35
TG-TW	1.23	1.07	**1.73**	1.21	0.68	0.51	1.17	1.09	0.38
ST-TW	1.33	1.07	1.44	1.01	0.88	0.51	1.26	1.07	0.30
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	IBM	Mean	Std.
G-n	1.25	**1.28**	1.23	1.17	1.07	1.21	0.78	1.14	0.22
GJR-n	**1.44**	**1.28**	**1.31**	**1.31**	1.09	1.29	0.76	1.21	0.30
TG-n	**1.52**	1.26	**1.31**	**1.41**	1.09	1.26	0.91	1.25	0.31
ST-n	**1.59**	**1.36**	**1.33**	**1.39**	1.11	1.16	0.89	1.26	0.34
G-t	1.02	1.01	1.04	0.74	0.84	0.98	**0.52**	0.88	0.21
GJR-t	1.21	1.07	1.17	0.87	0.84	1.06	**0.56**	0.97	0.21
TG-t	**1.29**	1.05	1.15	0.91	0.82	1.03	**0.49**	0.96	0.24
ST-t	**1.31**	1.13	1.15	0.93	0.90	1.00	**0.52**	0.99	0.23
G-SKT	1.16	1.21	1.12	1.13	1.05	1.18	0.89	1.10	0.14
GJR-SKT	**1.31**	1.21	1.15	**1.31**	1.01	1.21	0.93	1.16	0.21
TG-SKT	**1.35**	1.15	1.19	**1.39**	1.07	1.21	1.01	1.20	0.23
ST-SKT	**1.36**	1.25	1.15	1.25	1.09	1.08	0.99	1.17	0.20
G-AL	1.00	1.01	0.88	0.99	0.84	0.90	0.78	0.91	0.12
GJR-AL	1.02	0.95	0.85	1.01	0.84	0.93	**0.66**	0.89	0.16
TG-AL	1.08	0.91	0.85	1.11	0.88	0.98	**0.70**	0.93	0.15
ST-AL	1.10	0.91	0.87	1.03	0.94	0.98	**0.72**	0.93	0.13
G-TW	1.12	1.17	1.08	1.03	0.94	1.06	0.82	1.03	0.11
GJR-TW	1.23	1.11	1.10	1.13	0.88	1.00	**0.72**	1.02	0.16
TG-TW	1.27	1.13	1.21	1.35	0.97	1.18	0.85	1.14	0.21
ST-TW	**1.29**	1.25	1.19	1.27	1.03	1.06	0.85	1.14	0.20

Note: Boxes indicate the ratio closest to 1 in that market, bold indicates the model is rejected by the unconditional coverage test (at a 5% level), for each market.

Table 29: Chapter 3: Counts of model rejections for VaR models at $\alpha = 0.01, 0.05$.

	$\alpha = 0.01$				$\alpha = 0.05$			
Method	UC	CC	DQ	Total(out of 7)	UC	CC	DQ	Total(out of 7)
G-n	7	6	7	**7**	1	2	5	**5**
GJR-n	5	4	6	6	4	1	3	4
TG-n	5	4	5	5	3	2	3	3
ST-n	6	4	6	6	4	2	3	4
G-t	1	1	2	2	1	2	2	2
GJR-t	1	0	4	4	1	2	2	2
TG-t	3	0	3	3	2	2	2	2
ST-t	2	0	5	5	2	2	2	2
G-SKT	0	0	3	3	0	1	3	3
GJR-SKT	0	0	1	$\boxed{1}$	2	0	2	2
TG-SKT	2	0	2	2	2	1	1	2
ST-SKT	1	0	1	1	1	0	2	2
G-AL	1	1	3	3	0	0	2	2
GJR-AL	3	1	1	3	1	1	3	3
TG-AL	2	0	3	3	1	0	2	2
ST-AL	2	1	1	2	1	0	2	2
G-TW	0	0	2	2	0	0	3	3
GJR-TW	1	0	1	1	1	0	0	$\boxed{1}$
TG-TW	1	0	3	3	0	0	1	1
ST-TW	0	0	3	3	1	0	3	3

Note: Boxes indicate the favoured model, bold indicates the leaset favoured model(at a 5% level).

Table 30: Chapter 3: Estimated nominal ES levels δ_α at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	IBM
G-t	0.0036	0.0036	0.0037	0.0035	0.0035	0.0037	0.0034
GJR-t	0.0036	0.0037	0.0037	0.0035	0.0035	0.0037	0.0034
TG-t	0.0036	0.0037	0.0037	0.0035	0.0036	0.0037	0.0034
ST-t	0.0037	0.0037	0.0037	0.0035	0.0036	0.0036	0.0034
G-SKT	0.0036	0.0036	0.0037	0.0035	0.0036	0.0037	0.0034
GJR-SKT	0.0036	0.0037	0.0037	0.0035	0.0036	0.0037	0.0034
TG-SKT	0.0036	0.0037	0.0037	0.0035	0.0036	0.0037	0.0034
STG-SKT	0.0037	0.0037	0.0037	0.0035	0.0036	0.0036	0.0034
G-TW	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037
GJR-TW	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037
TG-TW	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037
ST-TW	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	IBM
G-t	0.0182	0.0184	0.0188	0.0179	0.0181	0.0187	0.0173
GJR-t	0.0185	0.0189	0.0190	0.0180	0.0181	0.0187	0.0174
TG-t	0.0186	0.0187	0.0190	0.0181	0.0182	0.0187	0.0175
ST-t	0.0186	0.0188	0.0190	0.0181	0.0182	0.0186	0.0175
G-SKT	0.0183	0.0185	0.0189	0.0180	0.0182	0.0187	0.0174
GJR-SKT	0.0186	0.0187	0.0190	0.0181	0.0182	0.0187	0.0175
TG-SKT	0.0186	0.0187	0.0190	0.0181	0.0182	0.0187	0.0175
STG-SKT	0.0186	0.0188	0.0190	0.0181	0.0182	0.0186	0.0175
G-TW	0.0188	0.0187	0.0188	0.0186	0.0188	0.0188	0.0187
GJR-TW	0.0188	0.0188	0.0189	0.0186	0.0187	0.0187	0.0187
TG-TW	0.0189	0.0188	0.0189	0.0186	0.0188	0.0188	0.0187
ST-TW	0.0189	0.0188	0.0189	0.0186	0.0188	0.0188	0.0187

Table 31: Chapter 3: ESRatios $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	IBM	Mean	Std.
G-n	**4.68**	**3.79**	**4.25**	**3.40**	**2.79**	2.01	**2.52**	3.35	2.51
GJR-n	**3.70**	**2.53**	**3.75**	**3.40**	**2.53**	2.01	**2.27**	2.88	2.00
TG-n	**4.68**	**2.28**	**4.50**	**3.14**	2.03	2.01	2.02	2.95	2.24
ST-n	**4.44**	**2.53**	**4.25**	**3.14**	2.03	1.00	**2.52**	2.84	2.16
G-t	2.13	1.35	**3.12**	1.15	1.92	0.35	0.86	1.55	1.01
GJR-t	**2.36**	1.84	**3.10**	1.14	1.93	0.35	1.71	1.77	1.12
TG-t	**2.34**	1.85	**3.10**	1.14	1.92	0.35	1.14	1.69	1.09
ST-t	**2.60**	2.11	**3.36**	1.70	1.65	0.35	1.14	1.84	1.24
G-SKT	1.06	1.34	**2.33**	1.14	1.09	0.35	0.86	1.17	0.58
GJR-SKT	1.30	1.59	1.81	0.85	1.09	0.35	1.14	1.16	0.47
TG-SKT	1.82	1.32	2.06	1.14	1.09	0.35	1.14	1.28	0.58
ST-SKT	1.82	1.59	**2.32**	0.85	1.10	0.35	1.70	1.39	0.73
G-AL	0.51	0.52	0.26	0.81	0.78	**0.00**	0.26	0.45	0.62
GJR-AL	0.51	0.26	0.26	0.54	0.78	**0.00**	0.26	0.37	0.67
TG-AL	0.51	0.52	0.51	0.27	0.78	0.34	0.26	0.46	0.57
ST-AL	0.51	0.26	0.26	0.54	0.78	**0.00**	0.26	0.41	0.64
G-TW	1.01	0.52	1.80	1.08	1.04	0.35	1.30	1.02	0.45
GJR-TW	1.01	0.78	1.28	0.54	0.79	0.35	0.78	0.79	0.35
TG-TW	1.52	0.78	1.54	0.82	1.04	0.34	1.30	1.05	0.41
ST-TW	1.27	1.56	1.80	0.81	1.31	0.34	1.30	1.20	0.49
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	IBM	Mean	Std.
G-n	**1.84**	**1.79**	1.42	**1.54**	1.44	1.38	1.29	1.53	0.56
GJR-n	**2.28**	**1.84**	**1.67**	1.34	**1.74**	1.32	0.94	1.59	0.71
TG-n	**2.57**	**1.79**	**1.77**	1.44	**1.64**	1.45	0.94	1.66	0.80
ST-n	**2.66**	**1.99**	**1.77**	**1.54**	**1.59**	1.45	1.04	1.72	0.86
G-t	**1.72**	**1.80**	1.38	1.23	1.29	1.24	0.95	1.37	0.46
GJR-t	**1.85**	**1.75**	**1.72**	1.12	1.46	1.31	1.06	1.47	0.55
TG-t	**2.09**	**1.77**	**1.77**	1.22	1.45	1.38	0.95	1.52	0.63
ST-t	**2.24**	**1.86**	**1.72**	1.17	1.45	1.31	1.06	1.54	0.67
G-SKT	1.24	**1.63**	1.37	1.06	0.96	0.96	0.95	1.17	0.29
GJR-SKT	1.43	**1.50**	**1.47**	1.06	1.07	1.03	0.95	1.22	0.31
TG-SKT	**1.58**	**1.56**	**1.52**	1.06	1.23	1.31	0.83	1.30	0.40
ST-SKT	**1.58**	**1.55**	**1.62**	1.11	1.18	1.11	0.94	1.30	0.39
G-AL	1.13	1.11	1.05	0.82	0.85	**0.42**	0.63	0.86	0.28
GJR-AL	1.03	1.00	1.15	0.88	0.85	**0.42**	0.69	0.86	0.27
TG-AL	1.03	0.79	1.10	0.93	0.79	**0.42**	**0.58**	0.81	0.30
ST-AL	1.03	0.79	1.1	0.88	0.85	**0.42**	0.74	0.84	0.27
G-TW	1.11	1.25	1.22	0.86	0.93	0.69	0.83	0.98	0.20
GJR-TW	1.26	1.04	1.33	0.81	0.83	0.55	0.67	0.93	0.28
TG-TW	1.26	1.19	1.37	0.97	1.04	0.82	0.88	1.08	0.20
ST-TW	1.26	1.24	**1.48**	1.03	0.88	0.75	0.83	1.07	0.26

Note: Boxes indicate the ratio closest to 1 in that market, bold indicates the model is rejected by the unconditional coverage test (at a 5% level), for each market.

Table 32: Chapter 3: Counts of model rejections for ES models at $\alpha = 0.01, 0.05$.

	$\alpha = 0.01$				$\alpha = 0.05$			
Method	UC	CC	DQ	Total (out of 6)	UC	CC	DQ	Total (out of 6)
G-n	6	5	6	6	3	4	6	**6**
GJR-n	6	4	7	**7**	4	3	5	5
TG-n	4	3	6	6	4	4	5	5
ST-n	5	3	7	**7**	5	4	5	5
G-t	1	1	4	4	2	3	5	5
GJR-t	2	1	3	3	3	3	6	**6**
TG-t	2	1	2	2	3	4	4	4
ST-t	2	1	2	2	3	3	4	4
G-SKT	1	1	3	3	1	1	5	5
GJR-SKT	0	0	1	1	2	1	2	2
TG-SKT	0	0	1	1	3	0	2	3
ST-SKT	1	0	0	1	3	0	2	3
G-AL	1	0	0	1	1	1	5	5
GJR-AL	1	0	0	1	1	0	2	2
TG-AL	0	0	0	$\boxed{0}$	2	0	3	3
ST-AL	1	0	0	1	1	1	3	3
G-TW	0	0	2	2	0	1	5	5
GJR-TW	0	0	0	$\boxed{0}$	0	0	0	$\boxed{0}$
TG-TW	0	0	1	1	0	1	2	2
ST-TW	0	1	1	1	1	1	2	2

Note: Boxes indicate the favoured model, bold indicates the leaset favoured model(at a 5% level).

Table 33: Chapter 3: Before the financial crisis: ratios of $\hat{\alpha}/\alpha$ at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	IBM	Mean	Std.
G-n	**2.37**	**3.03**	**2.16**	**2.44**	**2.32**	**2.49**	**1.88**	**2.38**	**1.42**
GJR-n	**2.37**	**2.74**	**2.16**	**2.13**	**2.32**	**2.04**	1.15	2.13	1.22
TG-n	**2.65**	**2.45**	**2.30**	**2.29**	**2.32**	1.59	1.44	2.15	1.22
STG-n	**2.65**	**2.45**	**2.44**	**2.13**	**2.32**	1.36	1.44	2.11	1.21
G-t	1.12	1.15	**1.72**	0.91	0.72	1.13	**0.29**	1.01	0.41
GJR-t	1.26	1.15	**1.87**	0.76	0.72	1.13	0.72	1.09	0.39
TG-t	**1.81**	1.30	**1.72**	1.07	0.72	0.68	**0.43**	1.11	0.50
ST-t	**1.81**	1.15	**1.87**	0.91	0.72	0.45	1.01	1.13	0.51
G-SKT	1.39	1.59	**1.72**	1.52	1.30	1.59	1.15	1.47	0.50
GJR-SKT	1.39	1.30	1.58	1.68	0.87	1.36	1.15	1.33	0.42
TG-SKT	1.67	1.15	**1.72**	**1.83**	0.87	0.68	1.30	1.32	0.52
ST-SKT	1.67	1.30	**1.72**	1.68	0.87	0.91	1.15	1.33	0.47
G-AL	0.56	0.58	0.57	0.61	0.58	**0.23**	0.87	0.57	0.46
GJR-AL	0.56	**0.43**	**0.43**	0.76	**0.43**	**0.23**	1.01	0.55	0.51
TG-AL	0.56	**0.29**	**0.43**	0.91	**0.43**	**0.23**	1.01	0.55	0.53
STG-AL	0.56	**0.43**	**0.43**	0.76	**0.43**	**0.23**	1.01	0.55	0.51
G-TW	1.26	1.44	1.44	1.52	0.72	0.45	1.01	1.12	0.40
GJR-TW	0.98	1.01	1.29	0.91	0.72	**0.23**	0.87	0.86	0.34
TG-TW	1.12	1.30	1.58	1.37	0.72	**0.23**	1.30	1.09	0.44
ST-TW	1.26	1.15	1.29	1.22	0.72	**0.23**	1.30	1.02	0.38
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	IBM	Mean	Std.
G-n	**1.37**	**1.36**	**1.26**	**1.25**	0.96	**1.22**	0.78	1.17	0.27
GJR-n	**1.45**	**1.36**	**1.21**	**1.37**	0.96	**1.36**	0.72	**1.20**	**0.32**
TG-n	**1.45**	**1.24**	1.18	**1.49**	0.90	1.18	0.87	1.19	**0.29**
ST-n	**1.56**	**1.39**	1.18	**1.46**	0.93	1.04	0.87	1.20	0.32
G-t	1.09	1.04	1.06	0.85	0.78	1.09	**0.58**	0.93	0.20
GJR-t	1.20	1.13	1.12	0.98	0.72	1.13	**0.61**	0.98	0.21
TG-t	**1.28**	1.04	1.06	0.98	**0.67**	1.00	**0.52**	0.93	0.25
ST-t	**1.31**	1.15	1.06	0.98	0.72	0.86	**0.55**	0.95	0.25
G-SKT	1.20	**1.24**	1.12	1.19	0.96	1.22	0.92	1.12	0.17
GJR-SKT	**1.31**	**1.27**	1.03	**1.37**	0.84	1.22	1.01	1.15	0.23
TG-SKT	**1.34**	1.13	1.09	**1.49**	0.90	1.13	0.98	1.15	0.24
ST-SKT	**1.37**	**1.27**	1.06	1.31	0.93	0.95	0.98	1.12	0.21
G-AL	1.00	1.01	0.80	1.07	0.70	0.86	0.78	0.89	0.17
GJR-AL	0.95	0.95	**0.69**	1.04	0.70	0.91	0.72	0.85	0.20
TG-AL	1.09	0.84	**0.69**	1.16	0.70	0.86	0.72	0.86	0.22
ST-AL	1.12	0.87	**0.69**	1.04	0.75	0.91	0.72	0.87	0.20
G-TW	1.17	1.18	1.06	1.13	0.81	1.09	0.84	1.04	0.15
GJR-TW	1.17	1.10	0.98	1.13	0.70	0.95	0.78	0.97	0.17
TG-TW	1.28	1.10	1.03	1.31	0.75	1.13	0.84	1.06	0.20
ST-TW	1.31	1.15	1.09	1.34	0.84	0.95	0.84	1.08	0.21

Note: Boxes indicate the ratio is closest to 1 in that market, bold indicates that the model is rejected by the UC test (at a 5% level), for each market.

Table 34: Chapter 3: Before the financial crisis: $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	IBM	Mean	Std.
G-n	**5.08**	**4.88**	**4.86**	**3.96**	**3.01**	**2.36**	2.25	**3.77**	**2.99**
GJR-n	**3.63**	**2.63**	**3.73**	**4.36**	**3.01**	1.77	**2.63**	3.11	2.25
TG-n	**4.71**	**2.63**	**4.11**	**3.57**	1.88	1.77	1.88	2.93	2.23
STG-n	**4.35**	**3.00**	**4.11**	**3.57**	1.88	0.59	**3.00**	2.93	2.28
G-t	**2.74**	1.60	**3.11**	1.30	2.04	0.62	0.85	1.75	1.14
GJR-t	2.31	1.95	**3.08**	1.30	2.04	0.62	2.12	1.92	1.17
TG-t	1.91	1.96	**2.70**	1.29	2.04	0.62	1.28	1.69	0.93
ST-t	2.29	2.35	**3.47**	2.15	1.63	0.62	1.27	1.97	1.28
G-SKT	1.55	1.59	**2.71**	1.30	0.81	0.62	0.85	1.35	0.74
GJR-SKT	1.53	1.57	1.54	0.86	0.81	0.62	1.27	1.17	0.41
TG-SKT	1.53	1.57	1.54	1.29	0.81	0.62	1.27	1.23	0.42
ST-SKT	1.53	1.57	1.93	0.86	0.82	0.62	2.11	1.35	0.64
G-AL	0.75	0.39	0.38	0.82	0.78	**0.00**	**0.00**	0.44	0.64
GJR-AL	0.75	0.75	0.38	0.41	0.78	**0.00**	**0.00**	0.44	0.64
TG-AL	0.75	0.39	0.77	0.41	0.78	0.61	**0.00**	0.53	0.54
STG-AL	0.75	0.39	0.77	0.41	0.78	**0.00**	**0.00**	0.44	0.64
G-TW	1.49	0.39	1.92	1.23	0.78	0.61	1.55	1.14	0.54
GJR-TW	1.12	0.77	1.15	0.41	0.78	0.61	0.78	0.80	0.31
TG-TW	1.49	0.77	1.15	0.82	0.78	0.61	1.55	1.02	0.35
ST-TW	1.49	1.93	1.53	0.82	0.78	0.61	1.55	1.24	0.53
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	IBM	Mean	Std.
G-n	**1.92**	**1.92**	**1.47**	**1.87**	**1.48**	1.39	1.40	1.64	0.68
GJR-n	**2.07**	**1.84**	**1.47**	**1.48**	**1.56**	1.39	0.88	1.53	0.63
TG-n	**2.57**	**1.70**	**1.47**	**1.71**	1.41	1.27	0.88	1.57	0.75
STG-n	**2.71**	**1.92**	**1.54**	**1.79**	1.41	1.16	1.03	**1.65**	**0.83**
G-t	**1.84**	**2.03**	1.36	**1.52**	1.27	1.45	1.09	1.51	0.59
GJR-t	**1.66**	**1.77**	**1.51**	1.35	1.36	1.46	1.08	1.45	0.50
TG-t	**1.95**	**1.69**	**1.51**	1.43	1.36	1.22	0.92	1.44	0.54
ST-t	**2.25**	**1.84**	**1.51**	1.26	1.36	1.10	1.00	1.47	0.62
G-SKT	1.29	**1.71**	1.36	1.27	1.19	1.21	1.09	1.30	0.36
GJR-SKT	1.35	**1.61**	1.29	1.26	1.20	1.22	1.00	1.28	0.32
TG-SKT	**1.58**	**1.54**	1.36	1.26	1.20	1.22	0.83	1.28	0.37
ST-SKT	**1.57**	1.46	1.44	1.35	1.20	1.10	1.00	1.30	0.36
G-AL	1.21	1.26	1.09	0.99	0.95	**0.37**	0.63	0.93	0.31
GJR-AL	0.99	0.99	0.94	0.99	0.87	**0.37**	0.71	0.84	0.27
TG-AL	0.99	0.71	0.94	1.16	0.63	**0.37**	0.63	0.77	0.34
STG-AL	0.99	0.78	1.09	1.08	0.79	**0.37**	0.78	0.84	0.28
G-TW	1.19	1.54	1.22	1.06	1.16	0.85	0.93	1.13	0.25
GJR-TW	1.19	1.00	1.14	0.98	0.85	0.61	0.62	0.91	0.23
TG-TW	1.18	1.00	1.14	1.14	1.00	0.72	0.85	1.01	0.16
ST-TW	1.18	1.23	1.22	1.23	0.77	0.60	0.85	1.01	0.24

Note: Boxes indicate the ratio is closest to 1 in that market, bold indicates that the model is rejected by the unconditional coverage test (at a 5% level), for each market.

Table 35: Chapter 3: Post-financial crisis: ratios of $\hat{\alpha}/\alpha$ at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	IBM	Mean	Std.
G-n	1.48	1.19	1.74	1.48	1.49	1.19	1.78	1.48	0.53
GJR-n	**3.25**	2.09	**3.20**	1.18	1.79	0.89	1.48	1.98	1.31
TG-n	**3.55**	**2.39**	**3.20**	1.18	2.08	1.49	1.48	2.20	1.46
STG-n	**3.55**	**2.99**	**3.20**	1.18	2.08	**2.68**	1.48	**2.45**	**1.67**
G-t	1.18	**0.30**	1.16	**0.30**	0.60	0.60	0.30	0.63	0.52
GJR-t	1.18	0.90	1.74	**0.30**	0.60	0.89	0.59	0.89	0.45
TG-t	1.78	0.60	2.03	**0.59**	0.60	0.89	0.30	0.97	0.62
ST-t	1.48	0.90	2.03	**0.30**	0.60	0.89	0.30	0.93	0.59
G-SKT	1.18	0.60	1.45	0.89	0.60	0.89	1.19	0.97	0.30
GJR-SKT	1.18	0.90	1.74	**0.59**	0.60	0.89	1.19	1.01	0.37
TG-SKT	1.78	1.19	2.03	0.89	0.60	0.89	0.89	1.18	0.52
ST-SKT	1.48	0.90	2.03	0.89	0.60	1.19	1.19	1.18	0.47
G-AL	0.30	**0.30**	**0.58**	**0.59**	0.60	**0.00**	**0.30**	0.38	0.65
GJR-AL	0.30	**0.30**	**0.58**	**0.30**	**0.30**	**0.00**	**0.59**	0.34	0.69
TG-AL	0.89	**0.30**	**0.58**	**0.30**	0.60	**0.00**	**0.30**	0.42	0.64
STG-AL	0.89	**0.30**	0.87	**0.30**	0.60	**0.00**	**0.30**	0.46	0.62
G-TW	1.18	**0.30**	1.16	0.89	0.60	0.60	1.19	0.84	0.36
GJR-TW	1.18	0.60	1.16	**0.30**	0.60	0.30	0.89	0.72	0.45
TG-TW	1.48	0.60	**2.03**	0.89	0.60	0.89	0.89	1.05	0.49
ST-TW	1.48	0.90	1.74	**0.59**	1.19	0.89	1.19	1.14	0.39
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	IBM	Mean	Std.
G-n	1.01	1.13	1.16	1.01	1.31	1.19	0.77	1.08	0.18
GJR-n	**1.42**	1.13	**1.51**	1.18	**1.37**	1.19	0.83	1.23	0.31
TG-n	**1.66**	1.31	**1.57**	1.24	**1.49**	**1.37**	1.01	**1.38**	**0.43**
ST-n	**1.66**	1.31	**1.63**	1.24	**1.49**	1.31	0.95	1.37	**0.43**
G-t	0.89	0.96	0.99	**0.53**	0.95	0.83	**0.47**	0.80	0.28
GJR-t	1.24	0.96	1.28	0.65	1.07	0.95	**0.47**	0.95	0.28
TG-t	1.30	1.07	1.34	0.77	1.13	1.07	**0.42**	1.01	0.30
ST-t	1.30	1.07	1.34	0.83	1.25	1.19	**0.47**	1.07	0.30
G-SKT	1.07	1.13	1.10	1.01	1.25	1.13	0.83	1.07	0.14
GJR-SKT	1.30	1.07	**1.40**	1.18	1.37	1.19	0.77	1.18	0.27
TG-SKT	1.36	1.19	**1.40**	1.18	1.43	1.31	1.07	1.28	0.30
ST-SKT	1.36	1.19	1.34	1.12	1.43	1.25	1.01	1.24	0.28
G-AL	1.01	1.01	1.05	0.83	1.13	0.95	0.77	0.96	0.12
GJR-AL	1.18	0.96	1.16	0.95	1.13	0.95	0.53	0.98	0.21
TG-AL	1.07	1.07	1.16	1.01	1.25	1.13	0.65	1.05	0.18
ST-AL	1.07	1.01	1.22	1.01	1.31	1.07	0.71	1.06	0.18
G-TW	1.01	1.13	1.10	0.83	1.19	1.01	0.77	1.01	0.14
GJR-TW	1.36	1.13	1.34	1.12	1.25	1.07	0.59	1.12	0.27
TG-TW	1.24	1.19	1.57	1.01	1.43	1.25	0.89	1.23	0.31
ST-TW	1.24	1.43	1.40	1.12	1.43	1.19	0.89	1.24	0.30

Note: Boxes indicate the ratio closest to 1 in that market, bold indicates that the model is rejected by the unconditional coverage test (at a 5% level), for each market.

Table 36: Chapter 3: Post-Financial Crisis: Ratios of $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	IBM	Mean	Std.
G-n	**3.85**	1.55	3.02	2.31	2.32	1.55	3.09	2.53	1.72
GJR-n	**3.85**	2.33	**3.78**	1.54	1.55	2.32	1.54	2.41	1.70
TG-n	**4.61**	1.55	**5.29**	2.31	2.32	2.32	2.31	**2.96**	**2.35**
STG-n	**4.61**	1.55	**4.53**	2.31	2.32	1.55	1.54	2.63	2.07
G-t	0.83	0.84	3.15	0.84	1.68	0.00	0.88	1.17	0.94
GJR-t	2.45	1.61	3.12	0.84	1.68	0.00	0.85	1.51	1.10
TG-t	**3.22**	1.63	**3.91**	0.83	1.69	0.00	0.85	1.73	1.48
ST-t	**3.24**	1.62	3.12	0.83	1.69	0.00	0.85	1.62	1.28
G-SKT	0.00	0.82	1.56	0.84	1.68	0.00	0.87	0.83	0.64
GJR-SKT	0.80	1.62	2.34	0.83	1.68	0.00	0.85	1.16	0.73
TG-SKT	2.41	0.81	3.12	0.83	1.68	0.00	0.85	1.39	1.07
ST-SKT	2.43	1.62	3.12	0.83	1.68	0.00	0.85	1.51	1.10
G-AL	0.00	0.80	0.00	0.79	0.80	0.00	0.79	0.45	0.67
GJR-AL	0.00	0.00	0.00	0.79	0.80	0.00	0.79	0.34	0.77
TG-AL	0.00	0.80	0.00	0.00	0.80	0.00	0.79	0.34	0.77
STG-AL	0.00	0.00	0.00	0.79	0.80	0.00	0.79	0.34	0.77
G-TW	0.00	0.80	1.55	0.80	1.59	0.00	0.80	0.79	0.63
GJR-TW	0.79	0.80	1.55	0.80	0.80	0.00	0.80	0.79	0.46
TG-TW	1.58	0.80	2.33	0.80	1.60	0.00	0.80	1.13	0.71
ST-TW	0.79	0.80	2.33	0.80	2.39	0.00	0.80	1.13	0.83
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	IBM	Mean	Std.
G-n	1.66	1.53	1.34	0.91	1.37	1.37	1.06	1.32	0.40
GJR-n	**2.72**	1.83	**2.08**	1.06	**2.13**	1.22	1.06	1.73	0.94
TG-n	**2.57**	**1.98**	**2.38**	0.91	**2.13**	1.67	1.06	1.81	**1.00**
STG-n	**2.57**	**2.14**	**2.23**	1.06	**1.98**	1.82	1.06	1.84	0.99
G-t	1.45	1.30	1.40	0.66	1.33	0.96	0.68	1.11	0.33
GJR-t	**2.23**	1.75	**2.15**	0.65	1.66	1.12	1.00	1.51	0.75
TG-t	**2.37**	**1.91**	**2.31**	0.81	1.65	1.59	1.00	1.66	0.86
ST-t	**2.21**	**1.91**	**2.15**	0.98	1.65	1.60	1.17	1.67	0.79
G-SKT	1.12	1.46	1.39	0.66	0.49	0.64	0.67	0.92	0.38
GJR-SKT	1.58	1.27	1.84	0.65	0.82	0.80	0.84	1.11	0.44
TG-SKT	1.58	1.59	1.84	0.65	1.32	1.44	0.84	1.32	0.51
ST-SKT	1.58	1.75	**1.99**	0.65	1.15	1.12	0.84	1.30	0.54
G-AL	0.97	0.81	0.95	0.48	0.65	0.49	0.65	0.71	0.34
GJR-AL	1.13	0.97	1.58	0.64	0.81	0.49	0.65	0.89	0.36
TG-AL	1.13	0.97	1.42	0.48	1.13	0.49	0.48	0.87	0.38
STG-AL	1.13	0.81	1.26	0.48	0.97	0.49	0.65	0.83	0.33
G-TW	0.94	0.64	1.23	0.48	0.48	0.48	0.63	0.70	0.40
GJR-TW	1.41	1.11	1.70	0.48	0.79	0.48	0.79	0.97	0.43
TG-TW	1.41	1.59	1.85	0.64	1.11	0.95	0.95	1.21	0.44
ST-TW	1.41	1.27	2.00	0.63	1.11	0.95	0.79	1.17	0.45

Note: Boxes indicate the ratio closest to 1 in that market, bold indicates that the model is rejected by the unconditional coverage test (at a 5% level), for each market.

Table 37: Chapter 3: Loss function for VaR at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	IBM	Mean
G-n	**50.3**	**46.9**	**50.0**	**62.7**	**44.3**	14.7	50.6	**45.6**
GJR-n	45.0	44.6	47.6	57.3	38.3	14.8	48.9	42.3
TG-n	46.1	45.5	49.5	58.2	38.8	14.3	47.9	42.9
STG-n	45.8	44.9	49.1	58.4	39.9	14.6	48.0	43.0
G-t	43.7	45.4	46.4	59.6	38.7	14.8	**53.4**	43.1
GJR-t	41.5	41.7	45.5	55.4	37.9	14.9	52.4	38.5
TG-t	42.2	43.7	45.2	54.5	37.4	14.4	51.8	41.3
ST-t	41.8	43.9	45.6	55.3	37.5	14.3	50.4	41.3
G-SKT	44.1	44.0	46.3	58.9	37.9	14.5	49.4	42.1
GJR-SKT	41.5	43.9	43.9	54.5	36.9	14.4	49.3	40.6
TG-SKT	42.1	42.9	45.6	55.9	37.0	14.3	48.9	40.9
ST-SKT	41.7	43.3	45.6	55.4	36.4	13.9	49.1	40.8
G-AL	44.6	46.7	45.7	57.8	37.9	**15.7**	50.6	42.7
GJR-AL	43.1	47.4	45.1	55.7	37.1	**15.7**	50.1	42.0
TG-AL	42.6	47.7	45.1	55.7	37.4	15.6	50.6	42.1
STG-AL	42.6	47.1	44.8	56.0	38.1	**15.7**	49.2	41.9
G-TW	43.8	44.2	45.6	58.0	37.8	14.7	49.0	41.9
GJR-TW	40.7	43.9	43.3	53.4	34.8	15.1	48.4	40.0
TG-TW	41.6	43.6	44.7	54.9	36.8	14.4	49.8	40.8
ST-TW	41.5	43.9	44.0	53.9	39.8	14.4	49.6	41.0
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	IBM	Mean
G-n	**155.7**	**165.4**	**163.0**	210.6	**123.4**	56.5	174.4	**149.9**
GJR-n	152.4	162.4	160.0	202.4	120.3	56.4	165.2	145.6
TG-n	154.0	163.2	161.3	204.8	121.6	56.6	165.6	146.7
ST-n	154.1	164.4	160.8	204.9	121.7	**57.6**	165.4	147.0
G-t	153.5	164.3	161.0	**211.6**	121.0	55.9	**177.5**	149.2
GJR-t	149.5	161.4	159.7	202.6	120.7	56.5	174.7	146.5
TG-t	150.4	161.4	159.8	202.6	120.6	56.0	174.0	146.4
ST-t	150.7	161.5	159.5	202.9	120.7	56.5	171.2	146.1
G-SKT	154.3	165.1	160.6	208.3	120.1	55.9	168.5	147.5
GJR-SKT	149.9	161.6	159.1	202.0	119.9	56.0	164.6	144.7
TG-SKT	151.1	162.1	160.3	205.4	121.0	56.6	165.6	146.0
ST-SKT	151.1	162.7	160.0	203.3	120.3	55.9	164.2	145.4
G-AL	153.4	164.5	161.0	207.4	119.8	55.4	169.5	147.3
GJR-AL	148.9	161.0	159.7	200.3	119.3	55.5	165.9	144.4
TG-AL	149.5	161.6	160.8	203.0	121.0	56.1	167.0	145.6
ST-AL	149.4	160.9	159.6	202.6	121.1	56.0	165.0	144.9
G-TW	153.9	164.5	160.4	207.8	119.9	55.5	168.9	147.3
GJR-TW	149.5	161.3	160.2	198.3	118.6	55.7	165.7	144.2
TG-TW	150.6	161.8	160.7	203.8	120.6	55.7	167.2	145.8
ST-TW	151.3	161.6	160.0	202.8	123.1	55.4	166.0	145.8

Note: Boxes indicate the loss function is the lowest in that market, bold indicates that the loss function is highest in that market, for each market.

Table 38: Chapter 3: Loss function for ES at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	EUR/US	IBM	Mean
G-n	**27.1**	**23.1**	**24.7**	**29.7**	**27.6**	6.1	**24.2**	**23.2**
GJR-n	21.3	21.7	22.3	25.3	21.1	6.2	23.1	20.1
TG-n	23.3	22.1	23.6	26.3	21.2	6.1	22.2	20.7
STG-n	23.0	21.4	23.3	26.3	22.1	5.9	22.3	20.6
G-t	19.5	20.0	20.7	23.6	19.6	5.9	20.2	18.5
GJR-t	18.5	19.8	20.3	21.1	19.1	5.9	20.1	17.8
TG-t	19.6	20.3	20.5	21.6	18.9	5.8	19.8	18.1
ST-t	19.5	20.5	20.8	22.1	19.3	5.7	19.3	18.2
G-SKT	19.0	19.6	19.6	23.4	19.0	6.0	19.9	18.1
GJR-SKT	18.0	19.8	18.4	21.7	18.2	6.0	19.9	17.4
TG-SKT	18.0	19.6	19.1	21.7	18.6	6.0	19.9	17.6
ST-SKT	17.9	19.7	19.1	21.5	18.1	5.9	19.7	17.4
G-AL	20.0	20.7	19.7	24.5	18.7	**7.1**	21.8	18.9
GJR-AL	18.6	20.8	19.2	24.3	18.4	**7.1**	22.3	18.7
TG-AL	18.6	21.0	18.9	24.2	18.4	7.0	22.5	18.7
STG-AL	18.6	20.8	18.8	24.4	18.4	**7.1**	21.9	18.6
G-TW	19.7	20.1	18.9	24.6	19.2	6.5	21.5	18.7
GJR-TW	17.8	19.9	18.7	23.3	17.7	6.8	21.8	18.0
TG-TW	18.3	20.2	18.7	23.1	18.7	6.4	21.6	18.1
ST-TW	18.5	20.2	18.7	23.4	20.7	6.4	22.1	18.6
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	EUR/US	IBM	Mean
G-n	**79.8**	**80.3**	81.2	**105.0**	**66.3**	26.3	**86.4**	**75.0**
GJR-n	75.1	78.2	80.3	97.1	61.7	26.0	81.9	71.5
TG-n	77.0	79.2	**81.8**	97.9	62.6	26.1	80.7	72.2
STG-n	76.6	79.1	81.5	98.3	63.1	**26.8**	80.8	72.3
G-t	71.8	74.2	78.2	94.5	58.2	24.4	75.5	68.1
GJR-t	70.7	74.1	77.9	90.2	58.1	24.9	75.9	67.4
TG-t	72.2	75.1	78.8	91.7	58.5	24.6	74.9	68.0
ST-t	71.9	75.0	78.9	92.0	58.9	24.9	74.8	68.1
G-SKT	71.3	73.4	77.6	95.1	57.8	24.6	75.6	67.9
GJR-SKT	69.4	72.7	75.0	90.4	56.4	24.5	75.7	66.3
TG-SKT	69.2	72.7	77.1	91.4	57.1	24.6	75.2	66.8
ST-SKT	69.0	73.6	77.0	91.2	55.9	24.3	74.7	66.5
G-AL	71.2	72.1	75.2	96.5	56.8	25.2	80.4	68.2
GJR-AL	68.1	71.3	72.1	91.4	55.6	25.2	80.1	66.3
TG-AL	68.3	71.4	73.0	93.2	56.6	24.8	80.7	66.9
STG-AL	68.2	70.6	73.4	93.1	56.9	25.2	79.1	66.7
G-TW	72.8	73.5	76.7	97.7	58.9	25.0	80.0	69.2
GJR-TW	69.1	73.0	74.8	89.7	55.7	25.0	79.6	66.7
TG-TW	69.5	72.0	76.4	94.1	57.6	24.5	80.9	67.9
ST-TW	69.4	72.8	74.7	92.8	60.8	24.7	80.8	68.0

Note: Boxes indicate the loss function is the lowest in that market, bold indicates that the loss function is highest in that market, for each market.

4 Forecasting Tail Risk via a Partitioned Distribution

The financial time series usually have a single mode, both unconditionally and conditionally. The two-sided Weibull distribution proposed in Chapter 3, although powerful at forecasting tail risks, does not satisfy the uni-modal property of returns. This inapt model fitting to the data may cause a problem in the model selection process; that is, the TW may not be favoured by formal model selection criteria, though it accurately forecast the risk levels.

Thus, the aims of this chapter are to:

1. Address the potential bimodal problem of a two-sided Weibull distribution in modelling the conditional return distribution.

2. Introduce a new distribution for capturing the skewness and heavy tails in the conditional distribution for financial time series.

3. Improve the model fitting of the tails as well as the centre of the conditional return distribution.

For these purposes, we develop a new distribution with Weibull tails and an AL centre with a single mode. In the literature, there are attempts to combine two or more distributions to form a new distribution: Behrens, Lopes, and Gamerman (2004) developed a truncated Gamma with a generalised Pareto distributions in the right tail for estimating the threshold of extreme events. Zhao, Scarrott, Oxley, and Reale (2010) considered a two-stage model, with a GARCH-type model in the first stage, and a normal distribution plus two generalised Pareto distribution in two tails modelling residuals in the second stage. So and Chan (2011) developed a mixture of distributions under a GARCH framework to model the tail asymmetry in financial time series. This mixture of distribution approach partitions the whole conditional distribution into three parts, with a normal distribution and two generalised Pareto distributions for the two tails. The partitioned distribution approach is more efficient in estimating tails than the common mixture distributions, such

as the mixture of Gaussian distributions, because the appropriate candidates for the tails can be pre-determined.

One of the contributions in this chapter is that the distribution proposed in this study adopts both non-Gaussian distributions for the tails and the centre, which is more general and flexbile, unlike those approaches in the studies mentioned above. So and Chan (2011) used standard MCMC and the random-walk MH algorithm for parameter estimations, this study employs an adaptive MCMC and MH algorithm with an adaptive mixture of the Student t (AdMit) of Hoogerheide, Kaashoek and Van Djik (2007) as the proposal density. I derived the formula for a number of properties of this distribution, including pdf, cdf, inverse cdf, VaR and ES functions, as well as the quantile level function for ES. The proposed distribution is used as an assumption for a conditional return distribution, combined with a classic GJR-GARCH volatility model to forecast dynamic financial tail risks.

The proposed model is illustrated by applying it to five return series and producing one-day-ahead VaR and ES forecasts. The forecasting performance is evaluated with formal and non-formal tests; further, the model-fitting performance is assessed by a range of hypothesis tests on the predictive residuals. In general, the proposed model is less favoured in forecasting than the TW distribution, but outperforms its competitors in model fitting, across several criteria.

4.1 A Partitioned Distribution

As the two-sided Weibull is proved to flexibly and adequately model the tails of the conditional return distributions in Chapter 3, Weibull is then a natural choice for the tail component of the partitioned distribution. The conditional distribution of the financial time series also has higher kurtosis than normal, thus an AL distribution is employed as the centre component of the partitioned distribution. This distribution has great flexibility in both tails and the centre.

The partitioned distribution of Weibull and AL (PWAL)'s shape and scale can be tuned by four Weibull parameters, one AL parameter, and two break points. The defini-

tion of a PWAL is, $Y \sim PWAL(\lambda_1, k_1, \lambda_2, k_2, p, \delta_1, \delta_2)$ if:

$$f(y|\lambda_1, k_1, \lambda_2, k_2, p, \delta_1, \delta_2) = \begin{cases} \left(\frac{-y}{\lambda_1}\right)^{k_1-1} \exp\left[-\left(\frac{-y}{\lambda_1}\right)^{k_1}\right] \ ; \ y < \delta_1 \\ \exp\left(\frac{y}{p}\right) \ ; \ \delta_1 \leq y < 0 \\ \exp\left(\frac{-y}{1-p}\right) \ ; \ 0 \leq y < \delta_2 \\ \left(\frac{y}{\lambda_2}\right)^{k_2-1} \exp\left[-\left(\frac{y}{\lambda_2}\right)^{k_2}\right] \ ; \ y \geq \delta_2. \end{cases} \tag{33}$$

where shape parameters satisfy $k_1, k_2 > 0$ and $0 \leq p \leq 1$, scale parameters $\lambda_1, \lambda_2 > 0$, and break points $\delta_1 < 0$, $\delta_2 > 0$.

The mean of the PWAL random variable can be derived as

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} y f(y|\lambda_1, k_1, \lambda_2, k_2, p, \delta_1, \delta_2) dy \\ &= \int_{-\infty}^{\delta_1} y \left(\frac{-y}{\lambda_1}\right)^{k_1-1} \exp\left[-\left(\frac{-y}{\lambda_1}\right)^{k_1}\right] dy + \int_{\delta_1}^{0} y \exp\left(\frac{y}{p}\right) dy \\ &+ \int_{0}^{\delta_2} y \exp\left(\frac{-y}{1-p}\right) dy + \int_{0}^{\infty} y \left(\frac{y}{\lambda_2}\right)^{k_2-1} \exp\left[-\left(\frac{y}{\lambda_2}\right)^{k_2}\right] dy \\ &= \int_{\left(\frac{-\delta_1}{\lambda_1}\right)^{k_1}}^{\infty} -\frac{\lambda_1^2}{k_1} \left[\left(\frac{-y}{\lambda_1}\right)^{k_1}\right]^{\left(\frac{1}{k_1}+1\right)-1} \exp\left[-\left(\frac{-y}{\lambda_1}\right)^{k_1}\right] d\left(\frac{-y}{\lambda_1}\right)^{k_1} + py \exp\left(\frac{y}{p}\right) |_{(\delta_1)}^{(0)} \\ &- p \int_{\delta_1}^{0} \exp\left(\frac{y}{p}\right) dy - (1-p) y \exp\left(\frac{-y}{1-p}\right) |_{(0)}^{(\delta_2)} + (1-p) \int_{\delta_2}^{0} \exp\left(\frac{-y}{1-p}\right) dy \\ &+ \int_{\left(\frac{\delta_2}{\lambda_2}\right)^{k_2}}^{\infty} \frac{\lambda_2^2}{k_2} \left(\frac{y}{\lambda_2}\right) \exp\left[-\left(\frac{y}{\lambda_2}\right)^{k_2}\right] d\left(\frac{y}{\lambda_2}\right)^{k_2} \\ &= -\frac{\lambda_1^2}{k_1} \Gamma\left(1 + \frac{1}{k_1}, \left(\frac{-\delta_1}{\lambda_1}\right)^{k_1}\right) + \frac{\lambda_2^2}{k_2} \Gamma\left(1 + \frac{1}{k_2}, \left(\frac{\delta_2}{\lambda_2}\right)^{k_2}\right) \\ &+ p(p - \delta_1) \exp\left(\frac{\delta_1}{p}\right) - (1-p)(1-p+\delta_2) \exp\left(\frac{-\delta_2}{1-p}\right) + 1 - 2p. \end{aligned} \tag{34}$$

Similarly, the expectation of Y^2 can be derived as

$$\begin{aligned} E(Y^2) &= \int_{-\infty}^{\infty} y^2 f(y|\lambda_1, k_1, \lambda_2, k_2, p, \delta_1, \delta_2) dy \\ &= \frac{\lambda_1^3}{k_1} \Gamma\left(1 + \frac{2}{k_1}, \left(\frac{-\delta_1}{\lambda_1}\right)^{k_1}\right) + \frac{\lambda_2^3}{k_2} \Gamma\left(1 + \frac{2}{k_2}, \left(\frac{\delta_2}{\lambda_2}\right)^{k_2}\right) + p\left(2p\delta_1 - \delta_1^2 - 2p^2\right) \exp\left(\frac{\delta_1}{p}\right) \\ &- (1-p)\left(\delta_2^2 + 2(1-p)\delta_2 + 2(1-p)^2\right) \exp\left(\frac{-\delta_2}{1-p}\right) + 2\left(p^3 + (1-p)^3\right). \end{aligned} \tag{35}$$

Thus, the variance of Y can be expressed as follows by using (34) and (35)

$$\text{Var}(Y) \;=\; E(Y^2) - E^2(Y).$$

4.1.1 Standardized partitioned distribution of the Weibull and asymmetric Laplace

Error distributions in volatility models should be standardised. A standardised PWAL distribution is equivalent to $\frac{Y}{\sqrt{\text{Var}(Y)}}$, which is defined above.

The pdf for an SPWAL random variable $X = \frac{Y}{\sqrt{\text{Var}(Y)}}$, where $Y \sim PWAL(\lambda_1, k_1, \lambda_2, k_2, p, \delta_1, \delta_2)$, is:

$$f(x|\lambda_1, k_1, \lambda_2, k_2, p, \delta_1, \delta_2) = \begin{cases} b_p \left(\frac{-b_p x}{\lambda_1}\right)^{k_1 - 1} \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right] \;;\; x < \frac{\delta_1}{b_p} \\ b_p \exp\left(\frac{b_p x}{p}\right) \;;\; \frac{\delta_1}{b_p} \leq x < 0 \\ b_p \exp\left(\frac{-b_p x}{1-p}\right) \;;\; 0 \leq x < \frac{\delta_2}{b_p} \\ b_p \left(\frac{b_p x}{\lambda_2}\right)^{k_2 - 1} \exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right] \;;\; x \geq \frac{\delta_2}{b_p}, \end{cases} \tag{36}$$

where $b_p^2 = \text{Var}(Y)$.

To ensure the pdf integrates to 1:

$$\frac{\lambda_1}{k_1} \exp\left(-\left(\frac{-\delta_1}{\lambda_1}\right)^{k_1}\right) + \frac{\lambda_2}{k_2} \exp\left(-\left(\frac{\delta_2}{\lambda_2}\right)^{k_2}\right) = p \exp\left(\frac{\delta_1}{p}\right) + (1-p) \exp\left(\frac{-\delta_2}{1-p}\right). \tag{37}$$

To ensure the pdf is continuous at the break points δ_1 and δ_2:

$$\begin{aligned} \left(-\frac{\delta_1}{\lambda_1}\right)^{k_1 - 1} \exp\left(-\left(\frac{-\delta_1}{\lambda_1}\right)^{k_1}\right) &= \exp\left(\frac{\delta_1}{p}\right) \;; \\ \left(\frac{\delta_2}{\lambda_2}\right)^{k_2 - 1} \exp\left(-\left(\frac{\delta_2}{\lambda_2}\right)^{k_2}\right) &= \exp\left(\frac{-\delta_2}{1-p}\right). \end{aligned} \tag{38}$$

Using the first equation in (38), δ_1 can be written into a function of λ_1, k_1 and p. Then given k_2 and solved δ_1, using (37) and the second equation in (38) together, constraints on δ_2 and λ_2 can be obtained. Thus, in this formulation there are only four free parameters, and we write $X \sim SPWAL(\lambda_1, k_1, k_2, p)$ where λ_2, δ_1 and δ_2 are fixed by (37) and (38) as explained above.

The $SPWAL(\lambda_1, k_1, \lambda_2, p)$ has cdf, obtained by direct integration,

$$F(x|\lambda_1,k_1,\lambda_2,k_2,p,\delta_1,\delta_2) = \begin{cases} \frac{\lambda_1}{k_1}\exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right] \;;\; x < \frac{\delta_1}{b_p} \\ \frac{\lambda_1}{k_1}\exp\left[-\left(\frac{-\delta_1}{\lambda_1}\right)^{k_1}\right] - p\exp\left(\frac{\delta_1}{p}\right) + p\exp\left(\frac{b_p x}{p}\right) \;;\; \frac{\delta_1}{b_p} \le x < 0 \\ \frac{\lambda_1}{k_1}\exp\left[-\left(\frac{-\delta_1}{\lambda_1}\right)^{k_1}\right] - p\exp\left(\frac{\delta_1}{p}\right) \\ +1-(1-p)\exp\left(\frac{-b_p x}{1-p}\right) \quad ;\; 0 \le x < \frac{\delta_2}{b_p} \\ \frac{\lambda_1}{k_1}\exp\left(-\left(\frac{-\delta_1}{\lambda_1}\right)^{k_1}\right) + \frac{\lambda_2}{k_2}\exp\left(-\left(\frac{\delta_2}{\lambda_2}\right)^{k_2}\right) - p\exp\left(\frac{\delta_1}{p}\right) \\ -(1-p)\exp\left(\frac{-\delta_2}{1-p}\right) + 1 - \frac{\lambda_2}{k_2}\exp\left(-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right) \quad ;\; x \ge \frac{\delta_2}{b_p} \end{cases} \tag{39}$$

In this parameterization,

$$c_{sk} = Pr(X<0) = \frac{\lambda_1}{k_1}\exp\left[-\left(\frac{-\delta_1}{\lambda_1}\right)^{k_1}\right] - p\exp\left(\frac{\delta_1}{p}\right) + p. \tag{40}$$

Thus, when $c_{sk} < 0.5$, the density is positively skewed to the right, while negative or left skewness occurs when $c_{sk} > 0.5$.

The inverse cdf or quantile function of an SPWAL is:

$$F^{-1}(\alpha|\lambda_1,k_1,\lambda_2,k_2,p,\delta_1,\delta_2) = \begin{cases} -\frac{\lambda_1}{b_p}\left[-\log\left(\frac{k_1}{\lambda_1}\alpha\right)\right]^{\frac{1}{k_1}} \;;\; 0 \le \alpha < F\left(\frac{\delta_1}{b_p}\right) \\ \frac{p}{b_p}\log\left[\begin{array}{l} \frac{\alpha}{p} + \exp\left(\frac{\delta_1}{p}\right) \\ -\frac{\lambda_1}{pk_1}\exp\left(-\left(\frac{-\delta_1}{\lambda_1}\right)^{k_1}\right) \end{array}\right] \;;\; F\left(\frac{\delta_1}{b_p}\right) \le \alpha < F(0) \\ -\frac{1-p}{b_p}\log\left[\begin{array}{l} \frac{\lambda_1}{(1-p)k_1}\exp\left(-\left(\frac{-\delta_1}{\lambda_1}\right)^{k_1}\right) \\ -\frac{p}{1-p}\exp\left(\frac{\delta_1}{p}\right) + \frac{1-\alpha}{1-p} \end{array}\right] \;;\; F(0) \le \alpha < F\left(\frac{\delta_2}{b_p}\right) \\ \frac{\lambda_2}{b_p}\left\{-\log\left[\frac{k_2}{\lambda_2}\left[\begin{array}{l} \frac{\lambda_1}{k_1}\exp\left(-\left(\frac{-\delta_1}{\lambda_1}\right)^{k_1}\right) \\ +\frac{\lambda_2}{k_2}\exp\left(-\left(\frac{\delta_2}{\lambda_2}\right)^{k_2}\right) \\ -p\exp\left(\frac{\delta_1}{p}\right) + 1 - \alpha \\ -(1-p)\exp\left(\frac{-\delta_2}{1-p}\right) \end{array}\right]\right]\right\}^{\frac{1}{k_2}} \\ \qquad\qquad ;\; \alpha \ge F\left(\frac{\delta_2}{b_p}\right) \end{cases} \tag{41}$$

The mean of an SPWAL, μ_X, is defined as:

$$\mu_X = \frac{1}{b_p}\left[\begin{array}{l} -\frac{\lambda_1^2}{k_1}\Gamma\left(1+\frac{1}{k_1}, \left(\frac{-\delta_1}{\lambda_1}\right)^{k_1}\right) + \frac{\lambda_2^2}{k_2}\Gamma\left(1+\frac{1}{k_2}, \left(\frac{\delta_2}{\lambda_2}\right)^{k_2}\right) \\ +p(p-\delta_1)\exp\left(\frac{\delta_1}{p}\right) - (1-p)(1-p+\delta_2)\exp\left(\frac{-\delta_2}{1-p}\right) + 1 - 2p \end{array}\right]. \tag{42}$$

Thus $Z = X - \mu_X$ has a shifted $SPWAL(\lambda_1, k_1, k_2, p)$ distribution with mean 0 and variance 1.

Figure 33 demonstrates a SPWAL density, a STW, and an AL densities with similar skewness, and their log densities. It is clear that at the extreme lower quantiles, the SPWAL can have tails between the tails of AL and STW, though very close to each other due to the scale.

4.1.2 VaR and tail conditional expectations for the standardized PWAL

Given the definitions of VaR and ES in (3) and (4) in Chapter 2, the VaR is thus simply the quantile given in (41) for SPWAL.

In practice, c_{sk} in (40) is estimated much closer to 0.5 than α, and in fact, $\frac{\lambda_1}{k_1}\exp\left[-\left(\frac{-\delta_1}{\lambda_1}\right)^{k_1}\right]$ is estimated closer to 0.5 than α as well; since risk management focuses on only the extreme tails of returns, particularly the cases $\alpha \leq 0.05$, thus only the case $\alpha < \frac{\lambda_1}{k_1}\exp\left[-\left(\frac{-\delta_1}{\lambda_1}\right)^{k_1}\right]$ in (41) is relevant here. In this context, the tail expectation of a SPWAL for the long position is the same as the ES for the STW derived in (26):

$$\mathrm{ES}_\alpha \;=\; \frac{-\lambda_1^2}{\alpha b_p k_1}\Gamma\left(1+\frac{1}{k_1},\left(\frac{-b_p VaR_\alpha}{\lambda_1}\right)^{k_1}\right)\;;\; 0 \leq \alpha < F\left(\frac{\delta_1}{b_p}\right) \tag{43}$$

where $\Gamma(s,x) = \int_x^\infty t^{s-1}e^{-t}dt$ is the upper incomplete gamma function.

The ES of SPWAL for short position is the same as that of the TW in (27).

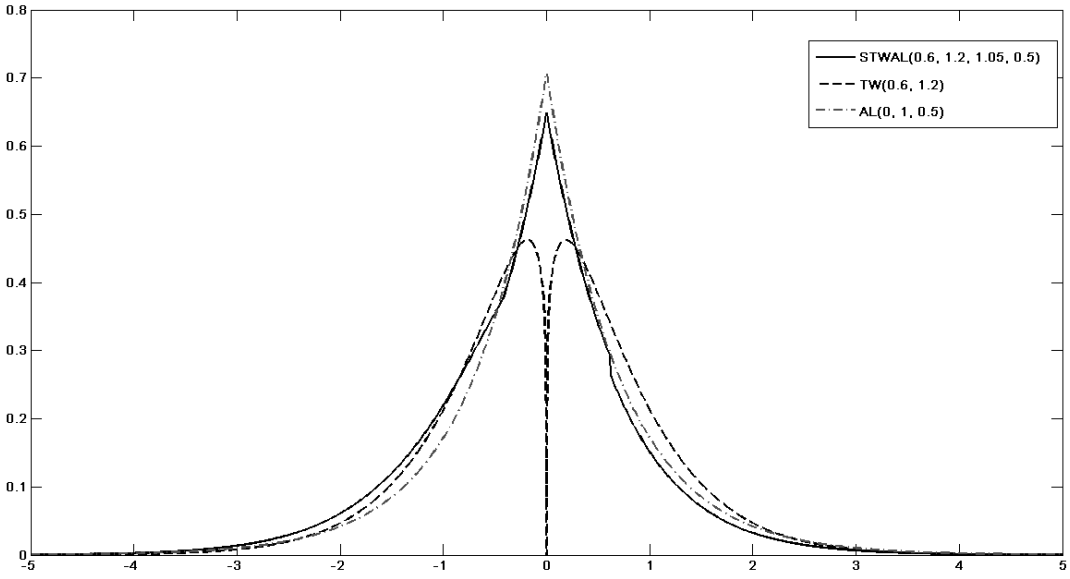
4.2 Model Specification and Estimation Methodology

This section specifies the models and discusses the Bayesian methods and MCMC procedures for estimating parameters and generating forecasting.

4.2.1 Volatility model

In Chapter 3, we discuss general non-linear volatility models for financial returns, as well as their impacts on VaR and ES forecasting. Since the GARCH volatility specifications are not as important as correct specification of conditional return distribution, and the SPWAL distribution is the focus of this chapter, we will focus on the GJR-GARCH model here, which is seen to be very effective in the empirical study in Chapter 3.

Densities



Log densities

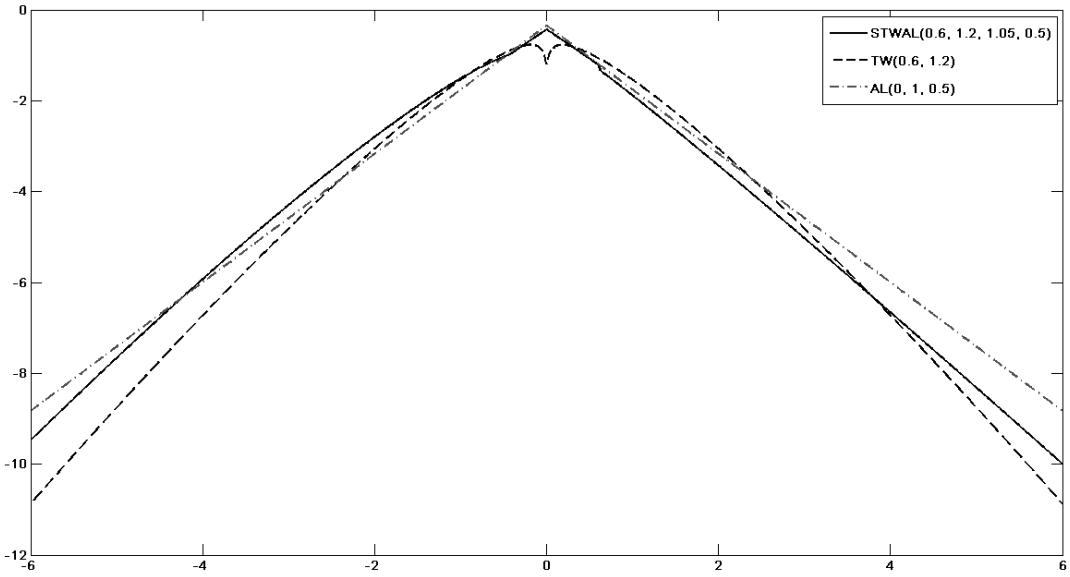


Figure 33: Chapter 4: SPWAL, TW, and AL densitites and log densities.

We follow the common assumption that the mean of a return series is (well approximated as) zero. The GJR-GARCH model for a financial return series $\mathbf{y}$ is:

$$\begin{aligned} y_t &= (\epsilon_t - \mu_\epsilon)\sqrt{h_t} \quad , \ \epsilon_t \overset{\text{i.i.d.}}{\sim} D(1), \\ h_t &= \alpha_0 + (\alpha_1 + \alpha_2 I(y_t < 0))\, y_{t-1}^2 + \beta_1 h_{t-1}, \end{aligned} \tag{44}$$

where $\text{Var}(y_t|\Omega_t) = h_t$ is the conditional variance and D is the conditional distribution and has variance 1 and mean μ_ϵ (often 0). The VaR and ES in this model are:

$$\text{VaR}_{t+1} = D_\alpha^{-1}\sqrt{h_{t+1}} \ ; \ \text{ES}_{t+1} = \text{ES}_\alpha^D \sqrt{h_{t+1}}, \tag{45}$$

where D_α^{-1} is the inverse cdf of D, and ES_α^D is the expected shortfall of D, at the $\alpha \times 100\%$ level. The Gaussian, Student t, the AL of Chen et al. (2011), the STW distribution, and the SPWAL are considered. Except for the first two, all have non-zero means that are subtracted in (44). Expressions for ES_α^D in the Gaussian and Student t cases can be found in McNeil, Frey and Embrechts (2005, pg. 45, 46), while for the AL see Chen et al (2011) and Chapter 2, and for the STW see Chapter 3, Section 1.

4.3 Estimation Methodology and Simulation

This section describes the Bayesian methods and MCMC procedure for estimating parameters and producing forecasting. Sufficient conditions for positivity and stationarity (also necessary) are

$$\begin{gathered} \alpha_0 > 0 \ ; \ 0 \leq \alpha_1 + \beta_1 + c_p\gamma < 1 \\ \alpha_1, \beta_1, \alpha_1 + \alpha_2 \geq 0, \end{gathered}$$

where $c_p = 0.5$, which applies whenever D is symmetric. Chen et al. (2011) and Appendix 1 derived expressions for c_p, in the case of the GJR-GARCH-AL model; c_p for the GJR-GARCH-STW is presented in Appendix 3. Again, although it will be better to estimate the c_p instead using 0.5 in the MCMC, as in practice, the values of c_p are usually very close to 0.5, indicating that the above condition is sufficient for stationarity most of time.

4.3.1 Bayesian estimation methodology

This section specifies the Bayesian methods and MCMC procedures for estimating parameters and generating forecasts.

In a Bayesian analysis, a likelihood function and a prior are usually required. The required likelihood follows from the choice of error distribution D and equation (44). We consider the priors for the GJR-GARCH model with SPWAL errors.

As in Chapter 3, we choose an uninformative prior over relevant ranges for each parameter. For the GJR-GARCH parameters $\boldsymbol{\theta}$:

$$\pi(\boldsymbol{\theta}) \propto I\left(\alpha_0 > 0, \alpha_1 + \beta_1 + c_p\alpha_2 < 1, \alpha_1, \beta_1, \alpha_1 + \alpha_2 \geq 0\right).$$

Then for SPWAL, the parameters λ_1, k_1, k_2, p have a flat prior:

$$\pi(\lambda_1, k_1, k_2, p) \propto I\left(0 < \lambda_1 < k_1, k_2 > 0, 0 < p < 1\right).$$

4.3.1.1 Adaptive MCMC and AdMit

None of the parameter groupings in the above section have a standard recognisable conditional posterior density and, as such, Metropolis and Metropolis-Hastings methods are required. Chapter 2 and Chapter 3 illustrate the efficiency and speed of mixing gains from employing an adaptive scheme where iterates in the burn-in period, simulated from standard random-walk Metropolis methods with tuning to achieve desired acceptance rates, are used to build a mixture of Gaussian proposal densities for use in the sampling period. Here, this method is adapted by replacing the mixture of Gaussian proposals in the sampling period with the more general and flexible 'AdMit' mixture of Student t proposal procedure, proposed by Hoogerheide, Kaashoek and van Dijk (2007).

The sampling scheme for $\boldsymbol{\theta}$ follows:

***Step*1.** RWM with mixture of Gaussian proposal during the burn-in period: $\boldsymbol{\theta}|\Omega$, where Ω is the information set. At MCMC iteration j,

$$\boldsymbol{\theta}^p = \boldsymbol{\theta}^{(j-1)} + \boldsymbol{\varepsilon}\ ;\ \ \boldsymbol{\varepsilon} \sim \rho N(\mathbf{0}, \text{diag}\{\boldsymbol{c_1}\}) + (1-\rho)N(\mathbf{0}, \omega\text{diag}\{\boldsymbol{c_1}\}),$$

where $\rho = 0.95$, $\omega = 100$, and $\boldsymbol{c_1}$ is a vector of positive numbers, all jointly tuned so that the observed acceptance rate lies $\in (0.15, 0.5)$, as recommended by Gilks et al. (1996). The proposed $\boldsymbol{\theta}^p$ is accepted as $\boldsymbol{\theta}^{(j)}$ with the usual RWM probability. Step 1 is iterated, at least until convergence is achieved, for $j = 1, \ldots, J$, and these iterates make up the burn-in MCMC sample.

Step2. We follow the steps in Hoogerheide et al. (2007) to determine the components in the mixture of Student t densities for the parameter. Compute the modes $\mu_h^{\boldsymbol{\theta}}$, scale or co-variance matrix $\Sigma_h^{\boldsymbol{\theta}}$ and weight p_h for each component of the mixture of Student t densities by optimizing the approximation of the proposal density to the target density kernal $p(\boldsymbol{\theta})$. Then the mixture of Student t proposal density is $nn(\boldsymbol{\theta}) = \sum_{h=1}^{H} p_h t\left(\boldsymbol{\theta}|\mu_h^{\boldsymbol{\theta}}, \Sigma_h^{\boldsymbol{\theta}}, \nu_h\right)$, where ν_h is the degrees of freedom for hth component, and $h = 1, \ldots, H$, where H is the number of components. In this study we use a mixture of Student t distribution with two components, as in most cases in the empirical study, one of the components already has a very large weight (over 0.85 on average); thus, the probability for a third component will be far more smaller than those of the other two.

The MCMC sampling period employs an IKMH algorithm.

Step3. Re-start the MCMC using an IK mixture-t proposal MH method for $\boldsymbol{\theta}|\Omega$. At MCMC iterate j :

$$\boldsymbol{\theta}^p \sim nn(\boldsymbol{\theta}).$$

Then accept $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^p$ with the usual IKMH probability. This step is iterated for $j = J+1, \ldots, J+K$, where K is chosen as sufficiently large for reliable inference.

4.3.2 Simulation study

A small scale simulation study is carried out to examine the effectiveness of the proposed AdMit MCMC sampling procedure.

4.3.2.1 Simulation Study for Mixture of Gaussian and Mixture of Student's t

Sample series of size $n = 2000$, the same size as the real data we consider in this chapter, are simulated from a GJR-GARCH-AL model:

$$\begin{aligned} \epsilon_t &\overset{\text{i.i.d.}}{\sim} AL(0, 1, 0.55), \\ \sigma_t^2 &= 0.05 + (0.05 + 0.05I\left(y_{t-1} < 0\right)) y_{t-1}^2 + 0.85\sigma_{t-1}^2, \\ y_t &= (\epsilon_t - \mu_{\epsilon_t})\sigma_t. \end{aligned}$$

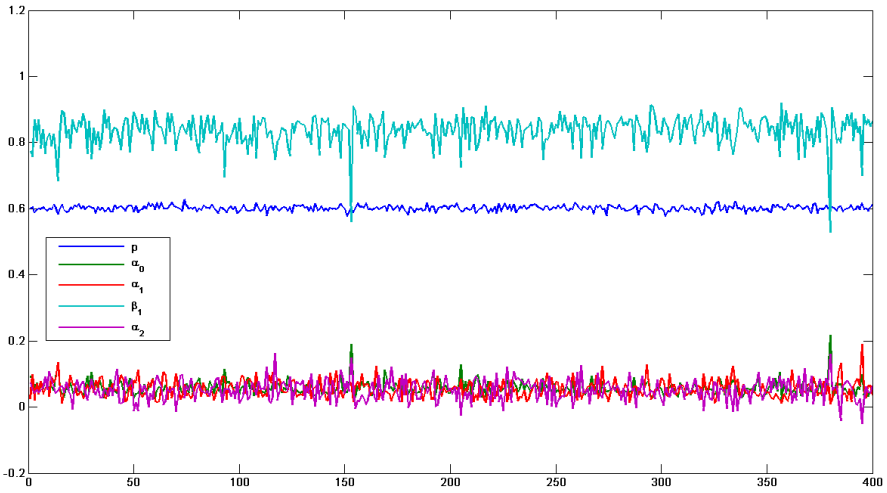
The adaptive MCMC method in the above section is used to estimate the parameters for each of 400 data replications. For each replicate the MCMC sample has 15 000 iterations in total: the first 5000 are the burn-in period using Step 1, and are discarded, as the MCMC chain is converged and stable after 5000 iterations; while Step 2 is used to generate the final sample of 10 000 iterates, used for estimation and inference. MCMC convergence is assessed via extensive examination of trace plots from multiple and varied starting positions for each parameter, from which the burn-in and sampling period sizes are chosen: in all simulated and real data examples we observed MCMC convergence well inside 5000 iterations.

The averages of the 400 posterior mean estimates ('Mean'), the standard errors ('Std') of the 400 estimates, and the observed coverage rate for the 100 95% credible intervals are shown. The summary statistics from the MCMC with a mixture of Gaussian (MiG) proposal are also presented for comparison. It is clear that the estimate from mixture of t (Mit) is closer to the true values, which are in bold.

From Table 39, the estimation of mixture of Gaussian and mixture of t are very close and accurate; Mit is marginally better than MiG, in terms of averages and standard errors of estimates. We note that the coverage rates for all parameters from MiG are higher than those from Mit. One possible explanation for this issue is that, both the components of MiG have modes centred around the sample mean, while one of the components of MiT may have a mode away from the sample mean. It may be improved by adding a third component in the mixture of Student t. However this won't be a problem for forecasting.

Figure 34 shows the estimates of 100 replications from MCMC with Mit and MiG

AdMit



Mixture of Gaussian

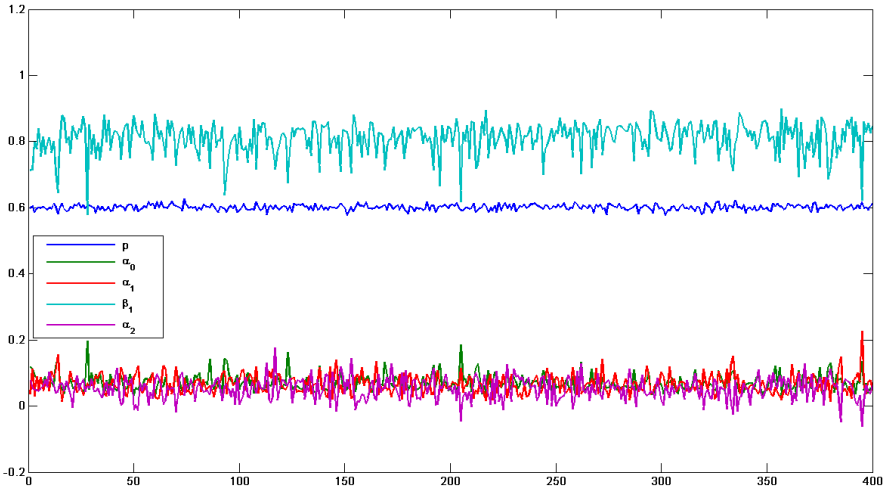


Figure 34: Chapter 4: Estimates of the GJR-GARCH-AL parameters using AdMit and MiG proposals.: 400 replications

Table 39: Chapter 4: Summary statistics for parameter estimates from 400 simulated data sets from the GJR-GARCH-AL model.

		Mean		Standard error		Coverage	
par.	True	MiG	Mit	MiG	Mit	MiG	Mit
p	0.6	**0.601**	**0.601**	**0.008**	**0.008**	**0.945**	0.935
α_0	0.05	0.070	**0.058**	0.024	**0.021**	**0.878**	0.783
α_1	0.05	0.065	**0.055**	0.028	**0.026**	**0.933**	0.763
γ_1	0.05	0.051	**0.050**	0.034	**0.031**	**0.965**	0.793
β_1	0.85	0.808	**0.835**	0.049	**0.045**	**0.865**	0.783

methods. The typical MCMC running times for Mit methods for the GJR-GARCH-AL model are around 20 seconds for the burn-in period, 20 seconds for the process of determining the Mit components, and 25 seconds for the sampling period, that is around 1 minute in total. The typical MH acceptance rates are between 15% and 35% in the burn-in period, and between 23% and 35% in the sampling period. Recall the simulation results in Chapter 2, the typical MCMC running times for MiG methods for the GJR-GARCH-AL model is around 1.5 minutes for the whole MCMC process, the typical MH acceptance rates are between 5% and 17% for the burn-in period, and between 50% and 68% in the sampling period. It is clear that by using the Mit, the actual MCMC running time is reduced, but extra time may be required to determine the Mit components, thus the running time for the whole estimation process may increase. Further, the Mit improved the typically MH acceptance rates in the burn-in period.

4.3.2.2 Simulation Study for GJR-GARCH-SPWAL

Sample series of size $n = 2000$, the same size as the real data we consider in this chapter, are simulated from a GJR-GARCH-SPWAL model:

$$\begin{aligned} \epsilon_t &\overset{\text{i.i.d.}}{\sim} SPWAL(0.6, 1.1, 1.05, 0.6), \\ \sigma_t^2 &= 0.05 + (0.05 + 0.05I\left(y_{t-1} < 0\right)) y_{t-1}^2 + 0.85\sigma_{t-1}^2, \\ y_t &= (\epsilon_t - \mu_{\epsilon_t})\sigma_t. \end{aligned}$$

The summary statistics from the MCMC with AdMit method are presented in Table 40. The estimates of all the parameters, except δ_1 are close to the true values. The

coverage rates for all the parameters except p are not close to 95. This relatively poor estimation may be due to the constraint on parameters in (37), which can be solved by relieving this constraint, taking the integration of the pdf of PWAL in (33), and scaling (33) accordingly to obtain the pdf of standardized PWAL.

Table 40: Chapter 4: Summary statistics for parameter estimates from 100 simulated data sets from the GJR-GARCH-PWAL model.

par.	True	Mean	Std	Coverage
λ_1	0.6	0.608	0.029	0.77
k_1	1.1	1.110	0.076	0.71
k_2	1.05	1.041	0.066	0.83
p	0.6	0.577	0.083	0.94
λ_2	0.43	0.428	0.026	0.83
δ_1	-0.6	-0.301	0.105	0.42
δ_2	0.17	0.212	0.073	0.82
α_0	0.05	0.054	0.021	0.73
α_1	0.05	0.048	0.023	0.75
γ_1	0.05	0.047	0.026	0.80
β_1	0.85	0.830	0.039	0.75

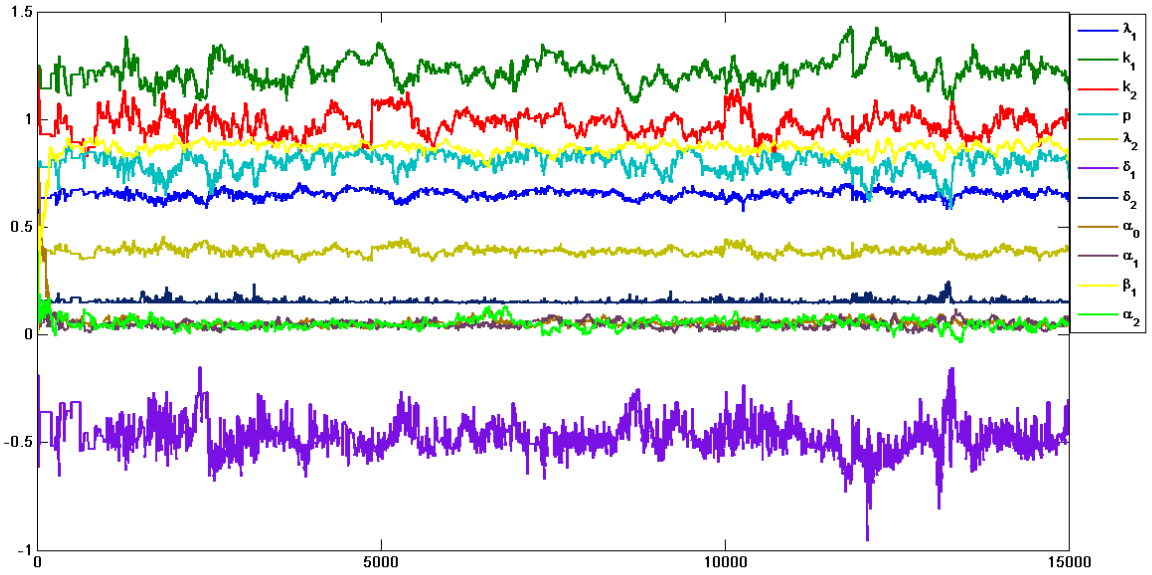
Figure 35 shows the MCMC plots in the burn-in and sampling period with a mixture of Student t method for the GJR-SPWAL model. The typical MCMC running times for the GJR-SPWAL model in Matlab are around 7 minutes for the burn-in period and less than 7 minutes for the sampling period, currently it takes around 2 minutes to determine the components of Mit. The typical MH acceptance rates are between 30 % and 35% for burn-in period, and between 9% and 14% in the sampling period.

Figure 36 shows the estimates from 100 replications of the GJR-PWAL model with the AdMit method.

4.3.3 VaR and ES forecasts

One-step-ahead forecasting is considered. The GJR-GARCH model in (44) provides one-step-ahead forecasts of volatility based on known parameter values. In MCMC methods, at each stage the entire parameter vector, denoted $\boldsymbol{\theta}$, has values simulated for it from

Burn-in period



Sampling period

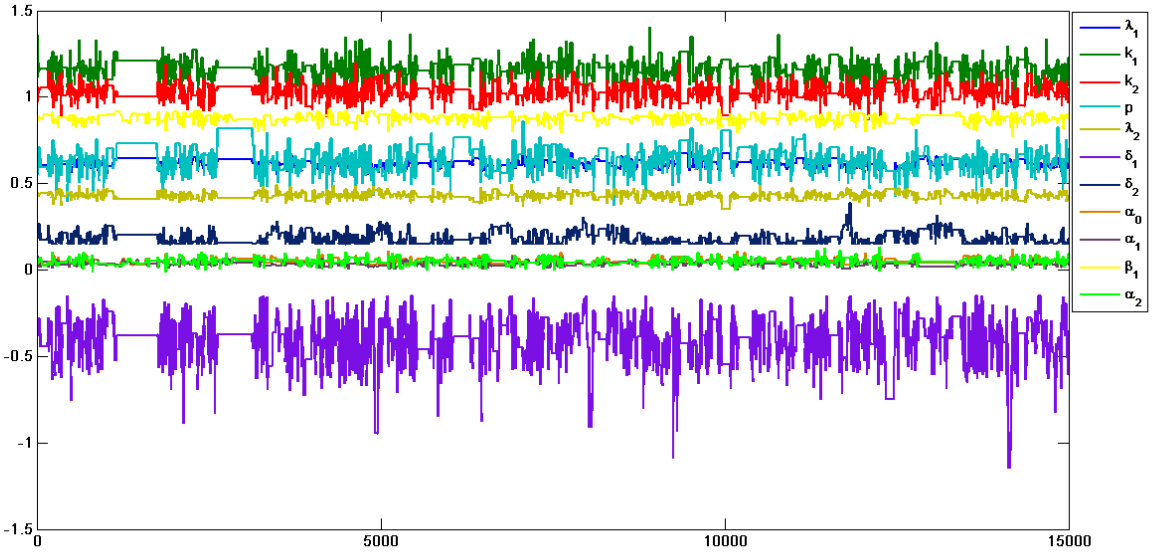


Figure 35: Chapter 4: Simulation study of the GJR-GARCH-PWAL model using AdMit.

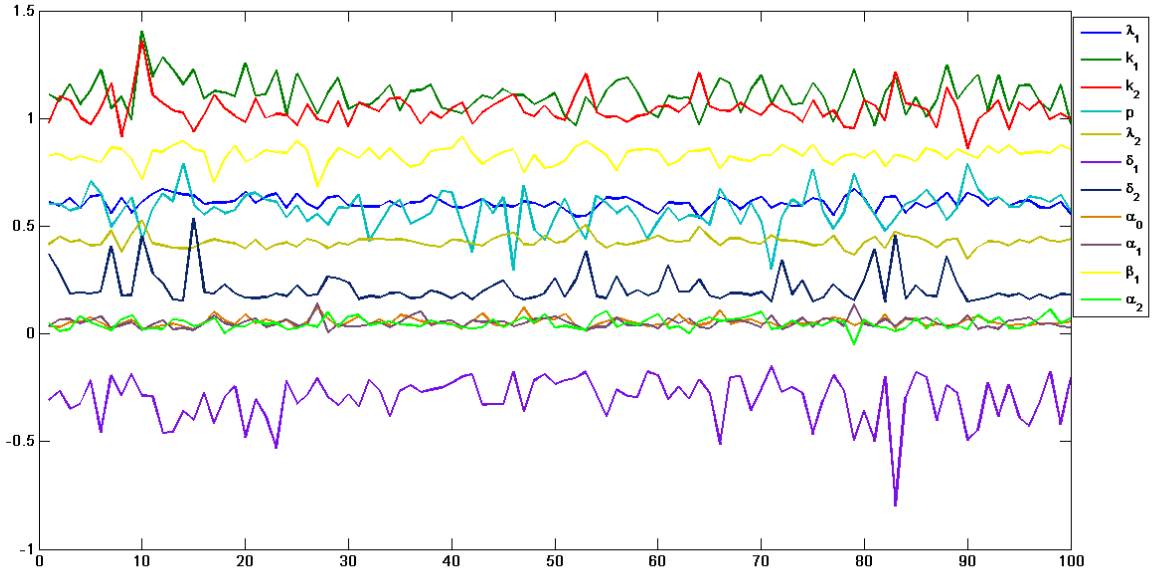


Figure 36: Chapter 4: Estimates from 100 replications of the GJR-GARCH-PWAL model using AdMit.

the posterior, combining to give a Monte Carlo sample $\boldsymbol{\theta}^{[1]}, \ldots, \boldsymbol{\theta}^{[N]}$, where N is the MC sample size. Each of these iterates provides a one-step-ahead forecast of h_t, which can be combined with (45) via, e.g. (41) and (43) for SPWAL errors, to give MC iterate forecasts of VaR and ES, i.e. $\text{VaR}^{[i]}$, $\text{ES}^{[i]}$ for $i = 1, \ldots, N$, for each model. These are simply averaged over the iterates in the sampling period of the MCMC scheme, to give a one-step-ahead forecast of VaR and ES for each model. This GJR-GARCH-SPWAL model is compared with: the GJR-GARCH model with Gaussian error (denoted as GJR-n), and Student t errors (denoted as GJR-t), and AL errors (denoted as GJR-AL), and STW errors (denoted as GJR-TW); and a semi-parametric nonlinear model called the asymmetric slope CAViaR (denoted as CAV; see Engle and Manganelli, 2004) as defined by (18) in Chapter 2, and the popular RiskMetrics model (RM) given in (19) in Chapter 2. The non-parametric historical methods (LT and ST models) are dropped because of their poor performance in risk forecasting in Chapter 2.

All the model estimation and forecasting are again realized in Matlab; I designed the code for GJR-SPWAL; GJR-n and GJR-t are again estimated by using the GARCH Toolbox in the Matlab; the code for CAViaR model is again provided by Dr. Richard

Gerlach.

The VaR and ES models are evaluated via the same standard backtesting methods employed in Chapter 2 and Chapter 3. For backtesting ES models, the ES quantile levels at $\alpha = 0.01, 0.05$ for Gaussian and AL distributions refers to Table 10 in Chapter 2.

The nominal levels for ES for the SPWAL distribution are approximated via the average estimated parameters during the forecast sample. In this way, the standard VaR backtesting methods can be applied to testing ES models as well.

4.4 Empirical Study

4.4.1 Data

The model is illustrated by applying it to daily return series from four international stock-market indices: the S&P 500 (US); FTSE 100 (UK); AORD All Ordinaries Index (Australia); the HANG SENG Index (Hong Kong); as well as one exchange rate series: the AU dollar to the US dollar. The data are obtained from Yahoo! Finance, covering twelve years, January 1998 to January 2010. The daily percentage log return series is $y_t = (ln(P_t) - ln(P_{t-1})) \times 100$, where P_t is the closing price/value on day t.

As in Chapter 3, the sample is initially divided into two periods: the period from January 1998 to December 2005, roughly the first 2000 returns, and is used as an initial learning period. The data from January 2006 to January 2010 are used as the forecasting period. The forecast sample sizes vary from 995 to 1050 days, due to different trading day holidays, etc., The forecast period is mostly more volatile and more fat-tailed because it contains the entire GFC, as the summary statistics show in Table 41. The plots and histograms of the return series are shown in Figures 20 to 24. The estimation results in each series again are mostly as expected and again coincide with the literature: high volatility persistence ($\alpha_1+\beta_1$); fat-tailed (e.g. $\nu < 10$ in Student t error model), and mildly negatively skewed (e.g. $\lambda_1/k_1 > 0.5$ in STW, $p > 0.5$ in AL) conditional distributions. It is to be noted that the skewness is determined in (40), which is a combination of TW skewness and AL skewness, and it is not necessary here for $\lambda_1/k_1 > 0.5$ and $p > 0.5$ to stand for negative skewness. In this case, both λ_1/k_1 and p are usually estimated as < 0.5

Table 41: Chapter 4: Summary statistics of five return series.

Index	Period	Mean	Std	Skewness	Kurtosis	Min	Max
Aus	98-05	0.028	0.73	-0.53	7.12	-5.85	3.39
	06-10	0.007	1.36	-0.54	7.23	-8.55	5.36
US	98-05	0.013	1.20	0.00	5.36	-7.04	5.57
	06-10	-0.011	1.65	-0.22	11.44	-9.47	10.96
UK	98-05	0.003	1.20	-0.11	5.19	-5.59	5.90
	06-10	-0.001	1.53	-0.10	10.00	-9.26	9.38
HK	98-05	0.019	1.64	0.20	8.64	-9.29	13.40
	06-10	0.033	2.13	0.09	9.26	-13.58	13.41
AU/US	98-05	0.006	0.72	-0.19	5.53	-4.45	4.82
	06-10	0.022	1.13	-0.72	15.10	-8.21	7.70

The initial learning period parameter estimates and standard errors (in brackets) from the GJR-GARCH-SPWAL model are given in Table 42.

4.4.2 VaR forecast comparison

Table 43 shows the ratios, and their summaries, of observed VRates to the true nominal levels $\alpha = 0.01, 0.05$ across all series; summaries shown are average ('Mean') and deviation ('Std') for each model and series. 'Std' is the square root of the average squared distance of the observed ratio away from the expected ratio of 1. For each series, the ratio closest to 1 is boxed, while the mean ratio and deviation closest to 1 over the models, for each series, is also boxed. Violation ratios that are significantly different from 1, at a 5% significance level by the UC test, are in bold. These ratios and their average are also presented in Figure 37.

First, it is clear that the GJR-TW and GJR-SPWAL models are close in violation rates to each other. At $\alpha = 0.01$, the RM and GJR-n models, which are with Gaussian errors, under-estimate risk levels in all series, and their average VRates are more than double the nominal 1%. On the other hand, the GJR-AL model over-estimates risk levels in all series with average VRates half of the nominal 1%; it is even rejected by the unconditional coverage test in two series for being too conservative; this is in agreement with the results in Chapter 2 and Chapter 3. The CAV and GJR-TW models are most

Table 42: Chapter 4: GJR-GARCH-SPWAL parameter estimates and standard errors (bracketed).

	Aus	US	UK	HK	AU/US
λ_1	0.553	0.629	00.579	0.514	0.499
	(0.062)	(0.107)	(0.071)	(0.039)	(0.043)
k_1	1.271	1.570	1.406	1.226	1.160
	(0.114)	(0.168)	(0.121)	(0.074)	(0.064)
k_2	1.754	1.714	1.958	1.460	1.627
	(0.158)	(0.207)	(0.204)	(0.088)	(0.109)
p	0.365	0.486	0.361	0.369	0.357
	(0.138)	(0.240)	(0.170)	(0.137)	(0.094)
λ_2	0.598	0.576	0.624	0.603	0.629
	(0.060)	(0.101)	(0.067)	(.035)	(0.043)
δ_1	-0.222	-0.412	-0.268	-0.239	-0.203
	(0.159)	(0.254)	(0.144)	(0.126)	(0.066)
δ_2	0.543	0.413	0.526	0.456	0.541
	(0.128)	(0.212)	(0.142)	(0.110)	(0.095)
α_0	0.010	0.013	0.011	0.014	0.011
	(0.002)	(0.004)	(0.003)	(0.005)	(0.004)
α_1	0.009	0.014	0.021	0.017	0.025
	(0.007)	(0.008)	(0.011)	(0.008)	(0.011)
α_2	0.115	0.141	0.095	0.074	0.021
	(0.010)	(0.015)	(0.017)	(0.015)	(0.011)
β_1	0.918	0.912	0.922	0.942	0.943
	(0.009)	(0.007)	(0.010)	(0.009)	(0.014)

favoured with VRates close to the nominal 1% in two series, respectively; however, the performance of CAV is not as consistent as the GJR-TW, and it is rejected for being too risky in the US market. The GJR-PWAL is the model running after the best performing model if it is not the best, in each series. It also ranks highest in terms of its average VRate ratios closest to 1 (1.07), closely followed by the GJR-t with 1.08; the GJR-TW ranks the third in average VRates but it is close to the nominal 1% from below. The GJR-TW has the minimum deviation from 1 (0.23), followed by the GJR-PWAL model. These two models generally have much lower deviation than the other models. In general, at $\alpha = 0.01$, GJR-TW and GJR-PWAL perform more consistently than other models, and have VRate ratios closely around 1 across five series.

Table 43: Chapter 4: Ratios of $\hat{\alpha}/\alpha$ at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	Mean	Std.
CAViaR	1.52	**2.33**	1.54	1.11	1.17	1.53	0.69
RM	**2.18**	**2.72**	**2.40**	**1.71**	**1.85**	2.17	1.23
GJR-n	**2.65**	**2.53**	**2.50**	**1.81**	**2.14**	**2.33**	**1.36**
GJR-t	1.23	1.07	**1.83**	0.60	0.68	1.08	0.45
GJR-AL	0.47	**0.39**	0.48	0.60	**0.39**	0.47	0.54
GJR-TW	1.04	0.88	1.25	0.70	0.68	0.91	0.23
GJR-PWAL	0.95	1.07	1.44	1.21	0.68	1.07	0.26
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	Mean	Std.
CAViaR	**1.61**	**1.34**	1.19	**1.37**	1.03	**1.31**	**0.36**
RM	1.14	1.25	1.21	1.13	1.11	1.17	0.17
GJR-n	**1.44**	**1.28**	**1.31**	**1.31**	1.09	1.29	0.31
GJR-t	1.21	1.07	1.17	0.87	0.84	1.03	0.16
GJR-AL	1.02	0.95	0.85	1.01	0.84	0.93	0.10
GJR-TW	1.23	1.11	1.10	1.13	0.88	1.09	0.15
GJR-PWAL	1.18	1.11	1.15	1.09	0.94	1.09	0.12

Note: Boxes indicate that the ratio is closest to 1 in that market, bold indicates that the model is rejected by the UC test (at a 5% leve), for each market.

It is clearer for $\alpha = 0.05$ that GJR-AL dominates the VaR forecasting, favoured by VRates close to nominal the 5% in three series out of five, followed by CAV and GJR-TW. The GJR-n still under-predicts risk levels and is rejected by the UC test in most series, while RM has improved performance than that at $\alpha = 0.01$ and is not rejected in any series. The performance of CAV is still not consistent and worse than that at $\alpha = 0.01$; it under-predicts risk levels in three series and has an average VRate furthest from the nominal 5% and the highest deviation of VRate ratios from 1. The GJR-t has its average VRate closest to the nominal 5%, followed by GJR-TW and GJR-PWAL; the GJR-AL has the minimum deviation, with GJR-PWAL ranking the second. The GJR-PWAL again usually ranks the second or the third in VRate ratios close to 1.

Figure 38 shows the 1% VaR forecasts from a number of models. It is clear that the forecasts are in the order of CAViaR, GJR-t, GJR-PWAL, and GJR-AL. The forecasts of GJR-t and GJR-PWAL are very close to each other most of time.

Table 44 shows counts of the number of rejections for each model, at a 5% significance level, across the five series, under the three formal backtests: the UC test, the CC test

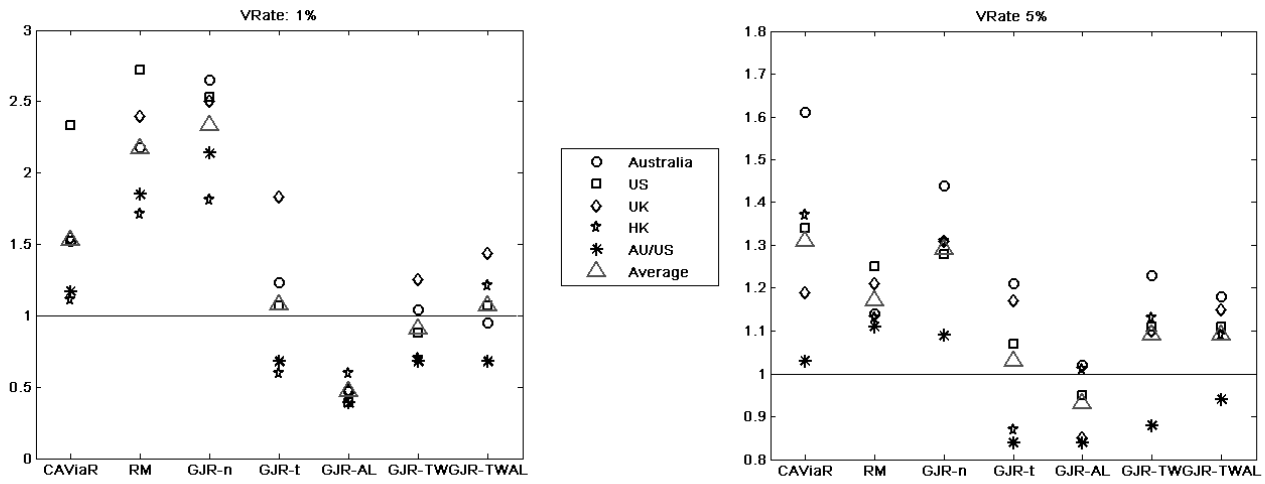
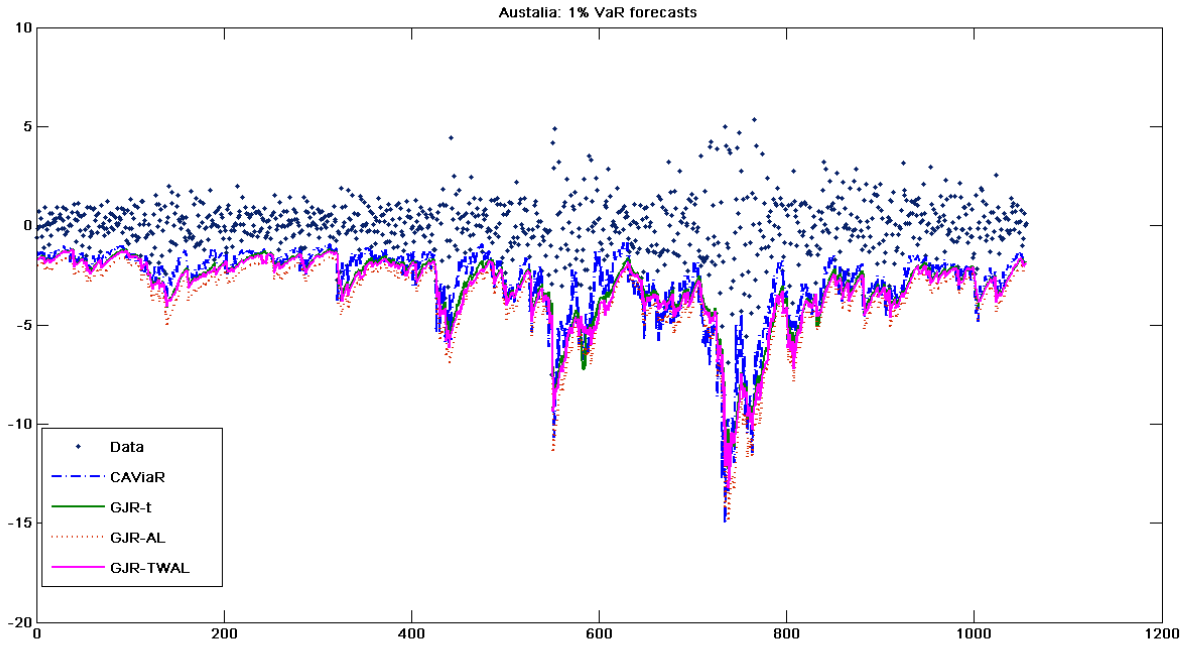


Figure 37: Chapter 4: Ratios of $\hat{\alpha}/\alpha$ at $\alpha = 0.01, 0.05$.

with a lag of 4, and the DQ test with a lag of 4, as in Chapter 2, in Chen et al. (2011) and Chapter 3. At $\alpha = 0.01$ the RM and GJR-n models with Gaussian error are rejected in all series, while the CAV and GJR-t models are rejected, in 3 out of 5 series. The best model is GJR-TW and it is rejected only in one series; the GJR-AL and GJR-PWAL are comparably the second best models. At $\alpha = 0.05$, the GJR-n model is rejected in 4 out of 5 series, while the RM has improved performance in predicting risk levels and is rejected in just two series, which is as good as the GJR-AL model. The CAV model has similar performance as at $\alpha = 0.01$. The GJR-TW and GJR-PWAL models are quite comparable and have no rejection in any series, thus are equally most favoured by the tests.

In summary, the GJR-t model has average VRates closest to the nominal at both $\alpha = 0.01, 0.05$; the GJR-PWAL has average VRates closest to the nominal at $\alpha = 0.01$, and ranks second closest and equally well with GJR-TW at $\alpha = 0.05$. In terms of deviation in VRate ratios from 1, GJR-TW and GJR-PWAL rank best and second best respectively at $\alpha = 0.01$, while GJR-AL and GJR-PWAL rank highest and second best respectively at $\alpha = 0.05$. In terms of the tests, for both $\alpha = 0.01, 0.05$, GJR-TW had the minimum number of rejections, and GJR-PWAL ranks second at $\alpha = 0.01$ and is equally most favoured with GJR-TW at $\alpha = 0.05$. The GJR-n model significantly under-predicts risk levels in most series at $\alpha = 0.01, 0.05$ by over 100% at $\alpha = 0.01$; CAV under-predicts risk

Australian market.



AU/US.

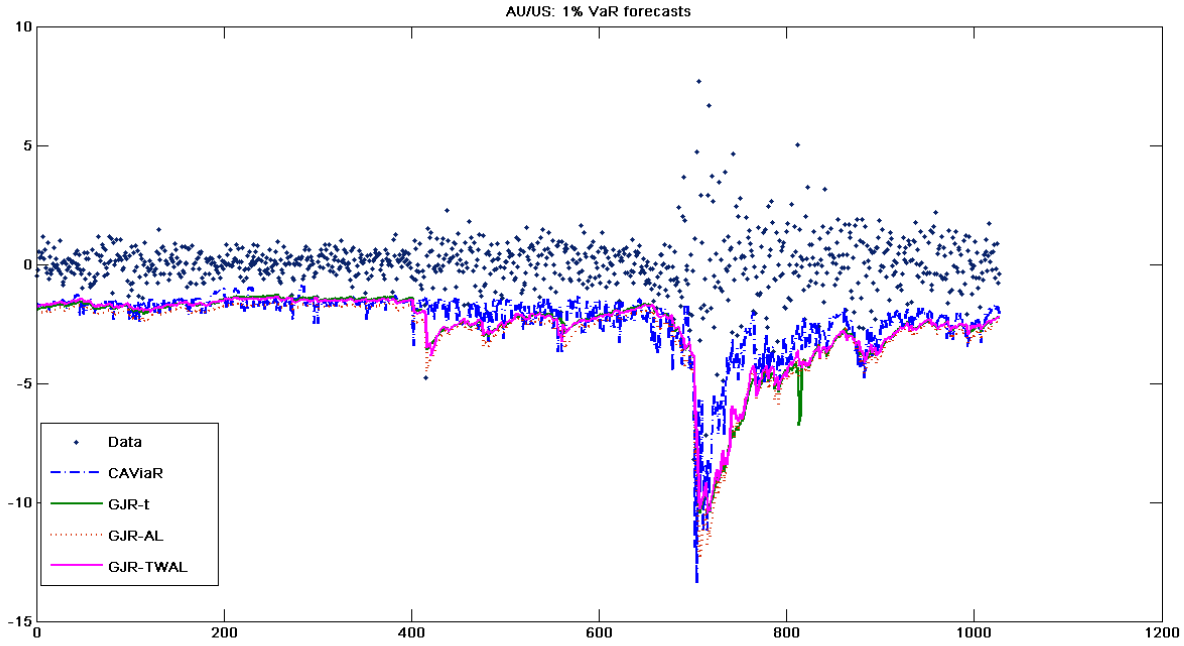


Figure 38: Chapter 4: 1% VaR forecasts from AS CAViaR, GJR-t, GJR-AL, and GJR-PWAL.

Table 44: Chapter 4: Counts of model rejections, at the 5% significance level, at risk level $\alpha = 0.01, 0.05$.

	$\alpha = 0.01$				$\alpha = 0.05$			
Method	UC	CC	DQ	Total (out of 5)	UC	CC	DQ	Total (out of 5)
CAViaR	1	1	3	3	3	1	3	3
RM	5	5	5	**5**	0	0	2	2
GJR-n	5	4	5	**5**	4	1	3	**4**
GJR-t	1	0	3	3	0	1	1	1
GJR-AL	2	0	0	2	0	0	2	2
GJR-TW	0	0	1	1	0	0	0	0
GJR-PWAL	0	0	2	2	0	0	0	0

Note: Boxes indicate the favoured model, bold indicates the least favoured model.

by 30% to 50%, on average, at both levels.

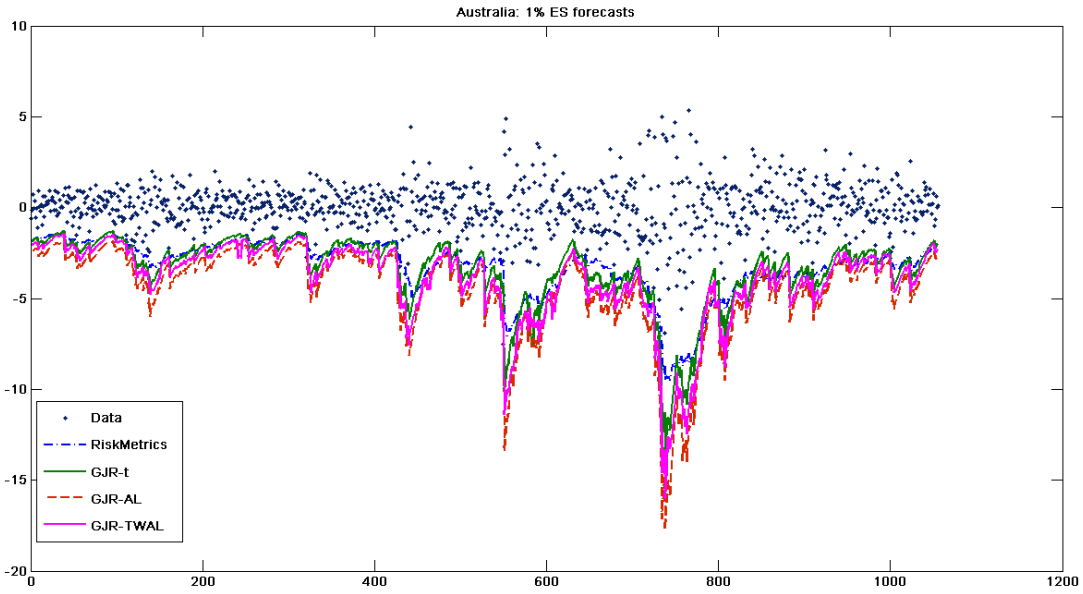
4.4.3 Expected Shortfall forecast comparison

The ES forecasts from several parametric models, for the returns on the Australian stock market and the AU to US dollar exchange rate, are shown in Figure 39.

The plots indicate a clear ordering in ES levels across models: the Gaussian is least extreme, followed by the Student t, STW, and SPWAL, while the AL distribution gives the most extreme ES forecasts. This pattern occurred consistently in each series, and agrees with the results in Chapter 3.

The quantile levels that ES occurs at, for various VaR quantile levels α, are well known and calculable in standard software for the Gaussian and Student t distributions, using their cdf functions; the ES quantile levels, constant for fixed α, for the AL distribution are given in Table 10 in Chapter 2; the ES quantile levels at α for the STW distribution are given in Table 30 in Chapter 3; the close forms for ES and the relation between ES and VaR for SPWAL are derived and given in (43), and its ES quantile levels are calculated by the same formula of the STW ES quantile level function. Table 45 shows the approximate quantile levels for ES from the GJR-t, GJR-TW, and GJR-PWAL models, obtained by using the average of the estimates of each distribution's parameters over the forecast period in each series.

Australian market.



AU/US.

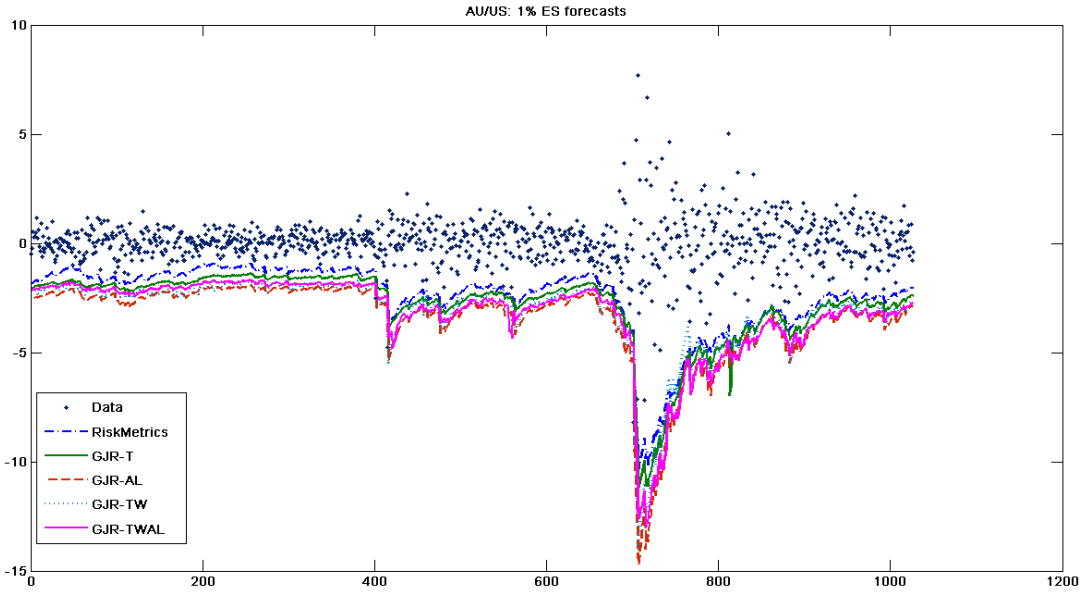


Figure 39: Chapter 4: 1% ES forecasts from RiskMetrics, GJR-t, GJR-AL, and GJR-PWAL.

Table 45: Chapter 4: Estimated nominal ES levels δ_α at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US
GJR-t	0.0036	0.0037	0.0037	0.0035	0.0035
GJR-TW	0.0037	0.0037	0.0037	0.0037	0.0037
GJR-PWAL	0.0037	0.0038	0.0038	0.0037	0.0037
$\alpha = 0.05$	Aust	US	UK	HK	AU/US
GJR-t	0.0185	0.0189	0.0190	0.0180	0.0181
GJR-TW	0.0188	0.0188	0.0189	0.0186	0.0187
GJR-PWAL	0.0189	0.0192	0.0190	0.0189	0.0186

Using these ES quantile levels, the ES violation rate, ESRate, is defined as:

$$\text{ESRate} = \frac{1}{m} \sum_{t=n+1}^{n+m} I(y_t < \text{ES}_t),$$

and a good model should have its ESRate very close to the nominal δ_α.

Table 46 contains the ratios of $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$ across all models and the five series in the forecast period. Again the best risk ratio, closest to 1, is boxed, and ESRates that are significantly different to nominal by the UC test are in bold. These ratios and their average and deviation from desired ration 1 are also presented in Figure 40. At $\alpha = 0.01$, it is clear that the RM and GJR-n models are always anti-conservative and that they significantly under-predict risk levels in all series: on average, ESRates are close to 3 times or more at the nominal 1%. Further, the GJR-t model also under-predicts risk, sometimes significantly; on average, its ES violation rates are double the nominal 1%. Alternatively, the GJR-AL again over-predicts risk levels, sometimes significantly; on average, ESRates are half the nominal 1%, and are thus conservative; agreeing with Chapter 3. The GJR-TW model ranks first with ESRate ratios closest to 1 in three series, followed by GJR-PWAL. The top ranked model in average ESRate is the GJR-PWAL, with 9% higher than the nominal ES quantile level, and ranked second is the GJR-TW, with 12% lower than the nominal. The GJR-TW also has the minimum deviation in ESRate ratios from 1, followed by the GJR-PWAL. It is clear that the performance of GJR-TW and GJR-PWAL are very close, while GJR-TW is marginally more favoured in this informal evaluation method.

At $\alpha = 0.05$, it is similar to the story for VaR at the 5% level. The RM and GJR-n

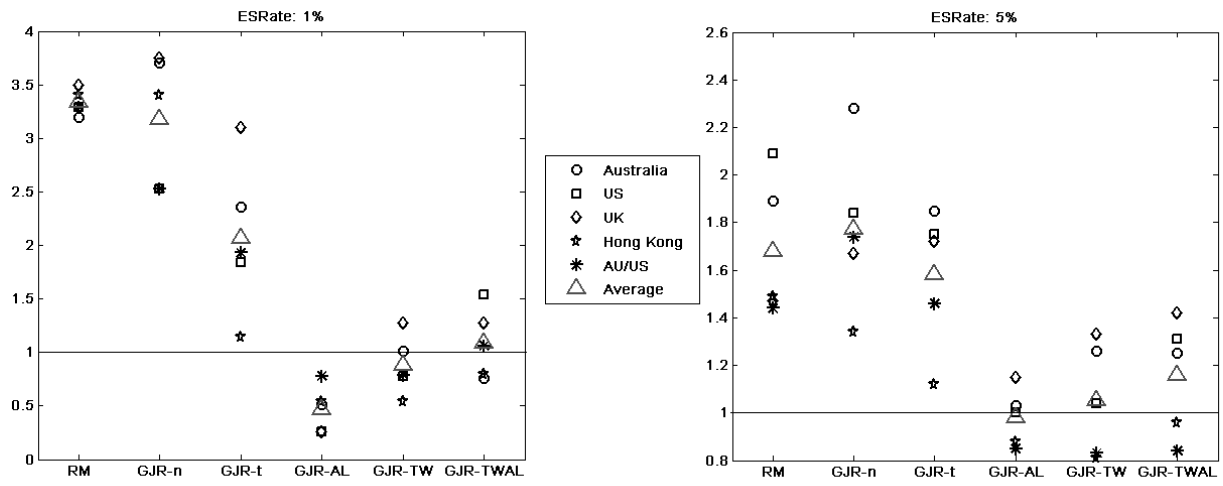


Figure 40: Chapter 4: ESRatios $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$.

are signifIcantly anti-conservative, but now by $\approx 70\%$ on average, and the GJR-t model performs similarly and is mostly rejected in 3 of the 5 series by the UC test. GJR-TW and GJR-PWAL ranks 2nd and 3rd, with average ESRates 5% and 16% higher than the nominal. The GJR-AL model now predicts risk levels most adequately and is favoured in 4 of 5 series, with ESRates, on average, only 2% below nominal and with minimum deviation in ESRate ratio from 1.

Table 47 shows counts of the number of rejections for each ES forecast model, at a 5% significance level, across the five series, under the three formal backtests: the UC, the CC, and the DQ test using the ES quantile levels discussed above. At $\alpha = 0.01$ and 0.05 the RM and GJR-n models are again rejected in all or most series by all tests, while the GJR-t model is rejected in less markets and ranks higher than RM and GJR-n models. The two most favourably performing models, which could not be rejected in any series at $\alpha = 0.01$ are GJR-AL and GJR-TW. GJR-PWAL is rejected by the DQ test in one series. At $\alpha = 0.05$, the GJR-TW is still the best model and could not be rejected in any series by any test. GJR-AL now is rejected once and GJR-PWAL is rejected twice at this level.

Overall, for forecasting ES during this forecast period, GJR-TW and GJR-PWAL have performed more favourably than other models at the 1% level, and GJR-AL is most favoured at the 5% level. The GJR-PWAL has its average ESRate closest to the nominal,

Table 46: Chapter 4: ESRatios $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	Mean	Std.
RM	**3.20**	**3.29**	**3.50**	**3.40**	**3.29**	**3.34**	**2.34**
GJR-n	**3.70**	**2.53**	**3.75**	**3.40**	**2.53**	3.18	2.25
GJR-t	**2.36**	1.84	**3.10**	1.14	1.93	2.07	1.25
GJR-AL	0.51	0.26	0.26	0.54	0.78	0.47	0.57
GJR-TW	1.01	0.78	1.28	0.54	0.79	0.88	0.28
GJR-PWAL	0.76	1.54	1.28	0.80	1.06	1.09	0.30
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	Mean	Std.
RM	**1.89**	**2.09**	**1.47**	**1.49**	1.44	1.68	0.73
GJR-n	**2.28**	**1.84**	**1.67**	1.34	**1.74**	**1.77**	**0.83**
GJR-t	**1.85**	**1.75**	**1.72**	1.12	1.46	1.58	0.64
GJR-AL	1.03	1.00	1.15	0.88	0.85	0.98	0.11
GJR-TW	1.26	1.04	1.33	0.81	0.83	1.05	0.22
GJR-PWAL	1.25	1.31	1.42	0.96	0.84	1.16	0.27

Note: Boxes indicate the ratio is closest to 1 in that market, bold indicates the model is rejected by the UC test (at a 5% level), for each market.

Table 47: Chapter 4: Counts of ES model rejections at $\alpha = 0.01, 0.05$

	$\alpha = 0.01$				$\alpha = 0.05$			
Method	UC	CC	DQ	Total (out of 6)	UC	CC	DQ	Total (out of 6)
RM	5	5	5	**5**	4	3	5	**5**
GJR-n	5	4	5	**5**	4	3	5	**5**
GJR-t	2	1	2	2	3	4	5	**5**
GJR-AL	0	0	0	0	0	0	1	1
GJR-TW	0	0	0	0	0	0	0	0
GJR-PWAL	0	0	1	1	0	1	2	2

Note: Boxes indicate the favoured model, bold indicates the leaset favoured model(at a 5% level).

though it ranks second or third for each series. GJR-TW and GJR-AL ES forecasts are least rejected by the formal tests, across the five return series. The test results seems to favour different models at different levels, and even for different risk measurement methods.

4.4.4 Pre-financial-crisis and post-financial-crisis forecast performance

As in Chapter 2 and Chapter 3, we again consider pre-crisis and post-crisis performance in detail, since model forecasting performance may be different in each period an/or suffer during the crisis.

Tables 48 and 49 present the ratios of VRate/α and ESRate/δ_α at $\alpha = 0.01, 0.05$ for the pre-crisis period for the VaR and ES forecast models, and for the post-crisis period, the ratios are presented in Tables 50 and 51. These ratios are shown in Figure 41 and Figure 42.

The results for the pre-crisis sample are clear in pattern, and mostly similar with those for the whole forecast sample, due to the larger overlapping sample size. GJR-TW model dominates the VaR and ES forecasts most of the time. It predicts most accurately at both risk levels for VaR and at the 5% level for ES, with VRate and ESRate averaging closest to 1, or with a minimum deviation in ratios from 1. The GJR-PWAL forecasts most accurately at 1% ES with average ESRate closest to 1. The GJR-TW is marginally more favoured than the other models, though the GJR-PWAL generally ranks second best in most series, as well as in the average VRate and in deviation in ratios from 1. Furthermore, the GJR-AL model is again the only always conservative risk forecaster for both VaR and ES.

Results for the post-crisis period tell a slightly different story. For VaR forecasting, GJR-t, GJR-TW, and GJR-PWAL perform comparably well at $\alpha = 0.01$. Particularly the CAV, GJR-t, and GJR-PWAL have average ratios closest to 1, across the five series. The GJR-PWAL model also has the lowest deviations about 1. The RM has much better performance than in the pre-crisis period. For the 5% level VaR forecast, the GJR-t has an average VRate closest to 1. At this level, GJR-PWAL performs marginally better than GJR-TW. For ES forecasting, the results are more consistent with those in the whole forecast sample as well as the pre-crisis period. The GJR-TW is clearly the best model post-crisis at 1%, with an average ratio closest to 1 and smallest deviation about 1. At the 5% risk level, however, the GJR-PWAL model has the average ratio closest to 1, followed by GJR-AL, which has the smallest deviation in ratios from 1.

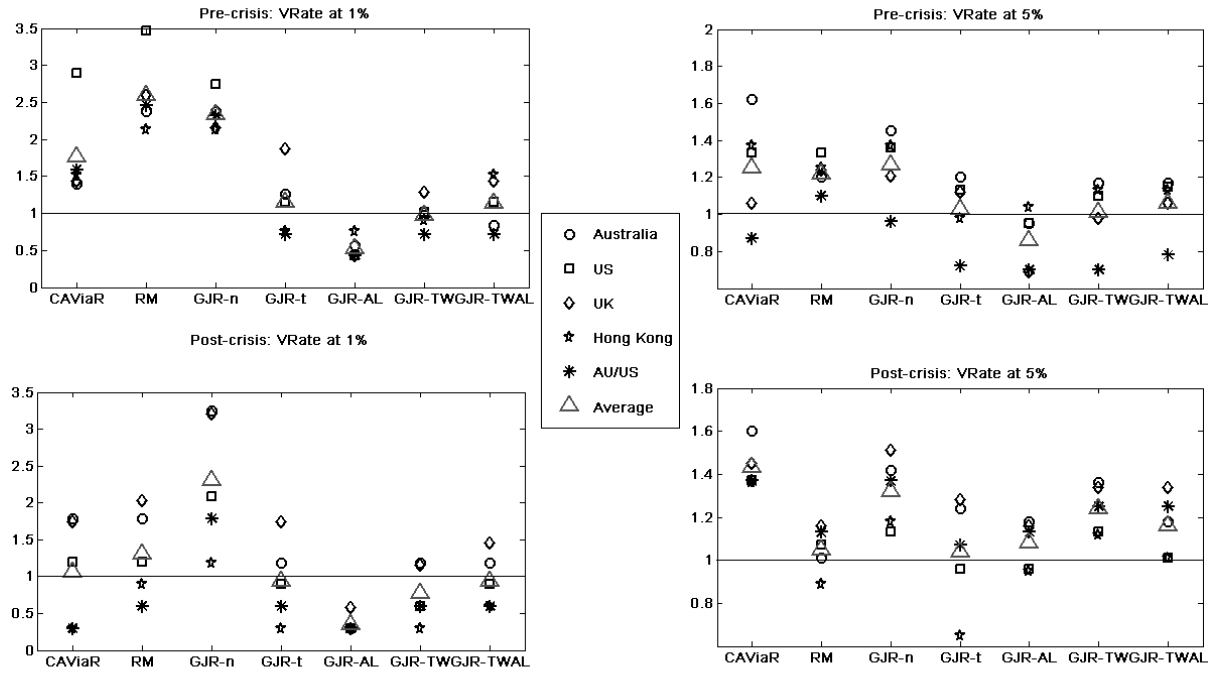


Figure 41: Chapter 4: Pre-crisis and post-crisis VRate of $\hat{\alpha}/\alpha$ at $\alpha = 0.01, 0.05$.

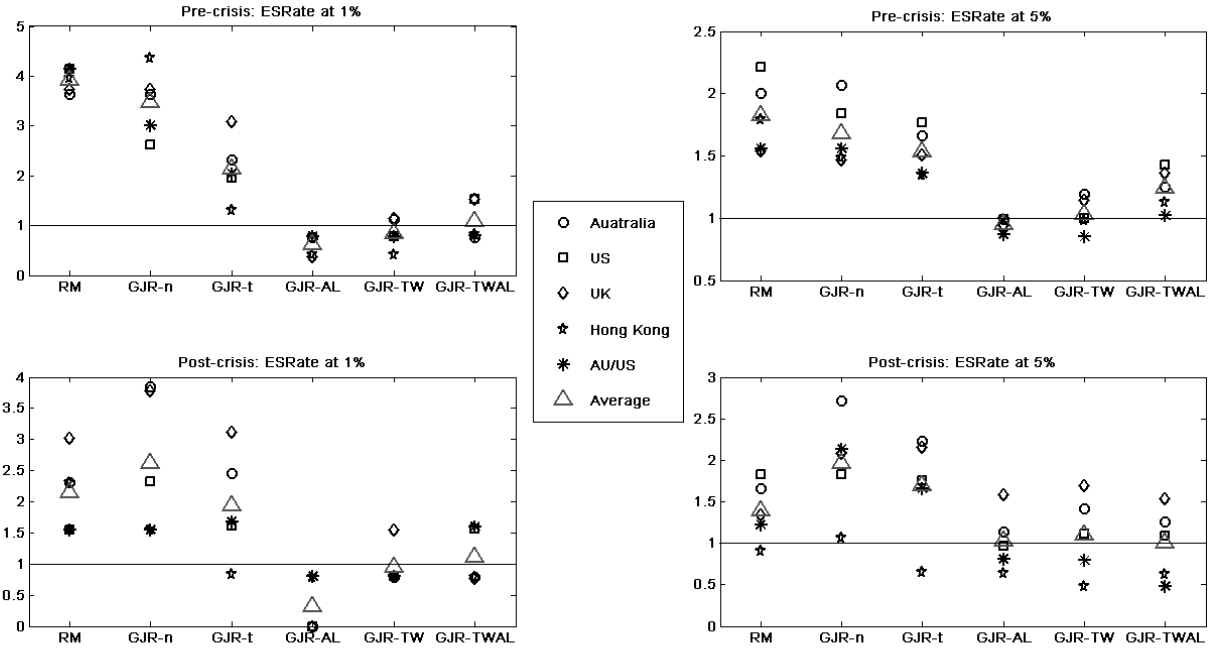


Figure 42: Chapter4: Pre-crisis and post-crisis: ESRates of $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$.

Table 48: Chapter 4: Before the financial crisis: ratios of VRate/α at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	Mean	Std.
CAViaR	1.39	**2.89**	1.44	1.52	1.59	1.77	0.95
RM	**2.37**	**3.46**	**2.59**	**2.13**	**2.46**	**2.60**	**1.67**
GJR-n	**2.37**	**2.74**	**2.16**	**2.13**	**2.32**	2.34	1.36
GJR-t	1.26	1.15	**1.87**	0.76	0.72	1.15	0.44
GJR-AL	0.56	0.43	**0.43**	0.76	0.43	0.52	0.49
GJR-TW	0.98	1.01	1.29	0.91	0.72	0.98	0.18
GJR-PWAL	0.84	1.15	1.44	1.52	0.72	1.14	0.34
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	Mean	Std.
CAViaR	**1.62**	1.33	1.06	**1.37**	0.87	1.25	0.36
RM	1.20	1.33	1.24	1.25	1.10	1.22	0.23
GJR-n	**1.45**	**1.36**	1.21	**1.37**	0.96	1.27	0.32
GJR-t	1.20	1.13	1.12	0.98	0.72	1.03	0.17
GJR-AL	0.95	0.95	0.69	1.04	0.70	0.86	0.20
GJR-TW	1.17	1.10	0.98	1.13	0.70	1.01	0.17
GJR-PWAL	1.17	1.15	1.06	1.13	0.78	1.06	0.16

Note: Boxes indicate that the ratio is closest to 1 in that market, bold indicates that the model is rejected by the UC test (at a 5% level), for each market.

Table 49: Chapter 4: Before the financial crisis: $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	Mean	Std.
RM	**3.63**	**4.13**	**3.73**	**3.96**	**4.14**	3.92	**2.93**
GJR-n	**3.63**	**2.63**	**3.73**	**4.36**	**3.01**	**3.47**	2.54
GJR-t	2.31	1.95	**3.08**	1.30	2.04	2.14	1.28
GJR-AL	0.75	0.75	0.38	0.41	0.78	0.61	0.43
GJR-TW	1.12	0.77	1.15	0.41	0.78	0.85	0.31
GJR-PWAL	0.75	1.52	1.53	0.81	0.79	1.08	0.37
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	Mean	Std.
RM	**2.00**	**2.21**	1.54	**1.79**	1.56	**1.82**	**0.86**
GJR-n	**2.07**	**1.84**	1.47	1.48	1.56	1.68	0.72
GJR-t	**1.66**	**1.77**	1.51	1.35	1.36	1.53	0.55
GJR-AL	0.99	0.99	0.94	0.99	0.87	0.95	0.07
GJR-TW	1.19	1.00	1.14	0.98	0.85	1.03	0.12
GJR-PWAL	1.25	1.43	1.36	1.13	1.02	1.24	0.28

Note: Boxes indicate that the ratio is closest to 1 in that market, bold indicates that the model is rejected by the unconditional coverage test (at a 5% level), for each market.

Table 50: Chapter 4: Post-financial crisis: ratios of $\hat{\alpha}/\alpha$ at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	Mean	Std.
CAViaR	1.78	1.19	1.74	0.30	0.30	1.06	0.66
RM	1.78	1.19	2.03	0.89	0.60	1.30	0.61
GJR-n	**3.25**	2.09	**3.20**	1.18	1.79	**2.30**	**1.53**
GJR-t	1.18	0.90	1.74	0.30	0.60	0.94	0.50
GJR-AL	0.30	0.30	0.58	0.30	0.30	0.35	0.66
GJR-TW	1.18	0.60	1.16	0.30	0.60	0.77	0.42
GJR-PWAL	1.18	0.90	1.45	0.59	0.60	0.94	0.34
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	Mean	Std.
CAViaR	**1.60**	1.37	1.45	1.36	1.37	1.43	**0.44**
RM	1.01	1.07	1.16	0.89	1.13	1.05	0.11
GJR-n	1.42	1.13	1.51	1.18	1.37	**1.32**	0.35
GJR-t	1.24	0.96	1.28	0.65	1.07	1.04	0.23
GJR-AL	1.18	0.96	1.16	0.95	1.13	1.08	0.13
GJR-TW	1.36	1.13	1.34	1.12	1.25	1.24	0.26
GJR-PWAL	1.18	1.01	1.34	1.01	1.25	1.16	0.20

Note: Boxes indicate that the ratio is closest to 1 in that market, bold indicates that the model is rejected by the unconditional coverage test (at a 5% level), for each market.

Table 51: Chapter 4: Post-financial crisis: ratios of $\hat{\delta}_\alpha/\delta_\alpha$ at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	Mean	Std.
RM	2.31	1.55	3.02	2.31	1.55	2.15	1.27
GJR-n	**3.85**	2.33	**3.78**	1.54	1.55	**2.61**	**1.91**
GJR-t	2.45	1.61	3.12	0.84	1.68	1.94	1.22
GJR-AL	0.00	0.00	0.00	0.79	0.80	0.32	0.79
GJR-TW	0.79	0.80	1.55	0.80	0.80	0.95	0.31
GJR-PWAL	0.79	1.57	0.77	0.79	1.60	1.11	0.41
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	Mean	Std.
RM	1.66	1.83	1.34	0.91	1.22	1.39	0.51
GJR-n	**2.72**	1.83	2.08	1.06	**2.13**	**1.96**	**1.10**
GJR-t	**2.23**	1.75	**2.15**	0.65	1.66	1.69	0.89
GJR-AL	1.13	0.97	1.58	0.64	0.81	1.03	0.32
GJR-TW	1.41	1.11	1.70	0.48	0.79	1.10	0.44
GJR-PWAL	1.26	1.09	1.54	0.62	0.48	1.00	0.39

Note: Boxes indicate that the ratio is closest to 1 in that market, bold indicates that the model is rejected by the unconditional coverage test (at a 5% level), for each market.

4.4.5 Loss function

Loss functions are again applied to assess quantile forecasts as in Chapter 2 and Chapter 3. Table 52 and Table 53 present the losses for the VaR and ES forecasts via various models across the five series in the overall forecast period. For each series, the lowest loss is boxed, while the lowest mean and standard deviation (Std.) across all the models are also boxed. Highest loss as well as mean and standard deviation are in bold. The results are also demonstrated in Figure 43 for a clearer view.

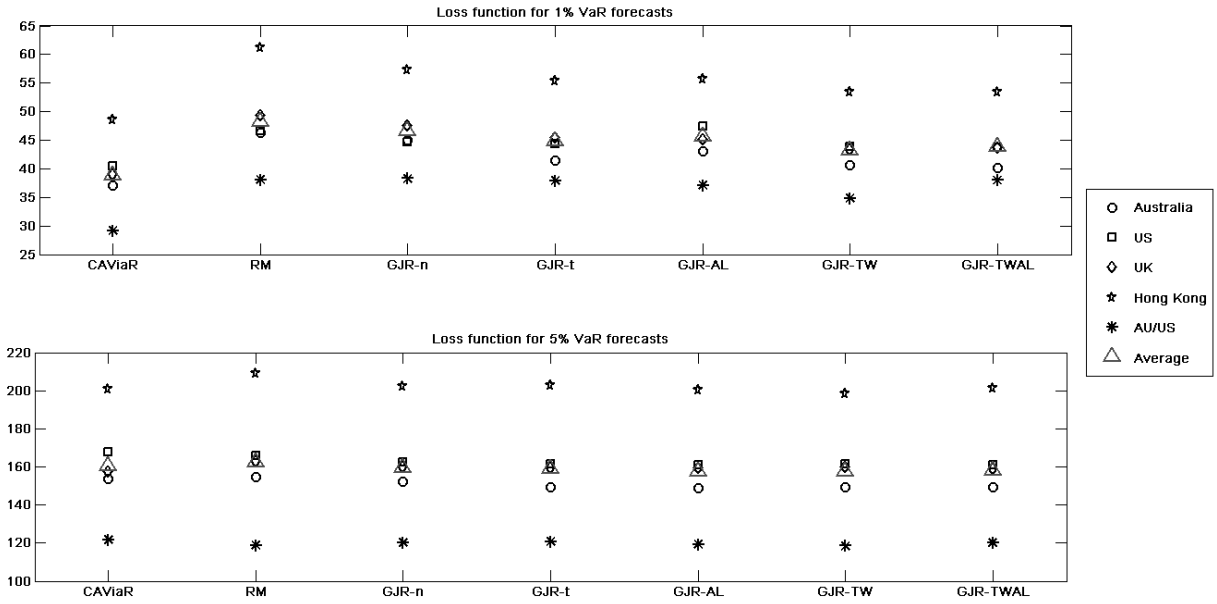
It is clear that the CAV has the smallest losses and lowest average loss for VaR at the 1% risk level. The GJR-TW has the lowest losses for the 5% VaR forecast in three series as well as the minimum average loss. For the ES forecast, the GJR-TW and GJR-PWAL have comparably lowest losses at the 1% level. The GJR-AL has the smallest losses for the 5% level ES forecast in three series, and has the minimum average loss. The RM has the largest loss in all or most of the series for both VaR and ES forecasts at both risk levels.

Table 52: Chapter 4: Loss function for VaR at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	Mean	Std.
CARViaR	37.1	40.48	39.0	48.5	29.2	38.9	6.9
RM	**46.3**	46.6	**49.4**	**61.2**	38.1	**48.3**	**8.4**
GJR-N	45.0	44.6	47.6	57.3	**38.3**	46.6	6.9
GJR-T	41.5	44.3	45.5	55.4	37.9	40.4	6.5
GJR-AL	43.1	**47.4**	45.1	55.7	37.1	45.7	6.8
GJR-TW	40.7	43.9	43.3	53.4	34.8	43.2	6.7
GJR-PWAL	40.1	43.7	43.7	53.5	38.0	43.8	6.0
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	Mean	Std.
CARViaR	153.8	**167.8**	157.7	200.9	**121.8**	160.4	28.5
RM	**154.8**	165.9	**163.0**	**209.0**	118.8	**162.3**	**32.2**
GJR-N	152.4	162.4	160.0	202.4	120.3	159.5	29.3
GJR-T	149.5	161.5	159.7	202.6	120.7	158.8	29.4
GJR-AL	148.9	161.0	159.7	200.3	119.3	157.8	29.1
GJR-TW	149.5	161.3	160.2	198.3	118.6	157.6	28.5
GJR-PWAL	149.1	-160.9	159.1	201.2	120.1	158.1	29.1

Note: Boxes indicate that the loss function is the lowest in that market, bold indicates that the loss function is highest in that market, for each market.

VaR at $\alpha = 0.01, 0.05$.



ES at $\alpha = 0.01, 0.05$.

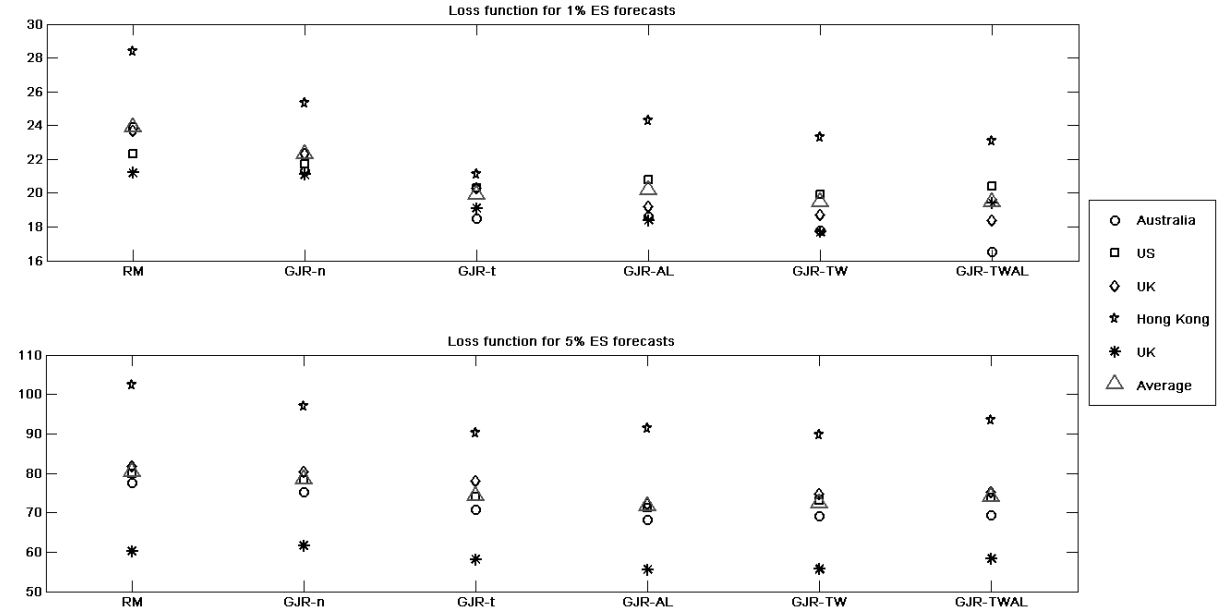


Figure 43: Chapter 4: Loss function for VaR and ES forecasts.

Table 53: Chapter 4: Loss function for ES at $\alpha = 0.01, 0.05$.

$\alpha = 0.01$	Aust	US	UK	HK	AU/US	Mean	Std.
RM	**23.9**	**22.3**	**23.7**	**28.4**	**21.2**	**23.9**	**2.7**
GJR-N	21.3	21.7	22.3	25.3	21.1	22.3	1.7
GJR-T	18.5	20.3	20.3	21.1	19.1	19.8	1.1
GJR-AL	18.6	20.8	19.2	24.3	18.4	20.2	2.4
GJR-TW	17.8	19.9	18.7	23.3	17.7	19.5	2.3
GJR-PWAL	16.5	20.4	18.4	23.1	19.4	19.5	2.5
$\alpha = 0.05$	Aust	US	UK	HK	AU/US	Mean	Std.
RM	**77.5**	**80.0**	**81.7**	**102.3**	60.3	**80.3**	**14.9**
GJR-N	75.1	78.2	80.3	97.1	**61.7**	78.5	12.7
GJR-T	70.7	74.0	77.9	90.2	58.1	74.2	11.7
GJR-AL	68.1	71.3	72.1	91.4	55.6	71.7	12.9
GJR-TW	69.1	73.0	74.8	89.7	55.7	72.5	12.2
GJR-PWAL	69.4	73.9	75.1	93.4	58.3	74.0	12.7

Note: Boxes indicate that the loss function is the lowest in that market, bold indicates that the loss function is highest in that market, for each market.

4.4.6 Further evaluations of model performance

This section discusses a range of formal methods and hypothesis tests used for evaluating the goodness-of-fit of models.

4.4.6.1 Log predictive likelihood

The predictive likelihood is a appropriate measure of fit for out-of-sample forecasts instead of in-sample model fitting. Here, we followed Geweke and Keane (2007) and Li, Villani and Kohn (2009) to compare financial time series models, using predictive likelihood.

Given the observed return series $y_1, \ldots, y_t$, at jth MCMC iterate, the estimates of conditional distribution parameters θ_1 and volatility parameters θ_2 are $\theta_1^{[j]}$ and $\theta_2^{[j]}$ respectively. Then the log predictive likelihood for y_{t+1} can be obtained, using the density function $f_D(\cdot)$ of the i.i.d. conditional distribution D and the $\theta_1^{[j]}, \theta_2^{[j]}$:

$$lpl_{t+1} = \log\left(f_D\left(y_{t+1}|\theta_1^{[j]}, \theta_2^{[j]}\right)\right).$$

Table 54 shows the sum of log predictive likelihood for the whole forecasting period of various models, across five series. The box indicates the highest log predictive likelihoods

in each series, and the lowest are in bold. The results are demonstrated in Figure 44; the average log predictive likelihood for each model is in a big grey triangle. There is an apparent pattern in the results: the GJR-t has the largest log predictive likelihood in each series, as well as the largest average log predictive likelihood, closely followed by those of the GJR-PWAL model; each of the GJR-RM and GJR-TW has the smallest log predictive likelihoods in two series, while the GJR-AL has the smallest average log predictive likelihood. Thus GJR-t fits the data best, and GJR-PWAL ranks second under this criterion.

Table 54: Chapter 4: Log predictive likelihood

	Aust	US	UK	HK	AU/US	Mean
RM	-1652.3	**-1620.9**	-1662.4	**-1931.8**	-1279.8	-1629.4
GJR-n	-1634.0	-1585.3	-1630.9	-1913.6	-1283.5	-1609.5
GJR-t	-1511.5	-1489.6	-1570.5	-1762.6	-1117.4	-1490.3
GJR-AL	**-1692.1**	-1589.7	-1672.4	-1928.0	-1302.2	**-1636.9**
GJR-TW	-1688.9	-1576.7	**-1678.5**	-1916.8	**-1294.4**	-1631.1
GJR-PWAL	-1630.2	-1577.4	-1624.9	-1907.4	-1258.2	-1599.6

Note: Boxes indicate that the log predictive likelihood is the highest in that market, bold indicates that the log predictive lieklihood is lowest in that market, for each market.

4.4.6.2 Diagnostic tests

A range of diagnostic hypothesis tests on the normalized predictive residuals are also considered: the Shapiro-Wilk test, the Lilliefors test, the Chi-square goodness-of-fit test, the Jacque-Bera test, the Ljung-Box on residuals and squared residuals, and the ARCH test of Engle (1982). The first three are used to test the null hypothesis that a sample $x_1, \ldots, x_n$ comes from a normally distributed population; the 4th and 5th measure the goodness-of-fit, while the Jacque-Bera also tests the normality of sample; the Ljung-Box tests assess the null hypothesis that there is no autocorrelations in the sample, while the ARCH test of Engle (1982) tests the null hypothesis that there exists no heteroscedasticity (ARCH effects) in the sample.

Here we follow the idea of Gerlach, Carter and Kohn (1999) to compute the test statistics by using MCMC and importance sampling. That is, for each MC iterate, with

newly updated jth sample of parameters $\theta_1^{(j)}$ and $\theta_2^{(j)}$, we generate (one-step-ahead) forecast of conditional variance. Using this conditional variance forecast, the estimate of conditional distribution parameters, as well as the actual returns in the forecast horizon (here one-day-ahead), we calculate the normalized residual. Repeat this process for K iterations, a vector of MC samples of normalized residuals can be obtained. The average of the MC samples is the estimate of normalized predictive residuals for tth forecast period, this process is repeated for the whole forecast period to obtained a vector of predictive residuals across the forecast period. Diagnostic tests then are applied to this residuals to evaluate the forecasting performance of the models. The above diagnostic methods are widely applied to evaluate financial time series models, for example, Diebold, Gunther and Tay (1998) and Diebold, Hahn and Tay (1999).

Table 55 to Table 61 present the test results of the above hypothesis tests. 1 indicates that the null is rejected, while 0 indicates the null cannot be rejected, at a 5% significance level; failure of rejection is indicated in a box. The p-values are in brackets, with the highest p-value in each series indicated by a "$*$". The counts of rejection for each model are recorded as well, with the least favoured models in bold, and most favoured models in boxes.

It is clear that the GJR-PWAL model is most favoured by the Shapiro-Wilk test, the Lilliefors test, the Chi-square goodness-of-fit test, and the Jacque-Bera test. Most of the time, the tests failed to reject the normality of the normalized predictive residuals from the GJR-PWAL model. Even in the case of rejection, the p-values for the residuals from the GJR-PWAL model are usually the highest in that series, compared to the p-values of other rejected models. Under these criteria, the GJR-PWAL is the best model fitting the data. However, the results of the Ljung-Box test on the residuals and squared residuals indicate that the GJR-PWAL model is among the most frequently rejected models. The GJR-AL, GJR-TW, and GJR-PWAL are equally least favoured by the Ljung-Box test on residuals; while the RM is the best under this test. For the squared residuals, the RM, GJR-AL, GJR-TW, and GJR-PWAL are least favoured. This implies there exists a group of autocorrelations in the residuals from the GJR-PWAL model. GJR-PWAL model is rejected twice by the ARCH test, which is less than the RM and GJR-TW, but more than

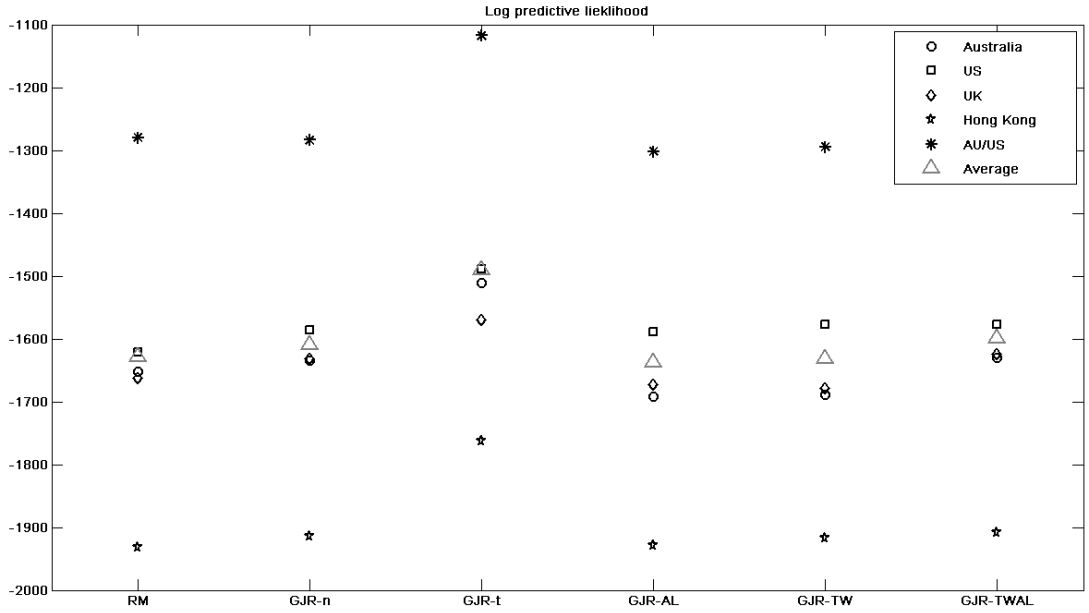


Figure 44: Chapter 4: Log predictive likelihood for six models across five markets.

the GJR-n and GJR-t, which are rejected only in one series. This implies there may exist heteroscedasticity in the residuals from GJR-PWAL. Overall, the GJR-PWAL model has least rejections (14) by all seven hypothesis tests, and thus is the best fitting model.

Figures 45 to 47 show the Quantile-Quantile (Q-Q) plots of residuals from six models fitting the return series of the Australian market. If the normalized residuals are from a normal population, the Q-Q plot should lie on the line $y = x$. First, all the models miss the normality in the right tails, except the RM. Since the attention of this thesis is on the left tails of the distribution, we focus on the interpretation of the left side of the Q-Q plots. Second, both tails of residuals from the GJR-AL model are far away from the normal line, and the residuals are significantly left skewed, which agrees with the fact that AL distribution has much fatter tails. Alternatively, residuals from the GJR-t model have their left tail closest to that of the normal distribution; that is, the left tail totally lies on the line $y = x$. The residuals from the GJR-PWAL model fits the series well in the sense that the residuals are slightly left skewed, and the left tail is less fat than that of the GJR-TW residuals. On the other hand, the residuals of the RM and GJR-n are

Table 55: Chapter 4: Shapiro-Wilk test.

	Aust	US	UK	HK	AU/US	Total
RM	1	1	1	1	1	**5**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
GJR-n	1	1	1	1	1	**5**
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	
GJR-t	1	1	1	1	1	**5**
	(0.000)	(0.000)	(0.000)	(0.004)	(0.000)	
GJR-AL	1	1	1	1	1	**5**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
GJR-TW	1	1	0	1	1	4
	(0.001)	(0.000)	(0.056)	(0.000)	(0.003)	
GJR-PWAL	1	1	0	0	1	3
	(0.010*)	(0.000)	(0.783*)	(0.524*)	(0.017*)	

Note: Boxes indicate that the null cannot be rejected, $*$ indicates the model with the highest p value in each market, bold indicates the leaset favoured model (at a 5% level).

Table 56: Chapter 4: Lilliefors test.

	Aust	US	UK	HK	AU/US	Total
RM	1	1	1	1	1	**5**
	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	
GJR-N	1	1	1	1	1	**5**
	(0.002)	(0.001)	(0.010)	(0.001)	(0.001)	
GJR-T	1	1	1	1	1	**5**
	(0.007)	(0.001)	(0.036)	(0.001)	(0.001)	
GJR-AL	1	1	1	1	1	**5**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
GJR-TW	1	1	0	1	0	3
	(0.010*)	(0.001)	(0.089)	(0.038)	(0.104)	
GJR-PWAL	1	0	0	0	0	1
	(0.006)	(0.147*)	(0.500*)	(0.402*)	(0.292*)	

Note: Boxes indicate that the null cannot be rejected and the favoured model, bold indicates the leaset favoured model (at a 5% level).

significantly right skewed, disagreeing with the negative skewness in the return series, which explains their consistent and significant under-prediction of risk levels. The fitting of other series exhibit similar patterns.

Overall, the results agree with those in Chapter 3. The GJR-TW model is the most

Table 57: Chapter 4: Chi-square goodness-of-fit test.

	Aust	US	UK	HK	AU/US	Total
RM	1	1	1	1	1	**5**
	(0.000)	(0.000)	(0.004)	(0.000)	(0.000)	
GJR-N	1	1	1	1	1	**5**
	(0.000)	(0.000)	(0.000)	(0.008)	(0.000)	
GJR-T	1	1	1	1	1	**5**
	(0.002)	(0.000)	(0.003)	(0.001)	(0.001)	
GJR-AL	1	1	1	1	1	**5**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
GJR-TW	1	1	0	1	1	4
	(0.007)	(0.001)	(0.068)	(0.000)	(0.051)	
GJR-PWAL	0	0	0	0	0	0
	(0.166*)	(0.185*)	(0.345*)	(0.563*)	(0.151*)	

Note: Boxes indicate that the null cannot be rejected and the favoured model, bold indicates the leaset favoured model (at a 5% level).

Table 58: Chapter 4: Jacque-Bera test.

	Aust	US	UK	HK	AU/US	Total
RM	1	1	1	1	1	**5**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
GJR-N	1	1	1	1	1	**5**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
GJR-T	1	1	1	1	1	**5**
	(0.001)	(0.001)	(0.001)	(0.033)	(0.001)	
GJR-AL	1	1	1	1	1	**5**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
GJR-TW	1	1	0	1	1	4
	(0.011)	(0.001)	(0.098)	(0.003)	(0.019)	
GJR-PWAL	1	1	0	0	0	2
	(0.031*)	(0.001)	(0.500*)	(0.430*)	(0.140*)	

Note: Boxes indicate that the null cannot be rejected and the favoured model, bold indicates the leaset favoured model (at a 5% level).

favourably performing risk forecaster for this forecast data period across the five series over both VaR and ES forecasting at both $\alpha = 0.01, 0.05$ levels. By almost all criteria, GJR-TW model ranks best or equal best, with violation rates closest to 1 by average and squared deviation, minimum number of model rejections by formal tests, both in

Table 59: Chapter 4: Ljung-Box test on residuals: a lag of 20.

	Aust	US	UK	HK	AU/US	Total
RM	1	[0]	[0]	[0]	[0]	[1]
	(0.017)	(0.074)	(0.275*)	(0.247*)	(0.509*)	
GJR-N	1	[0]	[0]	1	[0]	2
	(0.010)	(0.053)	(0.260)	(0.047)	(0.306)	
GJR-T	1	[0]	[0]	1	[0]	2
	(0.005)	(0.222*)	(0.176)	(0.027)	(0.341)	
GJR-AL	1	1	[0]	1	[0]	**3**
	(0.014*)	(0.038)	(0.157)	(0.022)	(0.357)	
GJR-TW	1	1	[0]	1	[0]	**3**
	(0.003)	(0.017)	(0.186)	(0.020)	(0.384)	
GJR-PWAL	1	1	[0]	1	[0]	**3**
	(0.001)	(0.0110)	(0.067)	(0.010)	(0.240)	

Note: Boxes indicate that the null cannot t be rejected and the favoured model, bold indicates the leaset favoured model (at a 5% level).

Table 60: Chapter 4: Ljung-Box test on squared residuals: 20 lags.

	Aust	US	UK	HK	AU/US	Total
RM	[0]	[0]	1	1	1	**3**
	(0.119)	(0.961*)	(0.000)	(0.022*)	(0.000)	
GJR-N	[0]	[0]	[0]	1	1	[2]
	(0.909*)	(0.803)	(0.071)	(0.020)	(0.000)	
GJR-T	[0]	[0]	[0]	1	1	[2]
	(0.695)	(0.222)	(0.061)	(0.000)	(0.000)	
GJR-AL	[0]	1	[0]	1	1	**3**
	(0.762)	(0.016)	(0.384*)	(0.001)	(0.000)	
GJR-TW	[0]	[0]	1	1	1	**3**
	(0.236)	(0.055)	(0.001)	(0.002)	(0.007*)	
GJR-PWAL	[0]	[0]	1	1	1	**3**
	(0.323)	(0.145)	(0.014)	(0.000)	(0.000)	

Note: Boxes indicate that the null can not be rejected and the favoured model, bold indicates the leaset favoured model (at a 5% level).

the entire period and in the pre- and post-GFC periods. It consistently performs well in VaR forecasting for $\alpha = 0.01$, while GJR-AL model is always conservative and exhibited violation rates usually below nominal, with comparatively small variation in violation rate ratios. CAV has the minimum losses for the 1% VaR forecast, while for the rest of the

Table 61: Chapter 4: ARCH test: a lag of 10.

	Aust	US	UK	HK	AU/US	Total
RM	1	$\boxed{0}$	1	1	$\boxed{0}$	**3**
	(0.026)	(0.486*)	(0.004)	(0.005*)	(0.745*)	
GJR-N	$\boxed{0}$	$\boxed{0}$	$\boxed{0}$	1	$\boxed{0}$	$\boxed{1}$
	(0.933*)	(0.382)	(0.468*)	(0.005*)	(0.388)	
GJR-T	$\boxed{0}$	$\boxed{0}$	$\boxed{0}$	1	$\boxed{0}$	$\boxed{1}$
	(0.809)	(0.137)	(0.300)	(0.000)	(0.110)	
GJR-AL	$\boxed{0}$	1	$\boxed{0}$	1	$\boxed{0}$	2
	(0.805)	(0.003)	(0.452)	(0.000)	(0.094)	
GJR-TW	$\boxed{0}$	1	1	1	$\boxed{0}$	**3**
	(0.249)	(0.002)	(0.002)	(0.002)	(0.643)	
GJR-PWAL	$\boxed{0}$	1	$\boxed{0}$	1	$\boxed{0}$	2
	(0.695)	(0.038)	(0.209)	(0.001)	(0.127)	

Note: Boxes indicate that the null cannot t be rejected and the favoured model, bold indicates the leaset favoured model (at a 5% level).

risk level prediction, the GJR-TW, GJR-PWAL, and GJR-AL are the best or equally best under this criterion. The GJR-PWAL model predicts marginally less accurate risk levels, but is very close to the GJR-TW and GJR-t. The GJR-PWAL model generally fits the five series in the forecasting period better than the other models. It maximizes the log predictive likelihood, which is an important Bayes factor for model selection. It also has the least rejections by a range of formal goodness-of-fit measures among all the models.

4.5 Conclusion

The two-sided Weibull distribution proposed in Chapter 3, outperformed a range of popular competing models in risk forecasting, but may not be chosen during a formal model selection process due to its inapt fitting to the data, due to the bimodal problem. For this purpose, we derived a partitioned distribution, combining Weibull tails and a uni-modal AL centre for the conditional return distribution, paired with the classic GJR-GARCH volatility model, to forecast tail risk levels as well as improve the model fitting to data. Properties of this distribution, including VaR and ES functions, were developed and presented.

An adaptive Markov chain Monte Carlo method, employing the Metropolis-Hastings

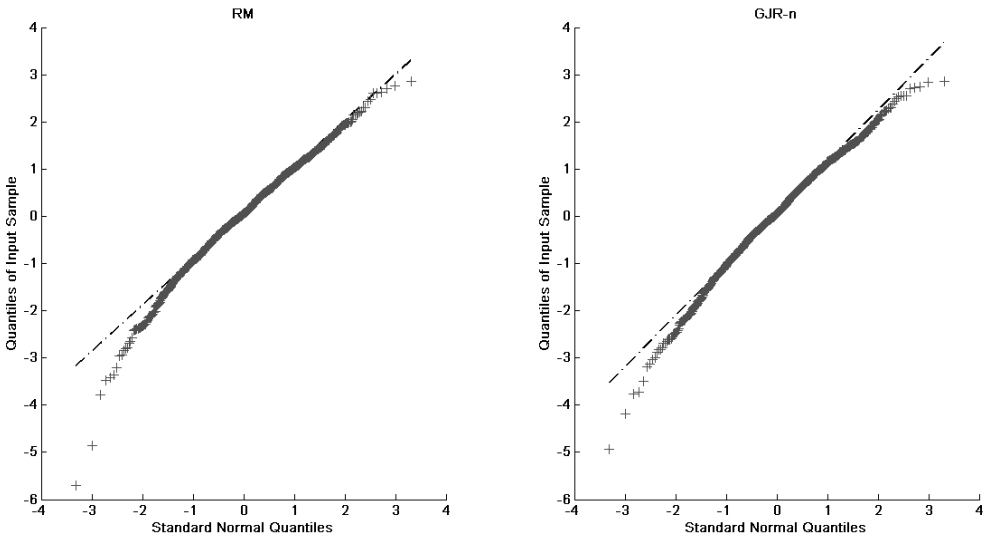


Figure 45: Chapter 4: Q-Q plots of residuals from the RiskMetrics and the GJR-GARCH-n model fitting of the return series of the Australian market.

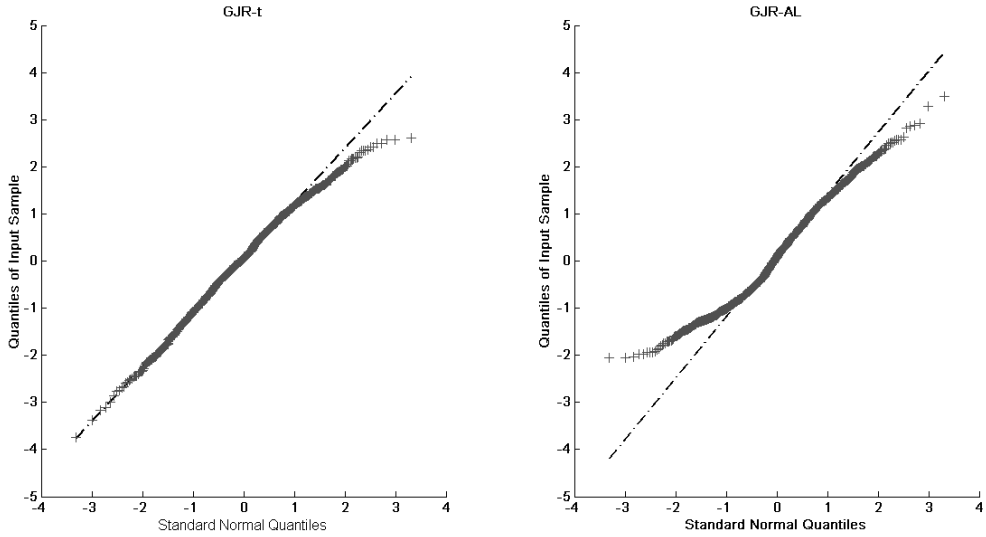


Figure 46: Chapter 4: Q-Q plots of residuals from the GJR-GARCH-t and the GJR-GARCH-AL model fitting of the return series of the Australian market.

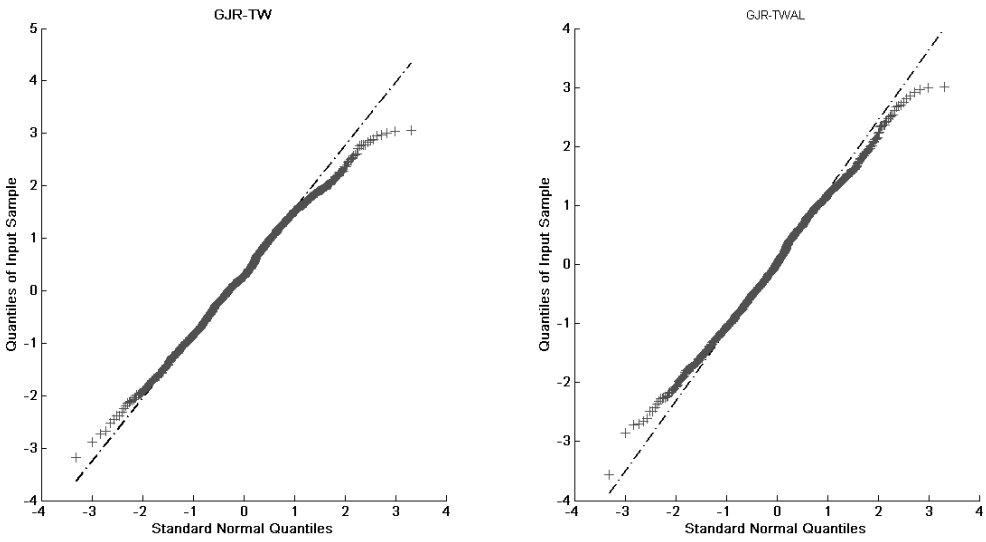


Figure 47: Chapter 4: Q-Q plots of residuals from the GJR-GARCH-TW and the GJR-GARCH-PWAL model fitting of the return series of the Australian market.

algorithm with a flexible mixture of Student t proposal densities, was used for estimation and forecasting. A simulation study indicated that the mixture of Student t proposals improves the estimation accuracy and speeds up the convergence in the Markov chain, and is more general than the mixture of Gaussian proposal adopted in Chapters 2 and 3.

An empirical study of five asset-return series found that the GJR-GARCH model with a conditional standardized partitioned distribution of Weibull and asymmetric Laplace (SPWAL) was very close to, though less accurate at forecasting VaR and ES levels than GJR-GARCH with a conditional two-sided Weibull errors. GJR-GARCH-PWAL could not be frequently rejected by several formal backtesting methods, compared to the Risk-Metrics, GJR-GARCH with Gaussian, Student t, and asymmetric Laplace conditional return distributions, as well as an asymmetric slope CAViaR model. The GJR-GARCH-PWAL model performed better during and after the global financial crisis compared to the pre-crisis period. This may be mostly affected by a larger number of parameters of SPWAL than the TW distribution, and inaccurate estimations of the break points in the SPWAL. Future work will be to improve the estimation of the break points.

Although it did not provide most adequate prediction for risk levels, the GJR-

GARCH-PWAL model did most appropriately fit the five return series in the forecasting period, and was most favoured by a range of formal goodness-of-fit measures and hypothesis tests. This interesting finding may prompt a future discussion on model selection: shall we choose the most adequate forecasting models or the best fitting models?

5 Discussion

Risk always exists in financial activity, it is not a bad thing itself; it is the misunderstanding and mismanagement of risk that may jeopardize the global financial system and the world economy. The recent global financial crisis casted a skeptical outlook on market participants' abilities to provide adequate coverage for dynamic changing risk levels. Basell II advocated using one of the most well-known risk measurement methods, the Value-at-Risk (VaR), to determine the benchmark of regulatory capital as risk buffers. A backtesting procedure was recommended to evaluate the performance of VaR models based on observed violations. However, VaR was criticized for its incoherency and its failure to calibrate the size of expected loss given the occurrence of violation. As a coherent risk measurement method, the conditional VaR or expected shortfall (ES) is able to measure the magnitude of loss in extreme cases. In literature, the backtesting ES is considered difficult and usually complicated, which may be one reason for its absence in regulations. However the benefit of this alternative method makes it an attractive topic in financial practice and academic literature. This thesis addresses the problem by evaluating ES at an equal quantile level based on the fact that ES falls at a certain quantile for parametric ES models.

This thesis focuses on both risk measurement methods, proposes parametric models for VaR and ES forecasts, and explores the preditive and fitting capabilities of a range of models, including non-parametric and semi-parametric methods. The parametric methods for VaR and ES have their focus on two aspects of financial time series: the time-varying property, such as volatility; and the shape of conditional distribution. The daily return series are observed mildly negatively skewed both unconditionally and conditionally. Several studies discussed the importance of including the skewness of return in financial decisions as early as the 1970s. This thesis proposed three distributions as candidates for the conditional return distributions to account for the potential skewness and heavy tails in financial returns. Further, besides using volatility models, extra dynamics in the financial time series, such as skewness and kurtosis or even higher moments, can be captured by allowing for time-varying shape parameters in these conditional distributions.

The first assumption on the conditional distribution is an asymmetric Laplace (AL) distribution with a single shape parameter. The first four moments, the pdf and cdf of AL distribution are reviewed. A GJR-GARCH volatility model, capturing the leverage effect, is combined with AL. Under this frame, the VaR, close-form ES, and ES quantile equation are derived. One finding is that the quantile level for ES is fixed and independent of the shape parameter in AL distributions. This proposed GJR-GARCH-AL model is applied to produce one-day-ahead VaR and ES forecasts for six real return series, over a two-year forecasting period, including the recent global financial crisis. This model is compared with several popular delegates in nonparametric historical methods, semi-parametric methods, and parametric methods. An informal ratio of violation rates, together with three conditional formal tests illustrate that the proposed model outperforms its competitors. In fact, the GJR-GARCH-AL model is the only always conservative risk model throughout the period studied.

The GJR-GARCH-AL model is a conservative model, because it frequently overestimates risk levels, thus has a much lower violation rate, that is, on average 50% lower than the nominal risk levels. It implies the financial institutions have to set aside a much higher regulatory capital based on this model. Regulators prefer conservative models, however, as another essential function of these financial entities is to make profit, the allocation of limited capital matters a lot. Ability to manage the risk and ability to make profit are both important for financial participants to stay competitive in the market. In this thesis, we argue that other than using an extremely conservative model, a more appropriate approach should be able to balance the allocation of regulatory capital and investment capital, by more accurately forecasting dynamic risk levels, thus carefully and properly increasing the investment opportunities in more profitable assets or during economic booms. The VaR and ES focus on the tail end of conditional return distributions, and the conservativeness of the GJR-GARCH-AL model mostly comes from the fat tails of the AL distribution. Therefore, we proposed the use of a standardized two-sided Weibull (STW) conditional return distribution, coupled with a volatility model.

STW distribution is a more general distribution, of which AL distribution is a special case. It is also more flexible with a wider range of skewness and kurtosis. With similar

negative skewness and kurtosis, we found that the left tail of a STW distribution is less heavy than that of AL, but fatter than that of a Student t or skewed Student t, between 2 to 6 standard deviations that we focus on. This thesis derives the first four moments, the pdf and cdf, as well as VaR and ES functions for the STW distribution, which were absent in the literature. The quantile functions for STW ES denpends on STW parameters, similarly as that the quantile function for the Student t ES depends on the degrees of freedom.

As various asymmetric conditional distribution and asymmetric volatility models have been developed and used for parametric VaR and ES models, we are interested in examining the impact of these asymmetric assumptions on the predictive accuracy for risk levels. Thus, this thesis compares a number of combinations of the following conditional distributions: Gaussian, Student t, skewed Student t and AL, that is two symmetric and two asymmetric distributions; GARCH, GJR-GARCH, threshold GARCH, and smooth-transition threshold GARCH, that is one symmetric and three asymmetric volatility models. They are applied to seven real daily return series, over a four-year forecasting period, covering the recent financial crisis. The finding is that models with conditional STW errors have produced highly accurate predictions for both VaR and ES levels and could not be frequently rejected or bettered across several criteria, both before the GFC hit markets as well as during and after the GFC period. The STW definitely improves the modelling of tails of conditional return distributions, thus, is a new option for both the regulators and the financial institutions.

From the empirical study, the estimated STW distributions are usually bimodal, which misses the fact that the return distribution has a single mode, unconditionally and conditionally. Although we focus on forecasting the tail risk, improving the fitting of the whole conditional distribution is also our task. For the purpose of better model fitting, we propose a partitioned distribution with Weibull tails and an AL centre. It is able to capture the potential skewness and heavy tails in the conditional return distribution at least as equally well as the Weibull and AL (PWAL). The mean and variance functions, pdf, cdf, VaR and ES functions are derived. The ES quantile of this standardized PWAL (SPWAL) distribution is in the same form of the STW ES quantile function. A model

with a GJR-GARCH volatility specification coupled with SPWAL errors is applied to forecast VaR and ES for five financial series. The model is compared with a number of alternative VaR and ES models, including semi-parametric and parametric methods. The proposed model ranked second or third, thus is less favoured than models with STW or AL errors by the non-formal and formal coverage tests and DQ tests. However, it is the best fitting model and cannot be rejected in most series by various goodness-of-fit criteria.

This thesis adopts Bayesian methods for the parameter estimation and forecasting. In Chapter 2 and Chapter 3 an adaptive Monte Carlo Markov chain sampling scheme with a mixture of Gaussian proposal distributions in the Metropolis-Hastings algorithm was employed. This method demonstrates improved inferential properties over a standard competitor. Then a more powerful computational method with a general and flexible mixture of Student t proposal densities are adopted, which speeds up the convergence of the Monte Carlo Markov chain and improves the estimation accuracy.

Last but not least, this thesis uses Bayesian methods to account for the parameter uncertainty. However, as recently demonstrated by Lahiri and Martin (2010), the Bayesian methods is also effective in dealing with model uncertainty. In future projects, Bayesian model averaging (BMA) will be considered to obtain Bayesian forecasting for risk measures. That is, instead of deriving forecasts from all models as in this thesis, and performing model choice using various criteria, forecasts can be generated from a combination of models, using a Bayesian model selection process. This mechanism is more coherent than the standard statistical methos used in this thesis, Hoeting, Madigan, Raftery and Volinsky (1999) had a comprehensive review of BMA.

Bibliography

Aas K. and Haff I. (2006), "The generalized hyperbolic skew Student's t-distribution," *Journal of Financial Econometrics*, **4**, 275-309.

Ait-Sahalia, Y., and Brandt, M.W. (2001), "Variable selection for portfolio choice," *Journal of Finance*, **56**, 1297-1355.

Allen, S.L. (2003), "Financial risk management: A practitioner's guide to managing market and credit Risk," John Wiley and Sons, New York.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2001), "The distribution of realized exchange rate volatility," *Journal of the American Statistical Association*, **96**, 42-55.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2003), "Modeling and forecasting realized volatility," *Econometrica*, **71**, 529-626.

Artzner, P., Delbaen, F., Eber, J.M., and Heath, D. (1997), "Thinking coherently," *Risk*, **10**, 68-71.

Artzner, P., Delbaen, F., Eber, J.M., and Heath, D. (1999), "Coherent measures of risk," *Mathematical Finance*, **9**, 203-228.

Azzalini, A., Capitanio, A. (2003), "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution," *Journal of the Royal Statistical Society*, Series B **65**, 367-389.

Bakshi, G. and Panayotov, G. (2007), "The capital adeqacy puzzle," working paper, Smith Business School, University of Maryland.

Barnea, A. and Downes, D. H. (1973), "A re-examination of the empirical distribution of stock price changes," *Journal of the American Statistical Association*, **168**, 348-50.

Behrens, C.N., Lopes, H.F. and Gamerman, D. (2004), "Bayesian analysis of extreme events with threshold estimation," *Statistical Modeling: An International Journal*, **4**, 227-44.

Berkowitz, J. (2001), "Testing density forecasts, with applications to risk management," *Journal of Business and Economic Statistics*, **19**, 465-474.

Berkowitz, J., Christoffersen, P., and Pelletier, D. (2010), "Evaluating Value-at-Risk models with desk-Level data," *Management Science*, in press.

BIS Committee on the Global Financial System, "A review of financial market events in Autumn 1998," October 1999.

Bjornstad, J.F. (1990), "Predictive likelihood: a review," *Statistical Science*, **5**, 242265.

Black, F. (1976), "Studies in stock price volatility changes," *American Statistical Association Proceedings of the Business and Economic Statistics Section*, 178-181.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroscedasticity," *Journal of Econometrics*, **31**, 307-327.

Bollerslev, T. (1987), "A conditionally heteroscedastic time series model for speculative prices and rates of return," *Review of Economics and Statistics*, **69**, 542-547.

Bollerslev, T. and Mikkelsen, H.O. (1996), "Modeling and pricing long memory in stock market volatility," *Journal of Econometrics*, **73**, 151-184.

Bollerslev, T. and Mikkelsen, H.O. (1999), "Long-term equity anticipation securities and stock market volatility dynamics," *Journal of Econometrics*, **92**, 75-99.

Brooks, C. (2001), "A double-threshold GARCH model for the French France/Deutschmark exchange rate," *Journal of Forecasting*, **20**, 135-145.

Butler, R.W. (1986), "Predictive likelihood inference with applications," *Journal of the Royal Statistical Society*, **Series B 48**, 138.

Caporin, M. and McAleer, M. (2006), "Dynamic asymmetric GARCH," *Journal of Financial Econometrics*, **4**, 385-412.

Chen, C.W.S, Gerlach, R., Choy, S.T.B. and Lin, C. (2010), "Estimation and inference for exponential smooth transition nonlinear volatility models", *Journal of Statistical Planning and Inference*, **140**, 719-33.

Chen, C.W.S., Gerlach, R. and Lin, E.M.H. (2008), "Volatility forecasting using threshold heteroskedastic models of the intra-day range," *Computational Statistics & Data Analysis*, **52**, 2990-3010.

Chen, C.W.S., Gerlach, R., Lin, E.M.H. and Lee, W.C.W. (2011), "Bayesian forecasting for financial risk management, pre and post the global financial crisis," *Journal of Forecasting*, forthcoming.

Chen, C.W.S. and So, M.K.P. (2006), "On a threshold heteroscedastic model," *International Journal of Forecasting*, **22**, 73-89.

Chen, Q., Gerlach, R. and Lu, Z. (2011), "Bayesian Value-at-Risk and expected shortfall forecasting via the asymmetric Laplace distribution," *Computational Statistics & Data Analysis*, in press.

Chen, Y.T. (2001), "Testing conditional symmetry with an application to stock returns," working paper, Institute for Social Science and Philosophy, Academia Sinica.

Christoffersen, P. (1998), "Evaluating interval forecasts," *International Economic Review*, **39**, 841-862.

Corsi, F. (2004), "A simple long memory model of realized volatility," working paper.

Davison, A.C. (1986), "Approximate predictive likelihood," *Biometrika*, **73**, 323332.

Dempster, MAH (2002), "Risk management: Value at Risk and beyond," *Cambridge University Press*, Cambridge.

Diebold, F.X., Gunther, T. and Tay, A.S. (1998), "Evaluating density forecasts with applications to finance and management," *International Economic Review*, **39**, 863-883.

Diebold, F.X., Gunther, T. and Tay, A.S. (1998), "Multivariate density forecast evaluation and calibration in financial risk management: high frequency returns on foreign exchange," *Review of Economics and Statistics*, **81**, 661-673.

Dowd, K. (1998), "Beyond Value-at-Risk: The new science of risk management," *Wiley*.

Duffie, D. and Pan, J. (1997) "An overview of Value-at-Risk," *Journal of Derivatives*, **4**, 7-49.

Engle, R. F. (1982), "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations," *Econometrica*, **50**, 987-1007.

Engle, R. F., and Manganelli, S. (2004), "CAViaR: conditional autoregressive value at risk by regression quantiles," *Journal of Business and Economic Statistics*, **22**, 367-381.

Engle, R.F. and Russell, J. (1998), "Autoregressive conditional duration: A new model for irregulatory spaced transaction data," *Econometrica*, **66**, 1127-1162.

Ewerhart, C. (2002), "Banks, internal models and the problem of adverse selection,", working paper, University of Bonn.

Fama, EUGENE F. (1965), "Portfolio analysis in a stable Paretian market," *Management Science*, **11**, 404-419.

Fofack, H. and Nolan, J. (1999), "Tail behavior, modes and other characteristics of stable distributions," *Extremes*, **2**, 39-58.

Friend, I. and Westerfield, R. (1980), "Co-skewness and capital asset pricing," *Journal of Finance*, **35**, 897-913.

Geman, S. and Geman., D. (1984), "Stochastic relaxation, Gibbs distribution and Bayesian restoration of images," *IEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721741.

Gencay, R. and Selcuk, F. (2004), "Co-skewness and capital asset pricing," *International Journal of Forecasting*, **20**, 287-303.

Geraci, M., Bottai, M. (2007), *Quantile regression for longitudinal data using the asymmetric Laplace distribution*, *Biostatistics*, **8**, 140-154.

Gerlach, R., Carter, C.K. and Kohn, R. (1999), "Diagnostics for time series Analysis," *Journal of Time Series Analysis*, **20**, 309-330.

Gerlach, R., Chen, C.W.S. (2008), "Bayesian inference and model comparison for asymmetric smooth transition heteroscedastic models," *Statistics and Computing*, **18**, 391-408.

Geweke, J. and Keane, M. (2007), "Smoothly mixing regressions," *Journal of Econometrics*, **138**, 252-290.

Gilks, W., Richardson, S., Spiegelhalter, D. (1996), "Markov chain Monte Carlo in practice," *Chapman and Hall*.

Giordani, P. and Kohn, R. (2008), "Efficient Bayesian inference for multiple change-point and mixture innovation models," *Journal of Business and Economic Statistics*, **26**, 66-77.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993), "On the relation between the expected value And the volatility of the nominal excess return on stock," *Journal of Finance*, **48**, 1779-1801.

Gonzalez-Rivera, G. (1998), "Smooth transition GARCH models," *Studies in Nonlinear Dynamics and Econometrics*, **3**, 61-78.

Griffin, J.E. and Steel, M.F.J. (2006), "Order-based dependent Dirichlet processes," *Journal of the American Statistical Association*, **101**, 179-194.

Guermat, C. and Harris, R.D.F. (2001), "Robust conditional variance estimation and value-at-risk," *Journal of Risk*, **4**, 25-41.

Haas, M., Mittnik, S. and Paolella, M.S. (2009), "Asymmetric multivariate normal mixture GARCH," *Computational Statistics & Data Analysis*, **53**, 2129-2154.

Hansen, B. E. (1994), "Autoregressive conditional density estimation," *International Economic Review*, **35**, 705-730.

Hastings, W. K. (1970), "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, **57**, 97-109.

Havery, C.P. and Siddique, A. (1999), "Autoregressive conditional skewness," *Journal of Financial and Quantitiative Analysis*, **34**, 465-488.

Havery, C.P. and Siddique, A. (2000), "Conditional skewness in asset pricing tests," *Journal of Finance*, **55**, 1263-1295.

Hinkley, D.V. and Revankar, N.S. (1977), "Estimation of the Pareto law from underreported data," *Journal of Econometrics*, **5**, 1-11.

Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T. (1999), "Bayesian model averaging - a tutorial," *Statistical Science*, **14**, 382-401.

Holton, G.A. (2003), "Value-at-Risk: theory and practice," *Academic Press*.

Hoogerheide, L.F., Kaashoek, J.F. and van Dijk, H.K. (2007), "On the shape of posterior densities and credible sets in instrumental variable regression models with reduced rank: an application of flexible sampling methods using neural networks," *Journal of Econometrics*, **5**, 1-11.

Hoogerheide, L.F., van Dijk, H.K. (2010), "Bayesian forecasting of Value at Risk and expected shortfall using adaptive importance sampling," *International Journal of Forecasting*, **26**, 231-247.

Jensen, M. J. and Maheu, J. M. (2010), "Bayesian semiparametric stochastic volatility modeling", *Journal of Econometrics*, **157**, 306-316.

Jondeau, E. and Rockinger, M. (2006), "The Copula-GARCH model of conditional dependencies: an international stock market application," *Journal of International Money and Finance*, **25**, 827-853.

Jorion, P. (2000), "Risk management lessons for long-term capital management," *European Financial Management*, **6**, 277-300.

Kass, R. and Raftery, A. (1995), "Bayes factors and model uncertainty," *Journal of the American Statistical Association*, **90**, 773795.

Kerkhof, F.L.J. and Melenberg, B. (2004), "Backtesting for risk-based regulatory capital," *Journal of Banking and Finance*, **28**, 1845-1865.

Koenker, Roger W. and Bassett, Gilbert. Jr. (1978), "Regression quantiles," *Econometrica*, **46**, 33-50.

Kraus, A. and Litzenberger, R.H. (1976), "Skewness preference and the valuation of risk assets," *Journal of Finance*, **31**, 1085-1100.

Kupiec, P. H. (1995), "Techniques for verifying the accuracy of risk measurement models," *The Journal of Derivatives*, **3**, 73-84.

Lahiri, K. and Martin, G.M. (2010), "Bayesian Forecasting in Economics: Editorial", Special Issue of *International Journal of Forecasting*, **26**, 211-215.

Li, C.W. and Li, W.K. (1996), "On a double-threshold autoregressive heteroscedastic time series model," *Journal of Applied Economics*, **11**, 253-274.

Li, F., Villani, M. and Kohn, R. (2009). "Flexible Modeling of Conditional Distributions using Smooth Mixtures of Asymmetric Student T Densities", forthcoming in *Journal of Statistical Planning and Inference*.

Lim, K. (1989), "A new test of the three moment capital asset pricing model," *Journal of Financial and Quantitative Analysis*, **24**, 205-216.

Liu, Shi-Miin and Brorsen, B. Wade (1995), "GARCH-stable as a model of futures price movements," *Review of Quantitative Finance and Accounting*, **5**, 155-67.

Lu, Z., Huang, H., Gerlach, R. (2010), "Estimating Value-at-Risk: from JP Morgan's standard-EWMA to skewed-EWMA forecasting", University of Sydney Working Paper.

Madan D. and Seneta E. (1990), "The variance gamma (v.g.) model for share market returns," *Journal of Business*, **63**, 511-524.

Malevergne, Y. and Sornette, D. (2004), "VaR-Efficient portfolios for a class of super and sub-exponentially decaying assets return distributions", *Quantitative Finance*, **4**, 17-36.

Mandelbrot, B. (1963), "The variation of certain speculative prices," *Journal of Business*, University of Chicago Press, **36**, 394.

McAleer, M. (2005), "Automated inference and learning in modeling financial volatility," *Econometric Theory*, **21**, 232-261.

McAleer, M. and da Veiga, B. (2008), "Single-index and portfolio models for forecasting Value-at-Risk thresholds," *Journal of Forecasting*, **27**, 217-235.

McCulloch, J. Huston (1985), "Interest-risk sensitive deposit insurance premia," *Journal of Banking and Finance*, **9**, 137-156.

McNeil, A.J., Frey, R. (2000), "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach," *Journal of Empirical Finance*, **7**, 271-300.

McNeil, A.J., Frey, R., Embrechts, P. (2005), "Quantitative risk management", *New Jersey: Princeton University Press*, 283.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equation of state calculations by fast computing machines," *Journal Chemical Physics*, **21**, 1087-1092.

Mittnik, M. and Paolella, M.S. (2000), "Conditional density and value-at-risk prediction of Asian currency exchange rates," *Journal of Forecasting*, **19**, 313-333.

Mittnik, M. and Rachev, S.T. (1989), "Stable distributions for asset returns," *Applied Mathematics Letters*, **2**, 301-304.

Morgan, J. P. (1996), "RiskMetrics", *Technical Document*, *fourth ed.*, New York.

Nelson, D. B. (1991), "Conditional heteroscedasticity in asset returns: a new approach econometrica," *Econometrica*, **59**, 347-370.

Nelson, W. (1982), "Applied life data analysis", John Wiley and Sons, New York.

Ohanissian, A., Russell, J. R., and Tsay, R. S. (2003), "Using temporal aggregation to distinguish between true and spurious long memory," working paper.

Panagiotelis, A., Smith, M. (2010), "Bayesian skew selection for multivariate models," *Computational Statistics and Data Analysis*, **24**, 1824-1839.

Poon, S. H. and Granger,C. (2003), "Forecasting volatility in financial markets: A review," *Journal of Economic Literature*, **41**, 478-539.

Richardson, M. and Smith, T. (1993), "Asymptotic filtering theory for univariate ARCH models," *Journal of Financial and Quantitative Analysis*, **24**, 205-216.

Smith, J. Q. (1985), "Diagnostic checks of non-standard time series models," *Journal of Forecasting*, **4**, 283291.

So, M. and Chan, K.S. (2011), "Forecasting tail risk in financial time series using a mixture of distribution approach," *4th CSDA International Conference on Computational and Financial Econometrics 2011*, London.

Sornette, D., Simonetti, P. and Andersen, J.V. (2000b), "Φ^q-field theory for portfolio optimization : fat-tails and non-linear correlations," *Physics Reports*, **335**, 19-92.

Taylor, S. (1986), "Modelling financial time series," *John Wiley & Sons*, New York.

Taylor, S. (2000), "Consequences for option pricing of a long memory in volatility," working paper.

Thavaneswaran, A., Peiris, S. and Singh, J. (2008), "Derivation of kurtosis and option pricing formulas for popular volatility models with applications in finance," *Communications in Statistics: Theory and Methods*, **37**, 1799 1814.

Tong, H. and Xia, Y. and Li, W.K. (2007), "Threshold variable selection using nonparametric methods," *Statistica Sinica*, **17**, 265-288.

Tsay, R.S. (2010), "Analysis of financial time series," 3rd Editioin, John Wiley and Sons, New York.

Vlaar, P. J. G. and Palm, F. C. (1993), "The message in weekly exchange rates in the European monetary system: mean reversion, conditional heteroscedasticity, and jumps," *Journal of Business and Economic Statistics*, **11**, 351-360.

Vrontos, I., Dellaportas, P. and Politis, D. N. (2000), "Full Bayesian inference for GARCH and EGARCH models," *Journal of Business & Economic Statistics*, textbf18, 187-98.

Weibull, W. (1951), "A statistical distribution function of wide applicability," *Journal of Applied Mechanics -Trans. ASME*, **18**, 293-297.

Wong, K. (2008), "Backtesting trading risk of commercial banks using expected shortfall," *Journal of Banking and Finance*, **32**, 1404-1415.

Yamai, Y. and Yoshiba, T. (2005), "Value-at-Risk versus expected shortfall: A practical perspective," *Journal of Banking and Finance*, **29**, 997-1015.

Yu, K., Lu, Z., Stander, J. (2003), "Quantile regression: applications and current research area," *The Statistician*, **52**, 331-350.

Yu, K., Zhang, J. (2005), "Distributions and applications – a three-parameter asymmetric Laplace distribution and its extension," *Communications in Statistics-Theory and Methods*, **34**, 1867-1879.

Yu, K., Moyeed, RA. (2001), "Bayesian quantile regression," *Statistics and Probability Letters*, **54**, 437-447.

Zakoian, J.M. (1994), "Threshold heteroscedastic models," *Journal of Economic Dynamics and Control*, **18**, 931-944.

Zhao, X., Scarrott, C., Oxley, L. and Reale, M. (2010), "Extreme value modellingfore forecasting market crisis impacts," *Applied Financial Economics*, **20**, 63-72.

Zhu, D., Galbraith, J. (2009), "Forecasting expected shortfall with a generalized asymmetric Student-t distribution," *CIRANO Working Papers*, **2009s-24**, CIRANO.

Appendices

Appendix 1

A necessary and sufficient condition for second-order stationarity of the GJR-GARCH-AL model

Theorem: A necessary and sufficient condition for the existence of a second-order stationary solution to the GJR-GARCH-ALD model in (12) is:

$$\alpha_1 + \beta_1 + \alpha_2 c_p < 1, \tag{46}$$

where

$$c_p = \begin{cases} 2\frac{p^3}{b_p^2} + 2\frac{(1-p)^3}{b_p^2}\left[1 - \exp\left(\frac{2p-1}{1-p}\right)\right] - \mu_\epsilon^2 \ ; \text{if } 0 \le p \le 0.5 \\ 2\frac{p^3}{b_p^2}\exp\left(\frac{1-2p}{p}\right) \ ; \text{if } 0.5 \le p \le 1. \end{cases}$$

Here $b_p = \sqrt{p^2 + (1-p)^2}$. If $p = 1/2$, then $c_p = 1/2$: i.e. in this symmetric error case, the necessary and sufficient condition reduces to that of the traditional GJR-GARCH specification.

Proof: First note from (12) that $I_{\{y_{t-1}<0\}} = I_{\{\epsilon_{t-1}<\mu_\epsilon\}}$, and

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + (\alpha_1 + \alpha_2 I_{\{\epsilon_{t-1}<\mu_\epsilon\}})(\epsilon_{t-1} - \mu_\epsilon)^2\sigma_{t-1}^2 + \beta_1\sigma_{t-1}^2 \\ &= \alpha_0 + \varphi_{t-1}\sigma_{t-1}^2 \ , \end{aligned} \tag{47}$$

where $\varphi_t = (\alpha_1 + \alpha_2 I_{\{\epsilon_t<\mu_\epsilon\}})(\epsilon_t - \mu_\epsilon)^2 + \beta_1$.

We first prove the necessity. Assume y_t is a second-order stationary solution to the GJR-GARCH-AL model. Then $E(y_t^2) = E(\sigma_t^2) < \infty$, which is independent of t. It follows from (12) together with (47) and $\text{Var}(\epsilon_t) = E[(\epsilon_{t-1} - \mu_\epsilon)^2] = 1$ that

$$\begin{aligned} E(y_t^2) &= E(\sigma_t^2) = \alpha_0 + E(\varphi_{t-1}\sigma_{t-1}^2) = \alpha_0 + E(\varphi_{t-1})E(\sigma_{t-1}^2) \\ &= \alpha_0 + E(\varphi_t)E(\sigma_t^2) \\ &= \frac{\alpha_0}{1 - E(\varphi_t)} \ , \end{aligned}$$

by the independence of ϵ_t and σ_t in (12). Therefore $E\varphi_{t-1} < 1$ for $\alpha_0 > 0$ and $E\sigma_t^2 > 0$.

$$\begin{aligned} E(\varphi_t) &= \alpha_1 + \beta_1 + \alpha_2 E\left(I_{\{\epsilon_t<\mu_\epsilon\}}(\epsilon_t - \mu_\epsilon)^2\right) \\ &= \alpha_1 + \beta_1 + \alpha_2 c_p \ . \end{aligned} \tag{48}$$

To show this, note that:

$$E\left(I_{\{\epsilon_t<\mu_\epsilon\}}(\epsilon_t - \mu_\epsilon)^2\right) = \int_{-\infty}^{\mu_\epsilon} (x-\mu_\epsilon)^2 f(x|p)\, dx \ , \tag{49}$$

where $f(x|p)$ is the AL density function given by (). For $p < 0.5$ we have $\mu_\epsilon > 0$, so that

$$\begin{aligned} E\left(I_{\{\epsilon_t<\mu_\epsilon\}}(\epsilon_t - \mu_\epsilon)^2\right) &= \int_0^{\mu_\epsilon} (x-\mu_\epsilon)^2 b_p \exp\left(-\frac{b_p x}{1-p}\right) dx \\ &\quad + \int_{-\infty}^{0} (x-\mu_\epsilon)^2 b_p \exp\left(\frac{b_p x}{p}\right) dx \ , \end{aligned} \tag{50}$$

while for $p > 0.5$ we have $\mu_\epsilon < 0$, so that

$$E\left(I_{\{\epsilon_t<\mu_\epsilon\}}(\epsilon_t - \mu_\epsilon)^2\right) = \int_{-\infty}^{\mu_\epsilon} (x-\mu_\epsilon)^2 b_p \exp\left(\frac{b_p x}{p}\right) dx \ , \tag{51}$$

where in each case integration by parts applied twice results in (47).

We now turn to sufficiency. Under the condition (46), note from (47) that through iterations, we find that

$$\sigma_t^2 = \alpha_0 + \alpha_0 \sum_{j=1}^{\infty} \prod_{i=1}^{j} \varphi_{t-j}. \tag{52}$$

Obviously, owing to the *i.i.d.* property of ϵ_t in (12), (52) is a second-order stationary solution to (47) if we can display

$$E(\sigma_t^2) = \alpha_0 + \alpha_0 \sum_{j=1}^{\infty} E\left(\prod_{i=1}^{j} \varphi_{t-j}\right) = \alpha_0 + \alpha_0 \sum_{j=1}^{\infty} (E(\varphi_t))^j < \infty. \tag{53}$$

This is seen to hold true on noticing (48) and (46). Hence it follows from (12) that there is a stationary solution to the model in (12) under (46).

Appendix 2

Derivation of STW and some of its properties

The mean of a two-sided Weibull random variable can be derived as

$$\begin{aligned}
E(Y) &= \int_{-\infty}^{\infty} y f(y|\lambda_1, k_1, \lambda_2, k_2) dy \\
&= \int_{-\infty}^{0} y \left(\frac{-y}{\lambda_1}\right)^{k_1-1} \exp\left[-\left(\frac{-y}{\lambda_1}\right)^{k_1}\right] dy + \int_{0}^{\infty} y \left(\frac{y}{\lambda_2}\right)^{k_2-1} \exp\left[-\left(\frac{y}{\lambda_2}\right)^{k_2}\right] dy \\
&= \int_{0}^{\infty} -\frac{\lambda_1^2}{k_1}\left(\frac{-y}{\lambda_1}\right) \exp\left[-\left(\frac{-y}{\lambda_1}\right)^{k_1}\right] d\left(\frac{-y}{\lambda_1}\right)^{k_1} \\
&+ \int_{0}^{\infty} \frac{\lambda_2^2}{k_2}\left(\frac{y}{\lambda_2}\right) \exp\left[-\left(\frac{y}{\lambda_2}\right)^{k_2}\right] d\left(\frac{y}{\lambda_2}\right)^{k_2} \\
&= \int_{0}^{\infty} -\frac{\lambda_1^2}{k_1}\left[\left(\frac{-y}{\lambda_1}\right)^{k_1}\right]^{\left(\frac{1}{k_1}+1\right)-1} \exp\left[-\left(\frac{-y}{\lambda_1}\right)^{k_1}\right] d\left(\frac{-y}{\lambda_1}\right)^{k_1} \\
&+ \int_{0}^{\infty} \frac{\lambda_2^2}{k_2}\left[\left(\frac{y}{\lambda_2}\right)^{k_2}\right]^{\left(\frac{1}{k_2}+1\right)-1} \exp\left[-\left(\frac{y}{\lambda_2}\right)^{k_2}\right] d\left(\frac{y}{\lambda_2}\right)^{k_2} \\
&= -\frac{\lambda_1^2}{k_1}\Gamma\left(1+\frac{1}{k_1}\right) + \frac{\lambda_2^2}{k_2}\Gamma\left(1+\frac{1}{k_2}\right). \qquad (54)
\end{aligned}$$

Similarly, the expectation of Y^2 can be derived as

$$\begin{aligned}
E(Y^2) &= \int_{-\infty}^{\infty} y^2 f(y|\lambda_1, k_1, \lambda_2, k_2) dy \\
&= \int_{-\infty}^{0} y^2 \left(\frac{-y}{\lambda_1}\right)^{k_1-1} \exp\left[-\left(\frac{-y}{\lambda_1}\right)^{k_1}\right] dy + \int_{0}^{\infty} y^2 \left(\frac{y}{\lambda_2}\right)^{k_2-1} \exp\left[-\left(\frac{y}{\lambda_2}\right)^{k_2}\right] dy \\
&= \int_{0}^{\infty} \frac{\lambda_1}{k_1} y^2 \exp\left[-\left(\frac{-y}{\lambda_1}\right)^{k_1}\right] d\left(\frac{-y}{\lambda_1}\right)^{k_1} + \int_{0}^{\infty} \frac{\lambda_2}{k_2} y^2 \exp\left[-\left(\frac{y}{\lambda_2}\right)^{k_2}\right] d\left(\frac{y}{\lambda_2}\right)^{k_2} \\
&= \int_{0}^{\infty} \frac{\lambda_1^3}{k_1}\left(\frac{-y}{\lambda_1}\right)^{2} \exp\left[-\left(\frac{-y}{\lambda_1}\right)^{k_1}\right] d\left(\frac{-y}{\lambda_1}\right)^{k_1} \\
&+ \int_{0}^{\infty} \frac{\lambda_2^3}{k_2}\left(\frac{y}{\lambda_2}\right)^{2} \exp\left[-\left(\frac{y}{\lambda_2}\right)^{k_2}\right] d\left(\frac{y}{\lambda_2}\right)^{k_2} \\
&= \int_{0}^{\infty} \frac{\lambda_1^3}{k_1}\left[\left(\frac{-y}{\lambda_1}\right)^{k_1}\right]^{\left(\frac{2}{k_1}+1\right)-1} \exp\left[-\left(\frac{-y}{\lambda_1}\right)^{k_1}\right] d\left(\frac{-y}{\lambda_1}\right)^{k_1} \\
&+ \int_{0}^{\infty} \frac{\lambda_2^3}{k_2}\left[\left(\frac{y}{\lambda_2}\right)^{k_2}\right]^{\left(\frac{2}{k_2}+1\right)-1} \exp\left[-\left(\frac{y}{\lambda_2}\right)^{k_2}\right] d\left(\frac{y}{\lambda_2}\right)^{k_2} \\
&= \frac{\lambda_1^3}{k_1}\Gamma\left(1+\frac{2}{k_1}\right) + \frac{\lambda_2^3}{k_2}\Gamma\left(1+\frac{2}{k_2}\right).
\end{aligned}$$

Thus, the variance of a two-sided Weibull random variable can be expressed as

$$\begin{aligned}
\mathrm{var}(Y) &= E(Y^2) - E^2(Y) \\
&= \frac{\lambda_1^3}{k_1}\Gamma\left(1+\frac{2}{k_1}\right) + \frac{\lambda_2^3}{k_2}\Gamma\left(1+\frac{2}{k_2}\right) - \left[-\frac{\lambda_1^2}{k_1}\Gamma\left(1+\frac{1}{k_1}\right) + \frac{\lambda_2^2}{k_2}\Gamma\left(1+\frac{1}{k_2}\right)\right]^2 \qquad (55)
\end{aligned}$$

The mean of a standardized two-sided Weibull random variable can be derived as

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f(x|\lambda_1, k_1, \lambda_2, k_2) dx \\
&= \int_{-\infty}^{0} x b_p \left(\frac{-b_p x}{\lambda_1}\right)^{k_1-1} \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right] dx + \int_{0}^{\infty} x b_p \left(\frac{b_p x}{\lambda_2}\right)^{k_2-1} \exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right] dx \\
&= \int_{0}^{\infty} \frac{\lambda_1}{k_1} x \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right] d\left(\frac{-b_p x}{\lambda_1}\right)^{k_1} + \int_{0}^{\infty} \frac{\lambda_2}{k_2} x \exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right] d\left(\frac{b_p x}{\lambda_2}\right)^{k_2} \\
&= \int_{0}^{\infty} \frac{-\lambda_1^2}{b_p k_1} \left(\frac{-b_p x}{\lambda_1}\right) \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right] d\left(\frac{-b_p x}{\lambda_1}\right)^{k_1} \\
&+ \int_{0}^{\infty} \frac{\lambda_2^2}{b_p k_2} \left(\frac{b_p x}{\lambda_2}\right) \exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right] d\left(\frac{b_p x}{\lambda_2}\right)^{k_2} \\
&= \int_{0}^{\infty} \frac{-\lambda_1^2}{b_p k_1} \left[\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right]^{\left(\frac{1}{k_1}+1\right)-1} \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right] d\left(\frac{-b_p x}{\lambda_1}\right)^{k_1} \\
&+ \int_{0}^{\infty} \frac{\lambda_2^2}{b_p k_2} \left[\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right]^{\left(\frac{1}{k_2}+1\right)-1} \exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right] d\left(\frac{b_p x}{\lambda_2}\right)^{k_2} \\
&= \frac{-\lambda_1^2}{b_p k_1} \Gamma\left(1+\frac{1}{k_1}\right) + \frac{\lambda_2^2}{b_p k_2} \Gamma\left(1+\frac{1}{k_2}\right). \qquad (56)
\end{aligned}
$$

Similarly, the expectation of X^2 can be derived as

$$
\begin{aligned}
E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x|\lambda_1, k_1, \lambda_2, k_2) dx \\
&= \int_{-\infty}^{0} x^2 b_p \left(\frac{-b_p x}{\lambda_1}\right)^{k_1-1} \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right] dx + \int_{0}^{\infty} x^2 b_p \left(\frac{b_p x}{\lambda_2}\right)^{k_2-1} \exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right] dx \\
&= \int_{0}^{\infty} \frac{\lambda_1}{k_1} x^2 \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right] d\left(\frac{-b_p x}{\lambda_1}\right)^{k_1} + \int_{0}^{\infty} \frac{\lambda_2}{k_2} x^2 \exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right] d\left(\frac{b_p x}{\lambda_2}\right)^{k_2} \\
&= \int_{0}^{\infty} \frac{\lambda_1^3}{b_p^2 k_1} \left(\frac{-b_p x}{\lambda_1}\right)^2 \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right] d\left(\frac{-b_p x}{\lambda_1}\right)^{k_1} \\
&+ \int_{0}^{\infty} \frac{\lambda_2^3}{b_p^2 k_2} \left(\frac{b_p x}{\lambda_2}\right)^2 \exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right] d\left(\frac{b_p x}{\lambda_2}\right)^{k_2} \\
&= \int_{0}^{\infty} \frac{\lambda_1^3}{b_p^2 k_1} \left[\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right]^{\left(\frac{2}{k_1}+1\right)-1} \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right] d\left(\frac{-b_p x}{\lambda_1}\right)^{k_1} \\
&+ \int_{0}^{\infty} \frac{\lambda_2^3}{b_p^2 k_2} \left[\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right]^{\left(\frac{2}{k_2}+1\right)-1} \exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right] d\left(\frac{b_p x}{\lambda_2}\right)^{k_2}
\end{aligned}
$$

$$= \frac{\lambda_1^3}{b_p^2 k_1}\Gamma\left(1+\frac{2}{k_1}\right)+\frac{\lambda_2^3}{b_p^2 k_2}\Gamma\left(1+\frac{2}{k_2}\right).$$

Thus, the variance of a standardized two-sided Weibull random variable is

$$\begin{aligned}
\mathrm{var}(X) &= E(X^2)-E^2(X) \\
&= \frac{\lambda_1^3}{b_p^2 k_1}\Gamma\left(1+\frac{2}{k_1}\right)+\frac{\lambda_2^3}{b_p^2 k_2}\Gamma\left(1+\frac{2}{k_2}\right)-\left[-\frac{\lambda_1^2}{b_p k_1}\Gamma\left(1+\frac{1}{k_1}\right)+\frac{\lambda_2^2}{b_p k_2}\Gamma\left(1+\frac{1}{k_2}\right)\right]^2 \\
&= 1. \qquad (57)
\end{aligned}$$

It can be derived as well that

$$\begin{aligned}
E(X^3) &= \int_{-\infty}^{\infty} x^3 f(x|\lambda_1,k_1,\lambda_2,k_2)dx \\
&= \int_{-\infty}^{0} x^3 b_p\left(\frac{-b_p x}{\lambda_1}\right)^{k_1-1}\exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right]dx+\int_0^{\infty} x^3 b_p\left(\frac{b_p x}{\lambda_2}\right)^{k_2-1}\exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right]dx \\
&= \int_0^{\infty}\frac{\lambda_1}{k_1}x^3\exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right]d\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}+\int_0^{\infty}\frac{\lambda_2}{k_2}x^3\exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right]d\left(\frac{b_p x}{\lambda_2}\right)^{k_2} \\
&= \int_0^{\infty}\frac{-\lambda_1^4}{b_p^3 k_1}\left(\frac{-b_p x}{\lambda_1}\right)^3\exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right]d\left(\frac{-b_p x}{\lambda_1}\right)^{k_1} \\
&+ \int_0^{\infty}\frac{\lambda_2^4}{b_p^3 k_2}\left(\frac{b_p x}{\lambda_2}\right)^3\exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right]d\left(\frac{b_p x}{\lambda_2}\right)^{k_2} \\
&= \int_0^{\infty}\frac{-\lambda_1^4}{b_p^3 k_1}\left[\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right]^{\left(\frac{3}{k_1}+1\right)-1}\exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right]d\left(\frac{-b_p x}{\lambda_1}\right)^{k_1} \\
&+ \int_0^{\infty}\frac{\lambda_2^4}{b_p^3 k_2}\left[\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right]^{\left(\frac{3}{k_2}+1\right)-1}\exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right]d\left(\frac{b_p x}{\lambda_2}\right)^{k_2} \\
&= -\frac{\lambda_1^4}{b_p^3 k_1}\Gamma\left(1+\frac{3}{k_1}\right)+\frac{\lambda_2^4}{b_p^3 k_2}\Gamma\left(1+\frac{3}{k_2}\right).
\end{aligned}$$

Thus, the skewness of a standardized two-sided Weibuul random variable is

$$\begin{aligned}
\mathrm{S}(X) &= E\left[(X-E(X))^3\right] \\
&= E\left[(X-E(X))\left(X^2-2XE(X)+E^2(X)\right)\right] \\
&= E\left[X^3-2X^2E(X)+XE^2(X)-X^2E(X)+2XE^2(X)-E^3(X)\right] \\
&= E\left[X^3-3X^2E(X)+3XE^2(X)-E^3(X)\right] \\
&= E(X^3)-3E(X^2)E(X)+3E^3(X)-E^3(X)
\end{aligned}$$

$$
\begin{aligned}
&= E(X^3) - 3E(X^2)E(X) + 2E^3(X) \\
&= -\frac{\lambda_1^4}{b_p^3 k_1}\Gamma\left(1+\frac{3}{k_1}\right) + \frac{\lambda_2^4}{b_p^3 k_2}\Gamma\left(1+\frac{3}{k_2}\right) \\
&- 3\left[\frac{\lambda_1^3}{b_p^2 k_1}\Gamma\left(1+\frac{2}{k_1}\right) + \frac{\lambda_2^3}{b_p^2 k_2}\Gamma\left(1+\frac{2}{k_2}\right)\right]\left[-\frac{\lambda_1^2}{b_p k_1}\Gamma\left(1+\frac{1}{k_1}\right) + \frac{\lambda_2^2}{b_p k_2}\Gamma\left(1+\frac{1}{k_2}\right)\right] \\
&+ 2\left[-\frac{\lambda_1^2}{b_p k_1}\Gamma\left(1+\frac{1}{k_1}\right) + \frac{\lambda_2^2}{b_p k_2}\Gamma\left(1+\frac{1}{k_2}\right)\right]^3. \qquad (58)
\end{aligned}
$$

The expectation of X^4 can be derived as

$$
\begin{aligned}
E(X^4) &= \int_{-\infty}^{\infty} x^4 f(x|\lambda_1, k_1, \lambda_2, k_2)dx \\
&= \int_{-\infty}^{0} x^4 b_p \left(\frac{-b_p x}{\lambda_1}\right)^{k_1-1} \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right] dx + \int_0^{\infty} x^4 b_p \left(\frac{b_p x}{\lambda_2}\right)^{k_2-1} \exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right] dx \\
&= \frac{\lambda_1^5}{b_p^4 k_1}\Gamma\left(1+\frac{2}{k_1}\right) + \frac{\lambda_2^5}{b_p^4 k_2}\Gamma\left(1+\frac{2}{k_2}\right).
\end{aligned}
$$

Therefore, the kurtosis of a standardized two-sided Weibull random variable is

$$
\begin{aligned}
\mathrm{K}(X) &= E\left[(X - E(X))^4\right] \qquad (59) \\
&= E(X^4) - 4E(X)E(X^3) + 6E^2(X)E(X^2) - 3E^4(X) \\
&= \frac{\lambda_1^5}{b_p^4 k_1}\Gamma\left(1+\frac{2}{k_1}\right) + \frac{\lambda_2^5}{b_p^4 k_2}\Gamma\left(1+\frac{2}{k_2}\right) \\
&- 4\left[-\frac{\lambda_1^2}{b_p k_1}\Gamma\left(1+\frac{1}{k_1}\right) + \frac{\lambda_2^2}{b_p k_2}\Gamma\left(1+\frac{1}{k_2}\right)\right]\left[\frac{-\lambda_1^4}{b_p^3 k_1}\Gamma\left(1+\frac{3}{k_1}\right) + \frac{\lambda_2^4}{b_p^3 k_2}\Gamma\left(1+\frac{3}{k_2}\right)\right] \\
&+ 6\left[-\frac{\lambda_1^2}{b_p k_1}\Gamma\left(1+\frac{1}{k_1}\right) + \frac{\lambda_2^2}{b_p k_2}\Gamma\left(1+\frac{1}{k_2}\right)\right]^2\left[\frac{\lambda_1^3}{b_p^2 k_1}\Gamma\left(1+\frac{2}{k_1}\right) + \frac{\lambda_2^3}{b_p^2 k_2}\Gamma\left(1+\frac{2}{k_2}\right)\right] \\
&- 3\left[-\frac{\lambda_1^2}{b_p k_1}\Gamma\left(1+\frac{1}{k_1}\right) + \frac{\lambda_2^2}{b_p k_2}\Gamma\left(1+\frac{1}{k_2}\right)\right]^4 \qquad (60)
\end{aligned}
$$

These formulas can be used to verify the range of skewness and kurtosis for the STW given in Section 2 Chapter 3.

Appendix 3

A necessary and sufficient condition for second-order stationarity of the GJR-GARCH-STW model

The GJR-GARCH-STW model is:

$$\begin{aligned} y_t &= \epsilon_t \sigma_t, \\ \epsilon_t &\overset{\text{i.i.d.}}{\sim} STW(\lambda_1, k_1), \\ h_t &= \alpha_0 + (\alpha_1 + \alpha_2 I(y_{t-1} < 0))\, y_{t-1}^2 + \beta_1 h_{t-1}. \end{aligned} \tag{61}$$

Theorem: A necessary and sufficient condition for the existence of a second-order stationary solution to the GJR-GARCH-STW model is:

$$\alpha_1 + \beta_1 + \alpha_2 c_p < 1, \tag{62}$$

where

$$c_p = \begin{cases} \mu_\epsilon^2 \frac{\lambda_1}{k_1} \exp\left[-\left(\frac{-b_p \mu_\epsilon}{\lambda_1} \right)^{k_1} \right] + 2\frac{\mu_\epsilon \lambda_1^2}{b_p k_1} \Gamma\left(1 + \frac{1}{k_1}, \mu_\epsilon\right) + \frac{\lambda_1^3}{b_p^2 k_1} \Gamma\left(1 + \frac{2}{k_1}, \mu_\epsilon\right) \text{ ; if } 0.5 \leq \frac{\lambda_1}{k_1} \leq 1 \\ \mu_\epsilon^2 \frac{k_1 - \lambda_1}{k_1} \exp\left[-\left(\frac{-b_p \mu_\epsilon}{k_1 - \lambda_1} \right)^{k_1} \right] + 2\frac{\mu_\epsilon (k_1 - \lambda_1)^2}{b_p k_1} \boldsymbol{\gamma}\left(1 + \frac{1}{k_1}, \mu_\epsilon\right) + \frac{(k_1 - \lambda_1)^3}{b_p^2 k_1} \boldsymbol{\gamma}\left(1 + \frac{2}{k_1}, \mu_\epsilon\right) \\ + \mu_\epsilon^2 \frac{\lambda_1}{k_1} + 2\frac{\mu_\epsilon \lambda_1^2}{b_p k_1} \Gamma\left(1 + \frac{1}{k_1}\right) + \frac{\lambda_1^3}{b_p^2 k_1} \Gamma\left(1 + \frac{2}{k_1}\right) \text{ ; if } 0 \leq \frac{\lambda_1}{k_1} \leq 0.5 \end{cases}$$

Here $b_p = \sqrt{\frac{\lambda_1^3}{k_1}\Gamma\left(1 + \frac{2}{k_1}\right) + \frac{\lambda_2^3}{k_1}\Gamma\left(1 + \frac{2}{k_1}\right) - \left[-\frac{\lambda_1^2}{k_1}\Gamma\left(1 + \frac{1}{k_1}\right) + \frac{\lambda_2^2}{k_1}\Gamma\left(1 + \frac{1}{k_1}\right) \right]^2}$, $\mu_\epsilon = \frac{-\lambda_1^2}{b_p k_1}\Gamma\left(1 + \frac{1}{k_1}\right) + \frac{\lambda_2^2}{b_p k_1}\Gamma\left(1 + \frac{1}{k_1}\right)$ and $\lambda_2 = k_1 - \lambda_1$. $\Gamma(s, x)$ is the upper incomplete gamma function and $\boldsymbol{\gamma}(s, x)$ is the lower incomplete gamma function. If $\frac{\lambda_1}{k_1} = 1/2$, the above incomplete gamma functions become gamma functions, and if $k_1 = 1$, then $c_p = 1/2$: i.e. in this symmetric error case, the necessary and sufficient condition reduces to that of the traditional GJR-GARCH specification. We consider the case for $\frac{\lambda_1}{k_1} \neq \frac{1}{2}$, as follows.

Proof: First note from (61) with STW errors that $I_{\{y_{t-1}<0\}} = I_{\{\epsilon_{t-1}<\mu_\epsilon\}}$, and

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + (\alpha_1 + \alpha_2 I_{\{\epsilon_{t-1}<\mu_\epsilon\}})(\epsilon_{t-1} - \mu_\epsilon)^2 \sigma_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ &= \alpha_0 + \varphi_{t-1} \sigma_{t-1}^2 \ , \end{aligned} \tag{63}$$

where $\varphi_t = (\alpha_1 + \alpha_2 I_{\{\epsilon_t<\mu_\epsilon\}})(\epsilon_t - \mu_\epsilon)^2 + \beta_1$.

We first prove the necessity. Assume y_t is a second-order stationary solution to the GJR-GARCH-TW model. Then $E(y_t^2) = E(\sigma_t^2) < \infty$, which is independent of t. It follows from (61) together with (63) and $\text{Var}(\epsilon_t) = E[(\epsilon_{t-1} - \mu_\epsilon)^2] = 1$ that

$$\begin{aligned} E(y_t^2) &= E(\sigma_t^2) = \alpha_0 + E(\varphi_{t-1}\sigma_{t-1}^2) = \alpha_0 + E(\varphi_{t-1})E(\sigma_{t-1}^2) \\ &= \alpha_0 + E(\varphi_t)E(\sigma_t^2) \\ &= \frac{\alpha_0}{1 - E(\varphi_t)} \ , \end{aligned}$$

by the independence of ϵ_t and σ_t in (61). Therefore $E\varphi_{t-1} < 1$ for $\alpha_0 > 0$ and $E\sigma_t^2 > 0$.

$$\begin{aligned} E(\varphi_t) &= \alpha_1 + \beta_1 + \alpha_2 E\left(I_{\{\epsilon_t<\mu_\epsilon\}}(\epsilon_t - \mu_\epsilon)^2\right) \\ &= \alpha_1 + \beta_1 + \alpha_2 c_p \ . \end{aligned} \tag{64}$$

To show this, note that:

$$E\left(I_{\{\epsilon_t<\mu_\epsilon\}}(\epsilon_t - \mu_\epsilon)^2\right) = \int_{-\infty}^{\mu_\epsilon} (x-\mu_\epsilon)^2 f(x|p)\, dx \ , \tag{65}$$

where $f(x|p)$ is the STW density function given by (22). For $\frac{\lambda_1}{k_1} < 0.5$ we have $\mu_\epsilon > 0$, so that

$$\begin{aligned} E\left(I_{\{\epsilon_t<\mu_\epsilon\}}(\epsilon_t - \mu_\epsilon)^2\right) &= \int_0^{\mu_\epsilon} (x-\mu_\epsilon)^2 b_p \left(\frac{b_p x}{\lambda_2}\right)^{k_1-1} \exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_1}\right] dx \\ &\quad + \int_{-\infty}^{0} (x-\mu_\epsilon)^2 b_p \left(\frac{-b_p x}{\lambda_1}\right)^{k_1-1} \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right] dx \ , \end{aligned} \tag{66}$$

while for $\frac{\lambda_1}{k_1} > 0.5$ we have $\mu_\epsilon < 0$, so that

$$E\left(I_{\{\epsilon_t<\mu_\epsilon\}}(\epsilon_t - \mu_\epsilon)^2\right) = \int_{-\infty}^{\mu_\epsilon} (x-\mu_\epsilon)^2 b_p \left(\frac{-b_p x}{\lambda_1}\right)^{k_1-1} \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right] dx \ , \tag{67}$$

where in each case integration by parts applied twice results in (63).

We now turn to sufficiency. Under the condition (62), note from (63) that through iterations, we find that

$$\sigma_t^2 = \alpha_0 + \alpha_0 \sum_{j=1}^{\infty} \prod_{i=1}^{j} \varphi_{t-j}. \tag{68}$$

Obviously, owing to the *i.i.d.* property of ϵ_t in (61), (68) is a second-order stationary solution to (63) if we can display

$$E(\sigma_t^2) = \alpha_0 + \alpha_0 \sum_{j=1}^{\infty} E\left(\prod_{i=1}^{j} \varphi_{t-j}\right) = \alpha_0 + \alpha_0 \sum_{j=1}^{\infty} (E(\varphi_t))^j < \infty. \tag{69}$$

This is seen to hold true on noticing (64) and (62). Hence it follows from (61) that there is a stationary solution to the model in (61) under (62).

Appendix 4

The VaR and ES of the skewed Student t

The skewed Student t distribution by Hansen (1994) is considered and has the following density function:

$$g(z|\nu,\zeta) = \begin{cases} bc\left(1+\frac{1}{\nu-2}\left(\frac{bz+a}{1-\zeta}\right)^2\right)^{-(\nu+1)/2} & ;\ z<-a/b, \\ bc\left(1+\frac{1}{\nu-2}\left(\frac{bz+a}{1+\zeta}\right)^2\right)^{-(\nu+1)/2} & ;\ z\geq -a/b \end{cases}$$

where $2<\nu<\infty$, and $-1<\zeta<1$. The constants a, b and c are given by

$$\begin{aligned} a &= 4\zeta c\left(\frac{\nu-2}{\nu-1}\right), \\ b^2 &= 1+3\zeta^2-a^2, \\ c &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)}. \end{aligned} \tag{70}$$

The variable is positively skewed to the right when $\zeta>0$, and negatively skewed when $\zeta<0$. The inverse cdf of the skewed Student t distribution $Skewt(\nu,\lambda)$ is:

$$F^{-1}(\alpha|\nu,\zeta) = \begin{cases} \frac{1-\zeta}{b}\sqrt{\frac{\nu-2}{\nu}}F_s^{-1}\left(\frac{\alpha}{1-\zeta},\nu\right)-\frac{a}{b}\ ;\ \alpha<\frac{1-\zeta}{2}, \\ \frac{1+\zeta}{b}\sqrt{\frac{\nu-2}{\nu}}F_s^{-1}\left(0.5+\frac{1}{1+\zeta}\left(\alpha-\frac{1-\zeta}{2}\right),\nu\right)-\frac{a}{b}\ ;\ \alpha\geq\frac{1-\zeta}{2}, \end{cases}$$

Here F_s^{-1} is the inverse cdf of the Student's t distribution.

The Expected Shortfall of a $Skewt(\nu,\zeta)$ at level α for a long position can be calibrated via:

$$ES_\alpha = \frac{-\frac{c(1-\zeta)^2}{b}\frac{\nu-2}{\nu-1}d^{\frac{\nu-1}{2}}-\frac{a(1-\zeta)}{b}F_s\left(\frac{bVaR_\alpha+a}{1-\zeta},\nu\right)}{\alpha}\ ;\ VaR_\alpha<-\frac{a}{b} \tag{71}$$

where

$$d = cos^2\left(arctan\left(-\frac{bVaR_\alpha+a}{(1-\zeta)\sqrt{\nu-2}}\right)\right)$$